Ontology construction: cooking domain

Fernando Batista
Joana Paulo
Nuno Mamede
Paula Vaz
Ricardo Ribeiro
(alphabetical order)

February 9, 2006

Abstract

This paper describes an ontology for the cooking domain, reporting on the ontology building process, its life cycle, applied methodologies, taken decisions and achieved results.

In the past, our research group built a generic dialogue system able to manage specific devices at home, such as TVs, lamps and windows. The cooking domain appeared as an interesting research area, where our technologies could be applied, and techniques could be explored in order to make the system more independent from new domains. The information in the field is vast, and no information could be found in a manner that we could provide for the system. Besides the process of collecting all information, we also became conscious that techniques for managing and organizing such knowledge were essential. The promising and emerging ontologies domain seemed the best compromise, which was the motivation for developing the ontology here presented.

The paper gives details on the steps performed for the building process, which mainly consisted in: specification, knowledge acquisition, conceptualization, implementation and evaluation. The sources of information, used in the knowledge acquisition phase, consisted mainly of books, the internet, and text analysis techniques, as well as brainstorming and cross-validation sessions. Conceptualization consisted on the identification of concepts and groups of concepts and in building classification trees. The knowledge model was formalized using Protégé, which was also used to automatically generate the ontology code. The resulting ontology comprehends four main modules covering the key concepts of the cooking domain – actions, food, recipes, and utensils – and two auxiliary modules units, measures, and equivalencies. All modules were cross-validated in several meetings and informal competency questions were used in order to check the usefulness of the ontology.

Contents

1 Introduction

The main idea of this project emerged when our research group built a generic dialogue system accepting voice commands for home appliances. At that point we had an electronic agent that could answer to voice requests allowing easier control of home devices using only our voice. To test the system independence of the domain we suggested extending the demo to a new domain. We needed something distant from home devices and that had not been thought at specification time. The cuisine domain came up in a group meeting and proved to be a good choice as it brought to our attention some interesting points we had not thought about before. Having chosen the new domain, we needed to instruct the system into that new field. Instead of doing all the work from scratch we decided to explore the recent technique of Ontologies. Doing so, if we managed to "teach" the system how to use the ontological information, we would gain more independence adding new domains, since this would be as easy as plugging an appropriate ontology.

If we were to use an Ontology, we had to build it. This report describes the development process of an ontology for the cooking domain as well as the ontology itself.

The report is structured as follows: section 2 briefly introduces the evolution of ontology development methodologies; section 3 presents related work on the cooking domain; the building process is described in section 4; section 5 presents the developed ontology; the remaining sections address future work and conclusions.

2 State of the art

The work on the ontologies field goes back to the beginning of 1990. From those early years, ontology building methodologies have evolved, and several have been proposed in order to achieve the current state of the art in ontology design.

The first ontologies consisted in practical examples, built from scratch and made available in order demonstrate the usefulness of such technology. By that time no methodologies or guidelines were available to lead or ease the building process. After some experience on the field, Gruber [1993] introduced some principles for the design of ontologies. Gruber's work was the first to describe the role of ontologies in supporting knowledge sharing activities, and presented a set of guidelines for the development of ontologies. The ontology building process became clearer, with the continuous development of several other ontologies. As a consequence, the first methodologies for building ontologies appeared in 1995, leading to the emergence of the ontological engineering field.

According to Pinto and Martins [2004], three different generations of methodologies can be distinguished:

- The first generation corresponds to the first attempt on understanding how ontologies could be built. The building process was a main important issue, letting aside problems, such as maintenance and reuse. Methodologies, such as the ones from TOVE [Grüninger and Fox, 1995] and ENTERPRISE [Uschold et al., 1995] ontologies are examples of ontology building methodologies belonging to this first generation. The lack of maintenance is an important characteristic in this group of methodologies.

- Despite many existing fully working ontologies, topics such a maintenance, the entire life cycle of the ontology, and evaluation became more important. Ontologies from this generation consider longer life cycles and were being used in more experiments. Specification, conceptualization, integration, and implementation are now performed as often as required, during the ontology lifetime. The evaluation, nonetheless considered an important research topic at the time of the first ontologies, should now also occur throughout the development of the ontology. The initial version of the METHONTOLOGY [Fernandez et al., 1997] methodology belongs to this generation.

- With the continuous improvements and appliance of the existing ontologies, many more activities were identified. For example, *reusability* and *configuration management*, became activities of the development process, in later versions of METHONTOLOGY. One must notice that, most of the time, an ontology must be modified in order to be used or integrated in other ontologies. Therefore, in order to allow ontology reuse, different versions of ontology may coexist. The current version of the METHONTOLOGY [López et al., 1999] can be included in this last generation of methodologies.

The process of building an ontology can be categorized into two main categories: either building from scratch or by means of reuse. While the first two of the described generations are concerned with ontologies built from scratch, reusability is more concerned with methodologies of the third generation. Nevertheless, while not directly addressed by most of existing building methodologies, *reusability* is recognized by those methodologies as part of the overall ontology building process.

Evaluation is another important task to be taken into consideration, for guaranteeing the quality of the resulting ontology. All methodologies for building ontologies, recognize *evaluation* as an important activity. One important proposal for ontology evaluation is OntoClean [Guarino and Welty, 2004], which assures an appropriate and consistent hierarchical structure.

Ontologies are now being constructed in collaborative platforms using rapid prototyping methodologies and frameworks, as for example OntoEdit [Sure et al., 2002], but difficulties still occur in activities, such as: ontology integration; use of inference engines; tools in developing stages, etc. Currently, no methodology was established as a standard, and none sufficient mature was found having a considerable user community. As concluded by Pinto and Martins [2004], at the moment "ontology building is more of a craft than an engineering task".

Currently, ontologies are mostly developed and used by a manual process. Most stages from ontology construction, extension, mapping, merging and populating are made by hand, which is ineffective and may cause major barriers to their large-scale. Being so, recent years have seen a surge of interest in the discovery and automatic creation of complex, multirelational knowledge structures. For example, the natural language community is trying to acquire word semantics from natural language texts; database researchers are tackling the problem of schema induction; people building intelligent information agents are researching the learning of complex structures from semi-structured input (HTML, XML); and machine learning community is pursuing the induction of more concise and more expressive knowledge structures. ECAI-2000, IJCAI-2001, ECAI-2002 workshops on Ontology Learning and the KCAP-2001, ECAI-2002, KCAP-2003 workshops on Knowledge Markup / Ontology Population have shown interesting results. A remaining challenge is to evaluate in a quantitative manner how useful or accurate the extracted ontology classes, properties and instances are. This is a central issue as it is currently very hard to compare methods and approaches, due to the lack of a shared understanding of the task at hand and its appropriate metrics.

3 Related ontologies

The motto *ontologies are built to be reused* [Fernandez et al., 1997] conveys in an appropriate manner the ideas originally proposed by Gruber [1993]. Therefore, one of the steps taken to achieve the proposed objective was to survey existing information sources of the cooking domain and check their adequacy. Of these sources, some fall in the spectrum of definitions, presented by [McGuinness, 2003], that details the concept of ontology introduced by Gruber [1993].

USDA National Nutrient Database for Standard Reference is a database made by the United Stated Department of Agriculture to be the major source of food composition data in the United States. In its eighteenth release (SR18) comprehends 7,146 food items and up to 136 food components [United Stated Department of Agriculture, 2005].

AGROVOC is a multi-lingual thesaurus made by the Food and Agriculture Organization of the United Nations (FAO) that has about 17,000 concepts and 3 types of relations (preferred term, related term and broader term) [FAO, 2004].

```
~11165~^~1100~^~Coriander (cilantro) leaves, raw~^~CORIANDER (CILANTRO)
  LEAVES,RAW~^~Cilantro, raw, Chinese parsley, raw~^~^~^~Y~^~Roots, old
  and bruised leaves~^15^~Coriandrum sativum~^6.25^2.44^8.37^3.57
~11167~^~1100~^~Corn, sweet, yellow, raw~^~CORN,SWEET,YELLOW,
  RAW~^~^~^~^~Y~^~35% husk, silk, trimmings; 29% cob~^64^~Zea
  mays~^6.25^2.44^8.37^3.57
~11168~^~1100~^~Corn, sweet, yellow, cooked, boiled, drained,
  without salt~^~CORN,SWT,YEL,CKD,BLD,DRND,
  WO/SALT~^~^~^~^~Y~^~Cob~^45^~^~6.25^2.44^8.37^3.57
```

Figure 1: Excerpt of USDA National Nutrient Database (SR18).

```
<kaon:Label rdf:ID="1036619972084-1331809654" kaon:value="CHICKEN MEAT">
   <kaon:inLanguage rdf:resource="&kaon;en"/>
   <kaon:references rdf:resource="#24000"/>
</kaon:Label>
<kaon:Label rdf:ID="1036619972084-1461706594" kaon:value="VIANDE DE POULET">
   <kaon:inLanguage rdf:resource="&kaon;fr"/>
   <kaon:references rdf:resource="#24000"/>
</kaon:Label>
```

Figure 2: Excerpt of a KAON ontology made from AGROVOC.

Noy and McGuinness [2001] present the development of a wine (main focus), food and appropriate combinations of wine with meals ontology.

```
<owl:Class rdf:ID="WhiteBordeaux">
   <owl:intersectionOf rdf:parseType="Collection">
     <owl:Class rdf:about="#Bordeaux" />
     <owl:Class rdf:about="#WhiteWine" />
   </owl:intersectionOf>
</owl:Class>
<owl:Class rdf:about="#WhiteBordeaux">
   <rdfs:subClassOf><owl:Restriction>
       <owl:onProperty rdf:resource="#madeFromGrape" />
       <owl:allValuesFrom><owl:Class>
           <owl:oneOf rdf:parseType="Collection">
             <owl:Thing rdf:about="#SemillonGrape" />
             <owl:Thing rdf:about="#SauvignonBlancGrape" />
           </owl:oneOf>
       </owl:Class></owl:allValuesFrom>
   </owl:Restriction></rdfs:subClassOf>
</owl:Class>
```

Figure 3: Excerpt of the (mainly) wine ontology of Noy and McGuinness [2001].

Graça et al. [2005] present a specialized wine ontology that covers maceration, fermentation processes, grape maturity state, wine characteristics, and classification according to country and region where the wine was produced.

```
<oxml:concept id="a:Tannage" abstract="false">
   <oxml:documentation language="pt">Processo de vinifica&#231;&#227;o
     em que as pel&#237;culas est&#227;o em contacto com o mosto durante
     a fermenta&#231;&#227;o.</oxml:documentation>
   <oxml:externalRepresentation language="en">
     http://vinho.pt#Tannage</oxml:externalRepresentation>
   <oxml:externalRepresentation language="pt">
     http://vinho.pt#Curtimenta</oxml:externalRepresentation>
   <oxml:subConceptOf concept="a:DEFAULT_ROOT_CONCEPT"/>
</oxml:concept>
```

Figure 4: Excerpt of the wine ontology developed by Graça et al. [2005].

Villarías [2004] describes an ontology of culinary recipes, developed to be used in a semantic querying system for the Web.

```
<rdfs:Class rdf:ID="NON-SWEET-FRUIT-COURSE">
  <daml:intersectionOf rdf:parseType="daml:collection">
     <daml:Restriction>
        <daml:onProperty rdf:resource="#FOOD"/>
        <daml:hasClass rdf:resource="#NON-SWEET-FRUIT"/>
     </daml:Restriction>
     <rdfs:Class rdf:about="#MEAL-COURSE"/>
  </daml:intersectionOf>
  <rdfs:subClassOf
     rdf:resource="#DRINK-HAS-DELICATE-FLAVOR-TO-CLASS-RESTRICTION"/>
  <rdfs:subClassOf
     rdf:resource="#DRINK-HAS-OFF-DRY-SUGAR-TO-CLASS-RESTRICTION"/>
</rdfs:Class>
```

Figure 5: Excerpt of the culinary recipes ontology developed by Villarías [2004].

These ontologies did not cover what was intended with this project: some were too specific, focusing on issues like wine (Noy and McGuinness [2001], Graça et al. [2005]) or nutrients itselves (USDA National Nutrient Database), others not deep enough (Villarías [2004]), concentrating (as stated in their objectives) in building a classification – adequate to a specific application – of part of the knowledge we intended to structure.

Several web pages also try to structure information related to the cooking knowledge domain: Allrecipes.com is a large collection of classified recipes; open directory [Nestscape Communications Corporation, n.a.] – an open project that aims to build the largest human-edited directory of the web – is a general directory that covers also cooking data; Food Resource Oregon State University [n.a.] is digital library hosted by the Oregon State University with food related information. Although structured, this information can only be used as a knowledge source.

4 Building process

The development of the cooking ontology did not follow a specific ontology development methodology, but was strongly influenced by the ideas presented by López et al. [1999]. It is also important to mention the study presented by Pinto and Martins [2004], as a reference to understand how methodologies have evolved and to identify and comprehend the key steps in ontology development.

Although, (inherently) familiar to every team member, there were no cooking domain experts. As so, to better plan the project, the first step was a brainstorm session to try to make a general birds eye-view snapshot of the domain. Six key areas were identified: (i) recipes; (ii) actions; (iii) food; (iv) utensils; (v) times; and, (vi) equivalencies. Considering the achieved insight about the domain and the team experience in the knowledge representation area – even though, none had previous experience in ontology development –, it was then possible to define the project time-chart and task distribution.

Subsequently, to better define the scope of the ontology, informal competency questions related to these areas were formulated. These questions addressed specifically each of the previously identified areas. Figure 6 shows some of the questions.

[recipes]
How do I make recipe *R*?
What are the quantities to use when making recipe *R* for 4 persons?

[actions]
How do I do action *A*?

[times]
Which are the recipes that take less than 10 minutes to make?

[food]
Which recipes have food item *F1*, but not *F2*?
Which are the recipes that have as main ingedient food item *F*?

[utensils]
Which utensils are used in recipe *R*?
Which recipes can be made using the microwave?

[equivalencies]
How many liters is a cup?

Figure 6: Competency questions.

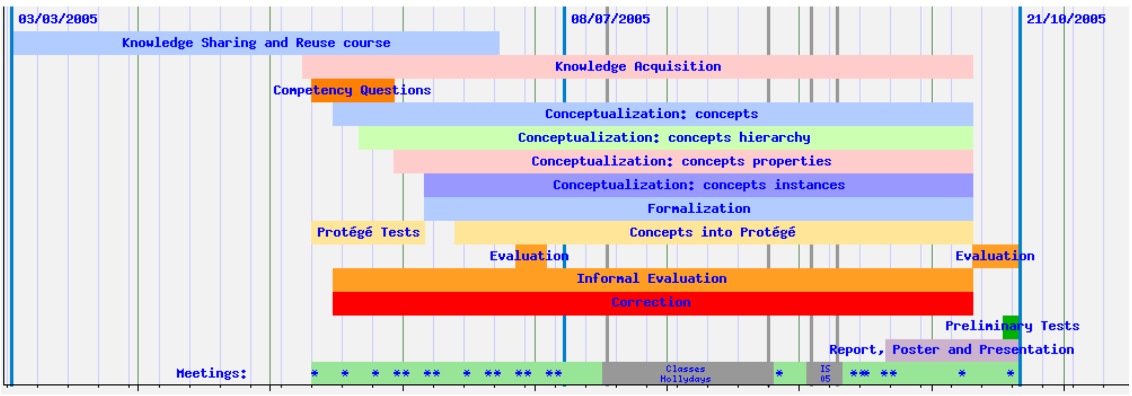

Figure 7: Temporal distribuition of the activities.

Figure 7 shows the temporal distribuition of the activities of the project. Activities like brainstorm sessions, knowledge validation and disambiguation, foundational conceptualizations and formalizations, or evaluation, among others, were done during the several meetings taken along the project. Acquisition, conceptualization, and formalization of knowledge was divided by each of the identified areas and assigned to team members as shown in table 1.

Table 1: Work distribution.

Knowledge area	**Number of team members**
recipes, times, and equivalencies	1
actions	2
food	1
utensils	1

The development of the ontology followed an evolving life cycle, illustrated in figure 8.

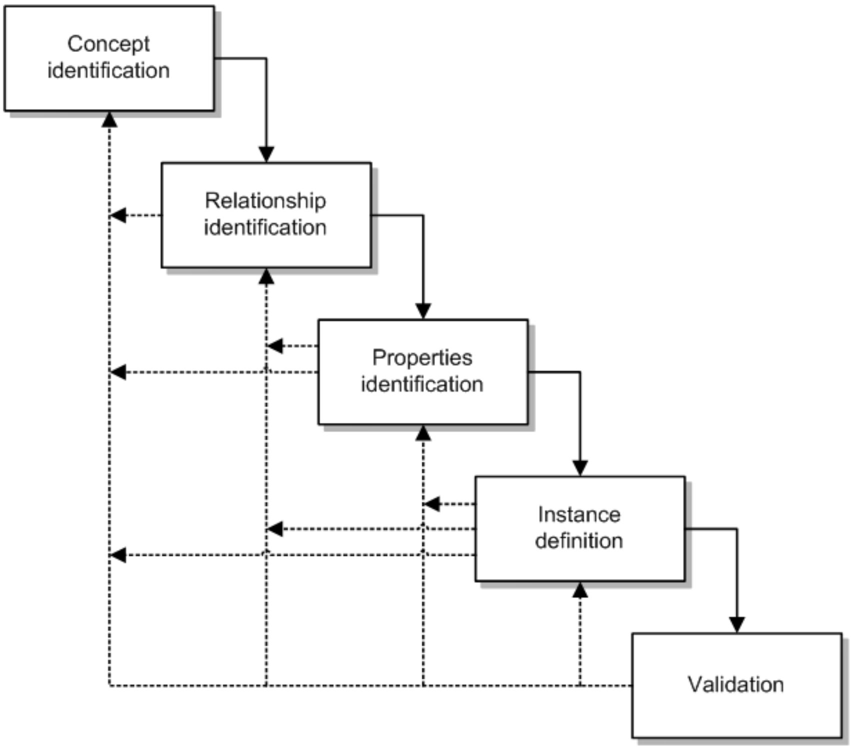

Figure 8: Ontology development.

After some work in the acquisition and conceptualization phases, it became apparent in the selected information sources that time and equivalences were not so relevant as actions, food, recipes or utensils concepts. In the end, the previously identified key areas of the cooking domain evolved into the current four main modules:

- Actions;
- Food;
- Recipes;
- Kitchen utensils;

and, two auxiliary modules: one for units and measures, and the other for equivalencies.

The dialogue system context, the earlier phases of knowledge acquisition, and the competency questions defined the requirements specification of the cooking ontology. Their satisfaction was continuously evaluated by one of the team members, that performed also the role of client (representing L^2F).

4.1 Knowledge acquisition

The knowledge acquisition activity is done simultaneously with the requirements specification phase and decreases as the ontology development process moves forward. The knowledge acquisition was extremely important, because this activity defined the extent of the ontology.

The knowledge acquisition, the first step of the building process, began with the reading and selection of available cooking books. After careful reading, the selection preyed upon "*Pantagruel*" [do Canto, 1997] and "*O grande livro ilustrado da Culinária*" [Terence and Couran, 1982]. The cooking book "*Pantagruel*" and "*O grande livro ilustrado da Culinária*" were took as the main knowledge source and defined the ontology extension and amount of detail. Other books like de Portugal [1987], Activa [2004], and Bernard [1964] and web sites like Arte Digital Design e Publicidade [1997], Wikipedia [2001b], Allrecipes.com [n.a.], and CondéNet [n.a.] were used to clarify concepts. The Google [1995] was used for disambiguation between terms. Typically, the term with the biggest number of occurrences was selected.

After careful reading of "*Pantagruel*" and "*O grande livro ilustrado da Culinária*", the information organization and selection problem was faced. "*Pantagruel*" and "*O grande livro ilustrado da Culinária*" had some inconsistence, even though were very complete. At this stage, the concepts organization was unclear, because concepts were very different and difficult to relate. The way to overcome this problem the other knowledge sources were consulted and the doubts clarified.

The knowledge sources had different views about subjects, but they all agreed on the separation of concepts. For example, almost every source had a description of: the kitchen tools; animal (cow, pork, rabbit, etc.) parts; and fish types. The first step to organize the concepts was this structured knowledge. The recipes were viewed as algorithms used to define a set of basic actions that later were used to describe Recipe concepts.

The knowledge structure found on the knowledge sources was divided in four main subjects: food, actions, utensils, and recipes. Subjects were discussed in every meeting. Later on, it turned out that each subject was like a small ontology, so in the end we had to preform integration.

4.2 Conceptualization

During conceptualization the acquired knowledge was structured. In the cooking ontology development process, the conceptualization phase started early (see figure 7) and it was strongly influenced by the decision of turning each main knowledge area of the domain into a module.

In this phase, each module was assigned to the same team member that had acquired the knowledge within a given sub-domain. The achieved conceptual model was then presented to the whole team and submited to validation. All concepts that interfaced between modules or did not clearly belong to a module were subject to discussion by the whole team. The main problem of this approach was that, although all team members participated in the conceptualization of all modules, the one responsible for the development of the module had in fact a stronger influence on the conceptual model of that module, what could bias the result.

The main activities of the conceptualization task taken in the development of each module were (i) identification of concepts and their properties; (ii) classification of groups of concepts in classification trees; (iii) description of properties; (iv) identification of instances; (v) description of instances. During this phase, the previously mention discussion and validation sessions were also used to indentify the relations between classification trees; debate the conceptualization as a concept or instance of ambivalent notions; and, harmonize the identified properties and their definitions.

A global problem of conceptualization was when to stop. This problem affected all the main modules. For example, food and utensils are clearly open domains: it is always possible to find new food items or develop new utensils. The adopted approach to this issue was to try to conceptualize as many different situations as possible. This would allow to easily include new concepts, since basic structure is already defined, as well as several example conceptualizations. In what concerns actions, the growing possibilities affect not only concepts modelled as classes, but also the instances (since actions are also modelled as instances).

The next subsections describe conceptualizations problems that affected each module of the ontology.

4.2.1 Food

The main difficulty within the conceptualization of the food module was to model the idea that a food item may be part of other food item. For example, a chicken leg is part of a chicken. One possible solution would be to only specialize a food item in its parts. The disadvantage of this conceptualization is that it does not help to understand that the specialized concepts have a PART-OF relation with the generalized one. That specialization and *Parte de Alimento Sólido* (Solid Food Part) concept state that relation in a proper way (for a more detailed view of this hierarchy, see figure 15, in section 5.1). As so, Chicken Leg concept was defined by specializing the Chicken and Leg concepts (that already specialized the Solid Food Part concept).

This also allows to model recipes where food items are not be used as a whole, and can be divided in parts that can be used in different ways. For example, when making chocolate mousse, eggs are separated in egg yolks and whites, which are used in different manners.

Another concept that was not easy to model was the notion of preparation. During the knowledge acquisition phase it was identified that preparations made of ingredients are used in the cooking process described by a recipe. *Preparado* (Preparation) models that notion, where identifiable food items cease to exist and are transformed in an amalgam of ingredients. The Preparation concept conceptualizes the idea of a composite food item, which composition it is the result of a subprocess.

4.2.2 Utensils

The definition of properties for each concept was a major problem in the utensils conceptualization process. Many interesting properties could be considered for each utensil, such as: dimensions, material, geometrical form, power consumption, etc. For example, consider the concept of a Chef's Knife as follows:

> *The Chef's Knife* – also known as the French Chef's Knife – has a very broad blade (called the Flat) and can range in length from six to twelve inches; the eight inch size being the most popular. The Chef's Knife is used for all the chopping, mincing and dicing tasks and is essential for vegetables. This knife is generally used on a cutting board by rocking it on its gently curving edge, using the tip as a stationary pivot. Its broad blade keeps knuckles from hitting the cutting board. Its back is useful to break chicken bones and scrape foods from the board. The flat side is useful for crushing things like garlic.

The previous definition persuaded the introduction of properties, such as: blade size; blade length; common use; type of edge; type of point, etc. For a general cooking knife, a broader set of properties could also be considered, namely: the handle material; blade material (ceramic, plastic, titanium, stainless steel, high-carbon steel), construction (forged, stamped, sintered), type of point (curved, sharp, thin), type of edge (flat ground, hollow ground or serrated), length, etc. The criteria was to define only the properties that could make an utensil, either applicable or not, to a given specific action, defined in the scope of the ontology.

Another major problem found in the utensils conceptualization process, appeared with the introduction of concepts such as *Fogão* (Stove). The stove utensil, broadly used in the kitchen, could not be introduced as is in the ontology, since it usually allows to simultaneously cook more than one preparation. How would someone specify to turn off the heater for just one of the preparations? Notice that the heating source is not actually the stove, but one of its components. The solution for this problem consisted on introducing the concept of *Bico de Placa* (Stove Heater). The figure 9 shows an excerpt of the ontology where the concept *Fonte de Calor* (Heat Source) appears.

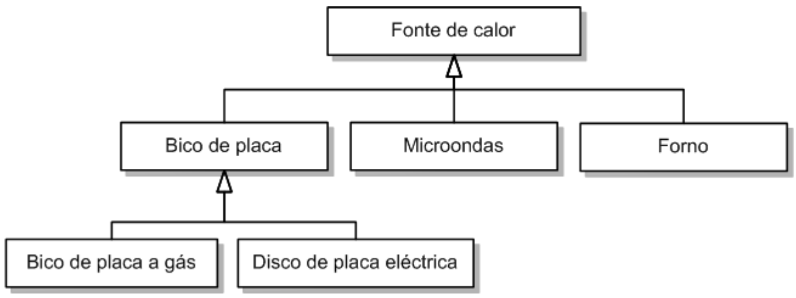

Figure 9: Excerpt of the utensils module, showing the concept *Fonte de Calor*.

Each cooking action can normally be performed with appropriate utensils. During the conceptualization phase, we found at least one cooking task that should always be performed using the hands (for example, *punched potatoes*). At first sight one could think that the hand could not be considered as an utensil, but after a short analysis we found out that the concept of hand could be defined as an instrument, thus satisfying the definition of cooking utensil. The team solved the problem, introducing the concept of "Hand" in the module. The excerpt represented on figure 18 illustrates our solution.

Finally, we found that some utensils could use other accessory utensils. In order to accommodate this kind of concepts, for example *Tampa de Recipiente* (Recipient Cover), we created the concept of *Acessório* (Accessory).

4.2.3 Actions

One of the difficulties of conceptualizing the actions module was to model the idea of repeating sequences of actions to describe actions like *scrambling the eggs until done, but not burned*. First, we decided to create an action called *Ciclo* (Cycle) that could define the sequence of numbered actions to be repeated, the frequency of repetition and the stop condition. This approach proved to be inefficient, because it was generating a great amount of *Acções ordenadas* (Numbered Actions) that could not be reused. *Acções compostas* (Complex Actions) that need the repeated sequence of *Acções primitivas* could be listed, so it was decided to add attributes to these actions that could state the *repetition frequency* and the *stopping condition*. On figure 10, the attribute *Estado de alimento* defines the stop condition. Food is turned until it reaches the inserted state.

A better solution to this problem could be the integration of a process ontology. Process ontologies give a suitable implementations for the cycling and sequencing problems. This solution was not an option, because integrating an ontology is a complex and time consuming task.

Another difficulty was to define time intervals that were dependent on the amount of food or on the power of the heat source. These two values were only known when defining the recipe. To solve this problem *Acções Primitivas Espera* (Waiting Simple Actions) were defined. A waiting action suggests that the user will have to wait until a transformation on the food occurs. The transformation can be visualized or measured, like waiting for the water to bubble or for the rice to toast in the oven.

4.2.4 Recipes

Having the main purpose for this ontology in mind, the Recipes are the key concept as they will be described to the user. In this sense, all the other modules have at least what this module needs to express its knowledge. Therefore, a few recipes were included in order to provide some

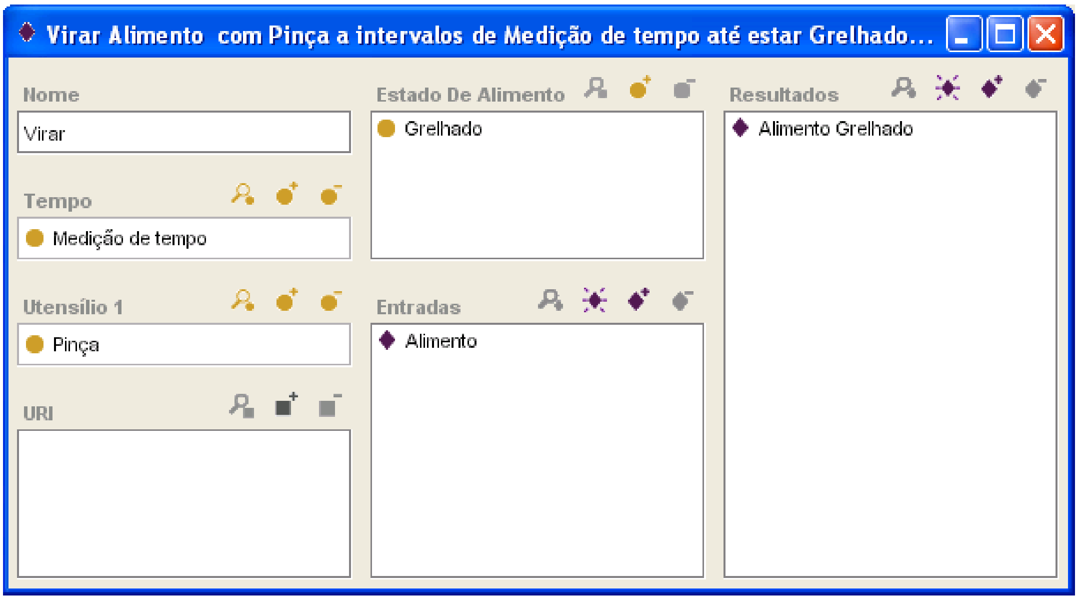

Figure 10: Simple Action *Virar*.

interesting cases. The idea here was to verify whether the ontology could include the relevant cases. Eventually "*Bitoque*" (fried beef and eggs) and "*Bife com Ovo*" (fried beef and fried eggs) were choosen. In the former, the eggs are fried in the same fat as the beef; in the later, they are fried in different fat. The distinction is stated by using different instances of fat. These recipes also needed the notion of "*Preparado*" (c.f. 4.2.1) to identify the fried fat that was fried with the beef and would be used to fry the eggs in the former case.

One of the major problems to solve was to find the best way to use (and reuse) the knowledge formerly stated on a single recipe while describing a new one. This led to the split of knowledge into simpler and smaller concepts in order to ease the reuse of instances (c.f. 5.4.4).

Another problem to solve was to find the best way to represent an ordered sequence of steps. Protégé allows multiple values and has the notion of order (one instance before or after another). The problem comes from the desirable compatibility with other tools as the majority of them don't have it. The concept of ordered list could have been introduced but that would bring more complexity. Instead of that, the solution found was to create auxiliary concepts that give an ordered number to a given instance. The correct usage of this knowledge needs to take this into account but that was considered as a minor problem. The solution is detailed in 5.4.4.

4.2.5 Auxiliary Modules

When describing the concept of Ingredient and Utensil, their measures were to be represented somehow (for example, "*10 grams of flour*" or "*a small knife*"). The first representation to state this knowledge was a slot of Integers. These proved to be insufficient as very different measures were found in the descriptions and the need for equivalences arose. To solve this, a concept was introduced and a module for Units and Measures was added. Later on, it was necessary to state things like "*a small knife has 10 to 15 centimeters*". This was included with some "informal" measures ("*small*", "*medium*", "*big*", etc.) and a concept to represent Equivalences. Also the relations between different units were included (for example, "*1 liter is equal to 1 cubicle decimeter*"). The instances of this module will be included in the database. This is why not all the known equivalences are included in the current version. Notions of equivalence depending on the density or on the utensil, and fuzzy or informal values are taken into consideration. An extended description is presented in 5.5.2.

Later on, to enrich the module of measures and equivalences, to make it independent of the application and to ease its reuse, other measures, not directly related to cooking were included. No similar module was found as a public resource to be imported and included. All the knowledge for this come from Wikipedia [2001c,a,b] and from the official web site of the General Conference on Weights and Measures [1998, 2000].

To categorize the Recipes the usage of the common division of meals was found to be useful. This includes those most seen in cooking books like appetizers, soups, sandwiches, meat, fish, etc. As this separation has a hierarchical organization and could be reused later on, a dedicated module was created (c.f. 5.5.1). Usually this corresponds to the index of a the cooking books and web sites, so eventually it would be one of the main search criteria for the recipes.

4.3 Formalization

Formalization begun with the brainstorms sessions being registered with UML notation [Fowler, 2003] in a blackboard. The idea was to separate the conceptualization and formalization from the implementation, as tool independence is essential to ease future reuse. However, after some initial work, the drag and drop facilities of Protégé [Gennari et al., 2003] were found to be a good help with the corrections eventually done. Indeed, after a few sessions and some tests with the tool, the knowledge was inserted right into it and formalization was made directly into Protégé integrated development environment (IDE). This option revealed to be a good one as corrections were easy to do. The problems here were related to the modules integration task, specially when we needed to merge work made over instances: some losses of work (and time) due to some IDE bugs were experienced.

One of the main issues of formalization concerned relations between concepts. As it was described before, several concepts (Food, Utensils, Actions, ...) belong to hierarchies. The concepts within these hierarchies were associated through IS-A relations. Attribute-based relations were used to associate concepts from the several hierarchies (Food, Utensils, Actions, ...) and the other concepts (Task, Recipe, ...). For example, a recipe uses utensils and that is stated as a slot in the Recipe class.

Another key issue in the formalization phase was to decide if concepts should be formalized as classes or instances. Several discussions and experiments were needed to understand the best way to formalize the previously defined conceptualization.

Some concepts were formalized as classes and their instances use the defined hierarchies as taxonomies (the values of the attributes are of class type). For example, a Recipe has several attributes: one of type class Utensils (that will refer classes); one of type Ingredient (that will be an instance). The figure 11 shows the main division used.

Food and Utensils concepts were formalized as classes. Food (abstract) classes will have no instances as they are used as a taxonomy to characterize Ingredient instances. The existence of Utensils instances depends on the usage of the ontology. For example, considering the context of a dialog system (like the one described earlier), Utensils instances would be the real utensils that the user can operate. Actions were formalized using classes and instances: classes to arrange the hierarchy and instances to describe the leaves of the classification tree.

4.4 Implementation

Protégé IDE is capable of generating, for example, OWL or RDF. In addition to this capability, the creation of a database with complementary information transforms the formalized model into an implemented one. The definition of the database can be similiar to the one used on Protégé. There are examples of it in the current version of the ontology.

One problem that came up is a consequence of dividing the ontology into modules. As each team member was responsible for one module, from time to time it was necessary to merge all the work done. To overcome this problem, team members in charge of this integration task needed to edit the implementation source files.

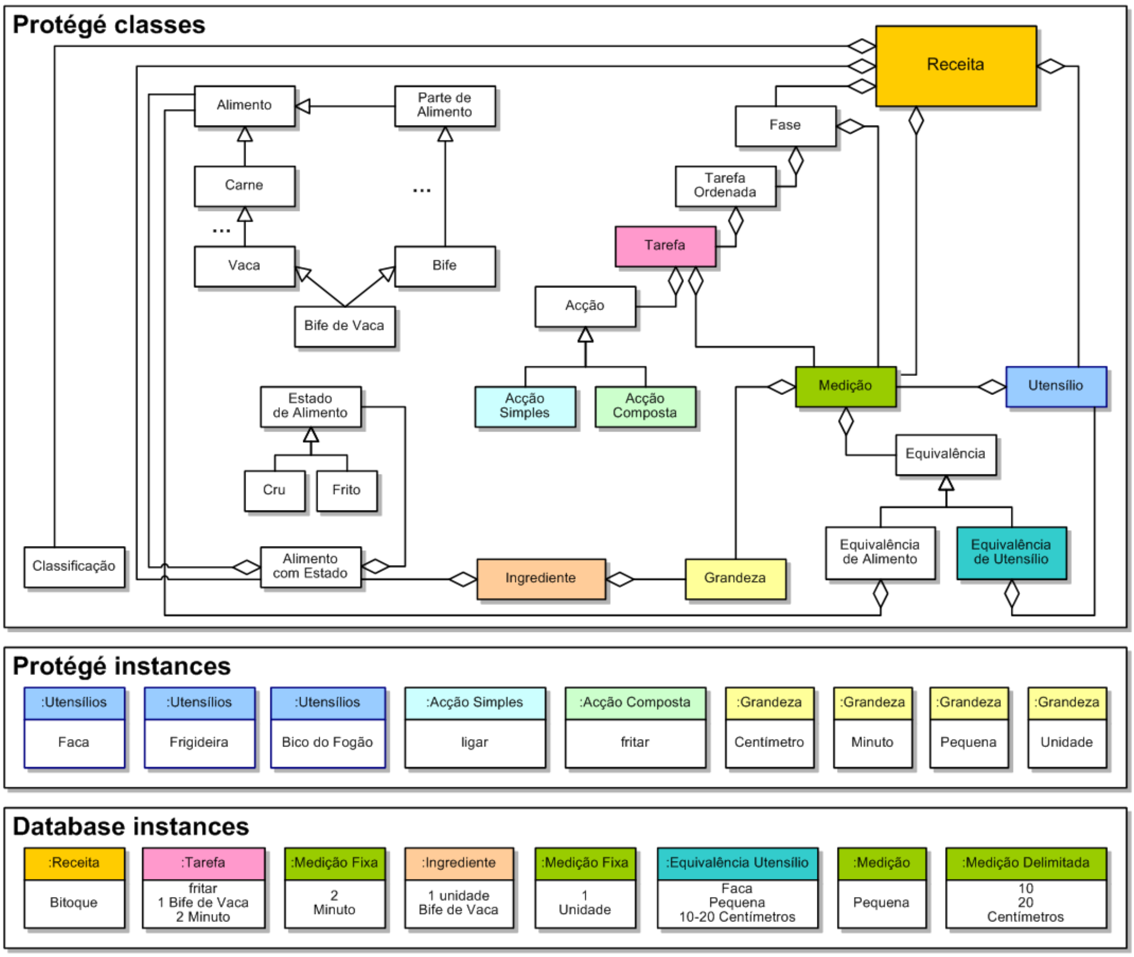

Figure 11: Formalization of the conceptual model.

4.5 Graphical Visualization

Later on, after the work has been done, a graphical representation of the knowledge was needed to have a poster version with a graphical visualization. Some plug-ins that come with the default instalation of Protégé were tested and made great part of the job. At the end, all that was necessary was to choose the proper amount of information to be included in each subset of images, and add some pictures and colors to increase readability.

4.6 Evaluation

In order to evaluate the work, two types of evaluation were performed, neither using a standard methodology like OntoClean [Guarino and Welty, 2004]: an internal evaluation performed by the whole team during the ontology life cycle, and an external evaluation performed by the client. The client took part in the development of the ontology as a supervisor: mainly by asking competency questions and checking if the ontology could answer them (assuming the needed knowledge had been introduced). Since no inference is available, all the verifications were done by checking whether the information was available and if the right relations existed in order to allow the future manipulation of the knowledge.

During the ontology development timeline (see figure 7) competency questions were used to evaluate the ontology. At this stage it was checked if the questions were being correctly answered.

5 Cooking ontology

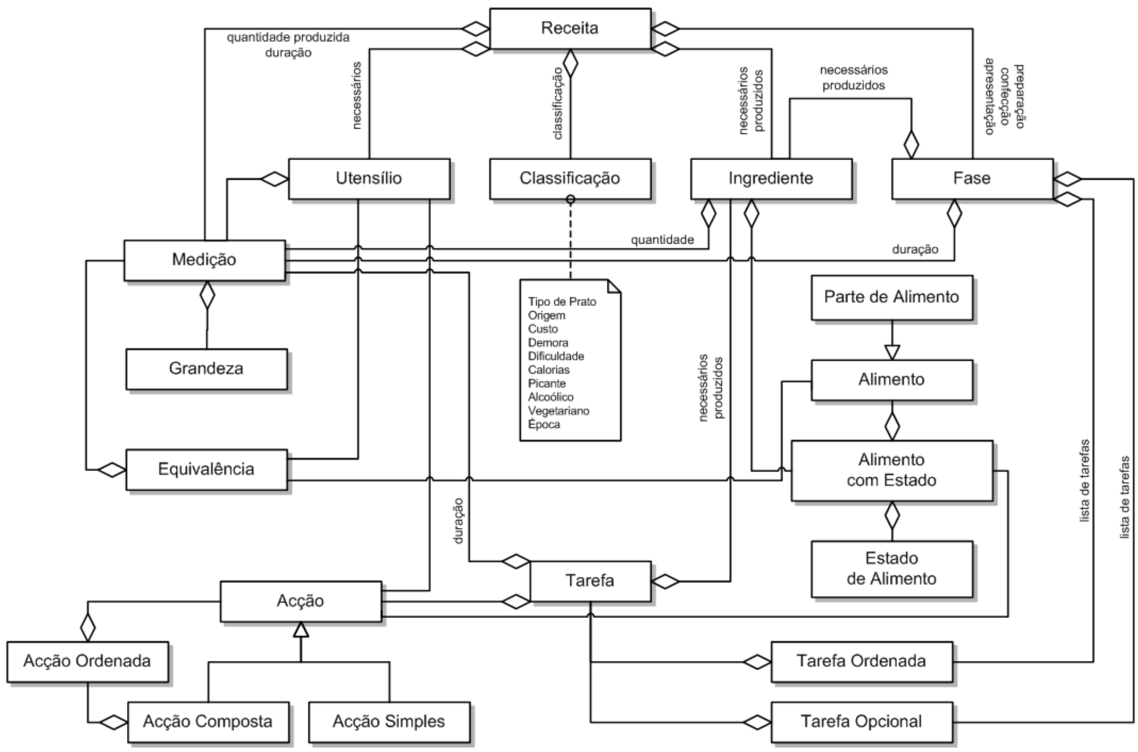

Figure 12: Main concepts.

The figure 12 shows the relations between the main concepts. A *Receita* (Recipe) is organized by *Fases* (phases), which consists on *preparation*, *cooking* and *presentation.* Each phase is a sequence of sorted *Tarefas* (tasks) that may include optional ones. These concepts specify in which position of the sequence each task should be executed. A *Tarefa* is composed by an action and incorporates information about needed and produced Ingredients and its duration time. The Ingredients of all Tasks are also part of the Phase concept. The Phase also has a duration. Each Recipe has a Classification, a list of Ingredients and required Utensils.

In its current state, the cooking ontology has several modules, with about 1151 classes, 92 slots (of which 52 establish relations between classes) and 311 instances.

The following subsections will present a detailed description of each module.

5.1 Food

In what concerns the food module, four main concepts were used in structuring the information:

1. *Alimento* (food);
2. *Parte de alimento sólido* (solid food part);
3. *Estado de alimento* (food state);
4. *Alimento com estado* (food with state);

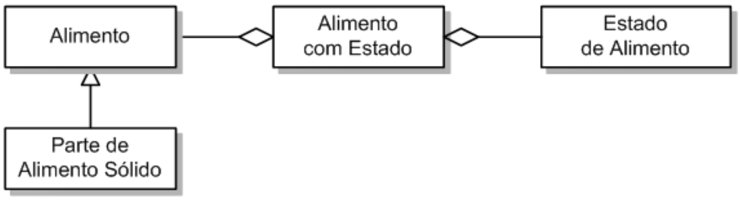

Figure 13: Main food concepts.

5.1.1 *Alimento* (food)

The food concept consists in a substance or a set of substances that are nourishing and digestable and consumed by humans in order to live, grow and have energy.

Food conceptualization is an hierarchy (with multiple inheritance) of food concepts. This organization reflects the one of a cooking book: the food is organized by kind. This approach to food classification can be denominated as a culinary approach. Concepts like fish, meat, milk derived food, cereal, fruit, etc inherit from the main food concept.

An excerpt of food conceptualization can be seen in figure 14.

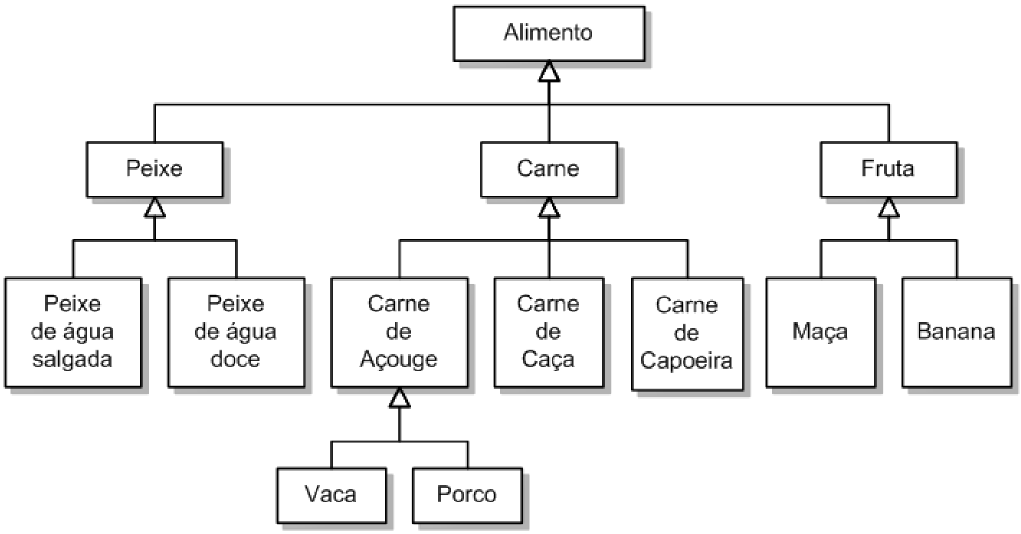

Figure 14: Excerpt of food conceptualization.

5.1.2 *Parte de alimento sólido* (solid food part)

Solid food is not usually used as a whole. All of the knowledge acquired about food parts is structured under the concept of *Parte de alimento sólido* (solid food part). The most specialized concepts of solid food parts are specializations of both a food part and a concept in the food hierarchy. For example, *Lombo de vaca* (beef sirloin) specializes both *Vaca* (beef) and *Lombo* (sirloin) – being sirloin a solid food part.

An excerpt of solid food part conceptualization can be seen in figure 15.

This classification complements the previous one, providing information about how a solid food item can be divided into parts.

5.1.3 *Estado de alimento* (food state)

Food state characterizes food in what concerns to its state. Four dimensions were considered:

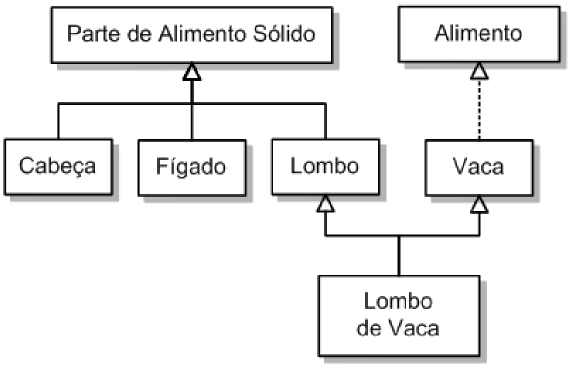

Figure 15: Excerpt of solid food part conceptualization.

1. Alimentary;
2. Physics;
3. Culinary;
4. Segmentation.

Estado alimentar (alimentary state) describes food before it suffers human action. According to this dimension food can be mature or green. *Estado físico* (physical state) is used to point out the physical state of a food item: solid, liquid or gaseous. *Estado culinário* (culinary state) is applied in the characterization of a food item after it suffered the effects of a culinary action: the food item can be boiled, roasted, fried, raw, etc *Estado de segmentação* (segmentation state) describes food after a segmentation process. It comprehends states like sliced, cut in half, opened, etc

Figure 16 shows an excerpt of the food state conceptualization.

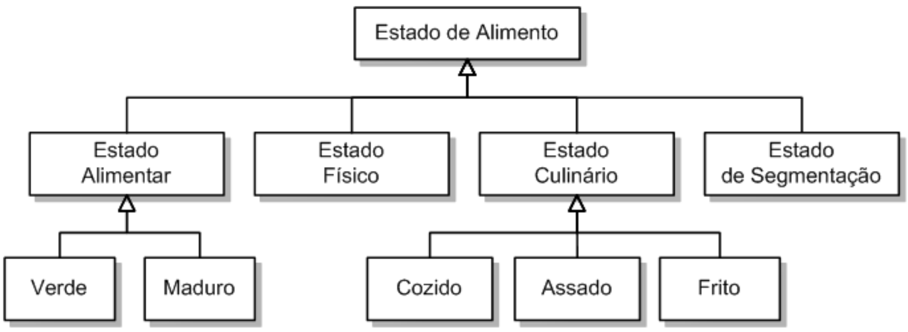

Figure 16: Excerpt of food state conceptualization.

5.1.4 *Alimento com estado* (food with state)

When cooking a meal, one acts over food items in a given state. Food with state is the conceptualization of that fact: it associates a food concept with a food state concept. Actions applied to food act over food items with state.

5.2 Kitchen utensils

The kitchen utensils module is a hierarchy of utensils that can be used in the kitchen. Besides all the instruments directly used when cooking, it also includes all the instruments or devices that might be used before or after cooking.

One of the competency questions, for which our ontology should give an answer, was to enumerate all the required utensils for a given recipe. This requirement made utensils one of the initial main modules, defined to be included in the ontology.

At the beginning of the knowledge acquisition process for the kitchen utensils module, a number of concepts were collected and structured in a basic hierarchy. Most of those concepts were suggested by people skilled on the area.

The utensils module uses the units and measures module in order to establish properties such as: power consumption, width, length and depth. Figure 17 focus this relation.

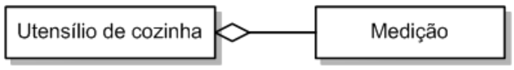

Figure 17: Utensils relation with other modules.

The hierarchy for kitchen utensils, which makes use of multiple inheritance, is mostly organized by the action performed with the instance. However, the first-level classes represent three big different groups of utensils plus the concept of “hand”, which could not be included in any of those groups. The figure 18 shows an excerpt of utensils conceptualization.

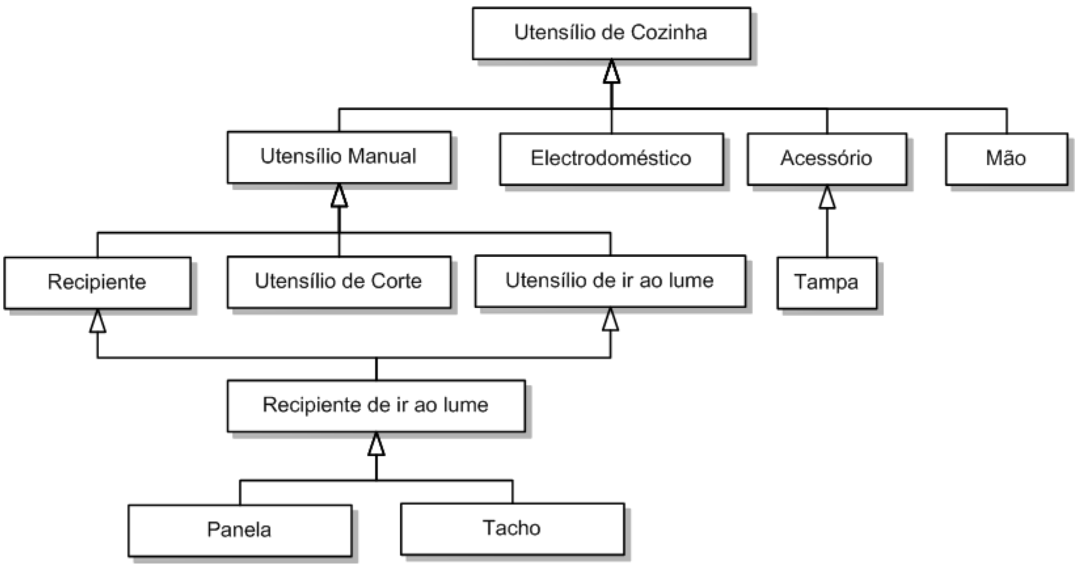

Figure 18: Excerpt of kitchen utensils conceptualization.

In a very first stage of the conceptualization process, the hierarchy was built based on several opinions from experts on the area. In a second stage, several books were used for refining our relations. Finally, we considered Terence and Couran [1982] as the main expert, because this book is one of the most complete reference on the subject.

Numerous kitchen utensils exist, and a difficult problem is to know when to stop. In the case of manual utensils, many specific and different utensils may be used in a kitchen, but the problem is even greater when considering electrical utensils, with all possible configurations and specifications. The criterion was based on three elements: a) most of the utensils named by a number of experienced people were included; b) the set of utensils described in the reference books; c) internet counts when searching for doubtful utensils.

5.3 Actions

The actions module has four main concepts to structure the information:

- *Acção* (action);
- *Acção primitiva* (simple action);
- *Acção ordenada* (ordered action);
- *Acção composta* (complex action).

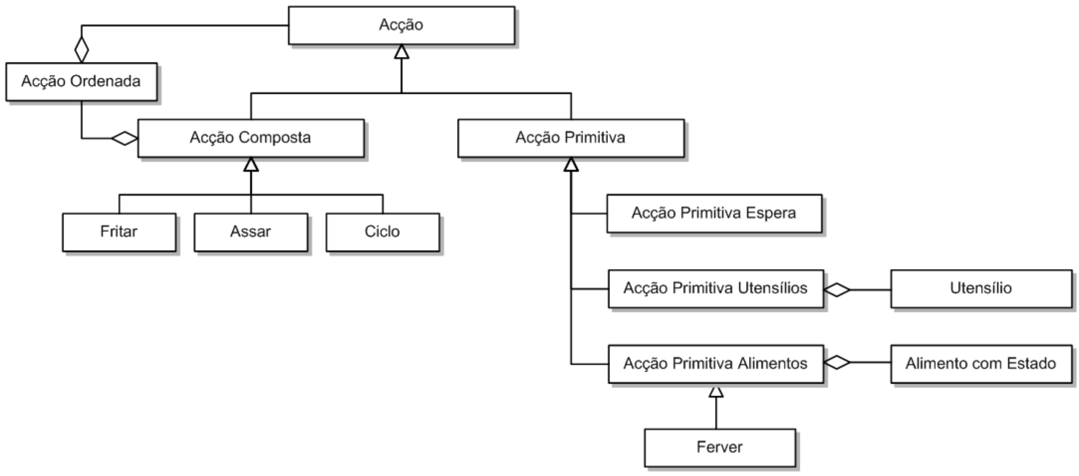

Figure 19: Actions concepts.

5.3.1 *Acção* (Action)

The *Acções* (Action) concept define operations executed on food and utensils. Recipes use *Acções* to describe the cooking process.

5.3.2 *Acção primitiva* (Simple Action)

Acções primitiva (Simple Actions) define the basic little things, typically, done when cooking, like *boiling*, *breaking*, *cutting*, or *peeling*. The ontology defines three major types of *Acções primitiva*: *Acções primitiva alimentos* (Food Simple Actions), *Acções primitiva utensílios* (Utensils Simple Actions), and *Acções primitiva espera* (Waiting Simple Actions).

Acções compostas (Complex Actions) and *Receitas* (Recipes) must use:

- *Acções primitiva alimentos* (Food Simple Actions) over food or food and utensils. Examples of *Acções primitiva alimentos*, are *putting* food inside a recipient, *cutting* food, *breaking* an egg, or *slicing* fruit.
- *Acções primitiva utensílios* (Utensils Simple Actions) over utensils. Examples of *Acções primitiva utensílios* are *switching on* a heat source, *covering* a recipient, or *putting* a recipience inside another recipient.
- *Acções primitiva espera* (Waiting Simple Actions) to express waiting times that depend, for example, on the food amount or the oven type. Examples of *Acções primitiva espera* are *boiling*, *frying*, or *roasting* waiting.

5.3.3 *Acção ordenada* (Numbered Action)

Acção ordenada (Numbered Action) define the sequence concept of *Acção*. *Acção ordenada* associate a number (integer) to a *Acção*. The process of each *Acção composta* is a sequence of *Acções ordenadas*. Figure 20, shows *Acção composta Fritar* (fry) as sequence of *Acções ordenadas*.

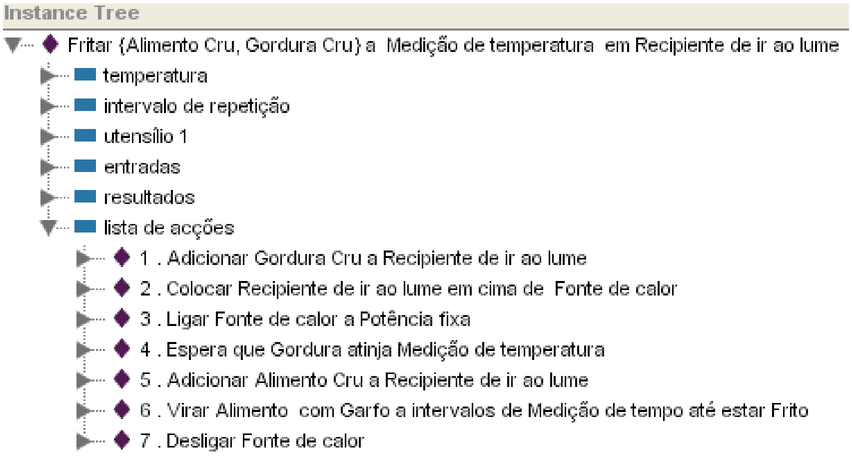

Figure 20: Acção composta Fritar.

5.3.4 *Acção composta* (Complex Action)

Acções compostas (Complex Actions) define complex tasks, typically, performed when cooking. *Receitas* use *Acções compostas* to describe the cooking process. *Acções compostas* use food and utensils and have a process that describes the *Acção composta*. The *Acção composta* process is a sequence of *Acções ordenadas*, as shown in figure 20. Example of *Acções compostas* are *roasting*, *frying*, and *stewing*.

5.4 Recipes

This is the main concept of a cooking ontology: a recipe will relate all other main concepts at a higher level in order to describe a cooking procedure. As shown on figure 21, a recipe takes some time to be done, produces an amount of food (both Meauseres); uses some Utensils; transforms some Ingredients into others; is slited into 3 Phases: preparation, cooking and presentation. All these is described in the next section.

5.4.1 *Receita* (recipe)

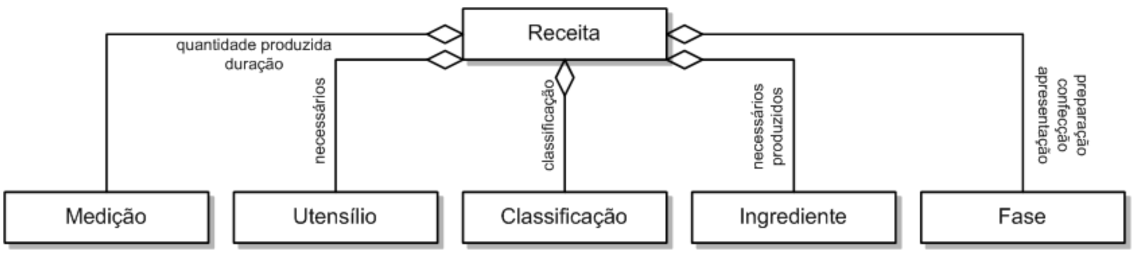

Figure 21: Main recipe concept and it's relations with other concepts.

A Recipe represents the algorithm that is found in a cooking book. It is described with nine slots (attributes):

Duração (duration): a Measure of the time needed to execute all the steps of the procedure;

Quantidade (quantity): a Measure of the ammount of food that is produced (for example, "for 1 person");

Utensílios (Utensils): a set of Utensils needed to execute the tasks of the recipe;

Classificação (Classification): information about the Recipe that can be used, for instance, as a search criteria;

Ingredientes necessários (needed Ingredients): a set of Ingredients taken as input for the Recipe;

Ingredientes produzidos (produced Ingredients): a set of Igredients that are produces as output of the Recipe. This is how the output of a Recipe is refered;

Preparação (preparation): Phase prior to the execution of a Recipe. Describes some steps needed before it is possile to do the cooking. This is needed because some Recipes needs some steps to be done, for example, on the day before;

Confecção (cooking): the main Phase of the Recipe;

Apresentação (presentation): Phase describing the steps needed to present the dish after it has been cooked.

These concepts and their implementations are described in the next sections.

5.4.2 *Classificação* (classification)

Each recipe is classified according to some criteria. To join them all we created a concept for the classification. As shown in figure 22, the recipe's classification considers:

Tipo de Prato (kind of dish): classification according to the common division of dishes: appetizers, meat, fish, salads, etc. This corresponds to an auxiliary module (c.f 5.5.1).

Origem (origin): geographical origin of the meal. Is the name of the country, region or city where the meal was created.

Custo (cost): how much will we spent to buy the ingredients for this meal? This is relative to the country where the application will be used. Discrete values: "cheap", "medium", "expensive".

Dificuldade (difficulty): how much expertise will you need to perform all the needed tasks? Discrete values: "easy", "medium", "difficult".

Calorias (calories): the energetic nutrition value of the meal. Discrete values: "light", "medium", "fat".

Picante (hot): whether the meal is spicy. "not hot', "medium", "spicy".

Alcoólico (alcoholic): is this meal alcoholic? Boolean value.

Vegetarino (vegetarian): whether this meal can be eaten by vegetarians. Boolean value. No differences between types of vegetarians were taken into consideration.

Época (season): is this meal for a proper time of the year? Discrete values: "Spring", "Summer", "Autumn", "Winter", "Christmas", "Easter", etc.

5.4.3 *Ingrediente* (Ingredient)

In any recipe, it is always possible to find a list of ingredients. The Ingredient concept (see figure 23) consists in certain quantity of a food item in a given state. That is, an instance of Ingredient will have an instance of Measure to fill its quantity attribute and an instance of Food with State, to which the measurement refers, that corresponds to the Food item and Food State that characterize the Ingredient. Ingredients are used by Recipe, Phase and Task to state the amount needed for each and the resulting products.

Figure 22: Recipe's classification.

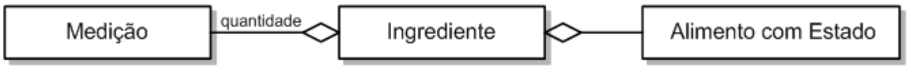

Figure 23: Ingredient concept definition.

5.4.4 *Tarefa* (Task) and *Fase* (Phase)

As said before, a Phase joins a list of Tasks representing the steps needed to cook a given dish. A *Tarefa* (Task) corresponds to that concept relating the Action to be performed, the needed Ingredients, a Measure of the time needed and the produced Ingredients. A restriction is applied between the Food used by the Action of the Task and the Food stated by the Ingredient: the Ingredients of a Task are quatities of the Food items refered by the Action of that Task.

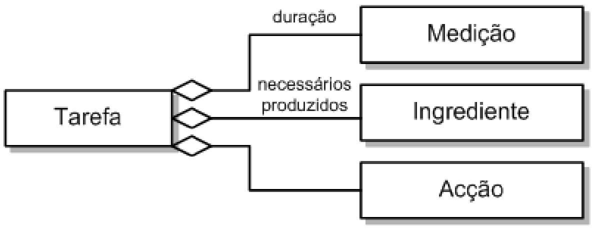

Figure 24: Task concept.

As the Tasks should be performed in a certain order, the concept of *Tarefa Ordenada* (Ordered Task) was introduced. It represents a task executed in a specific order (given by an order number). This decision is due to the lack of the list concept in the majority of the Ontology Languages and Tools as it would increase the complexity.

In order to represent similar Recipes without repeating all that they have in common the concept of *Tarefa Opcional* (Optional Task) was included. It represents a Task that can be skipped. For example, scrambling some spices into an egg before frying would be an Optional Task.

A set of Ordered Tasks represents a procedure that describe a Phase. A Phase collects all the Tasks needed to prepare, cook and present a meal, as described in the recipe. As presented in figure 27 it aggregates the needed Ingredients, the Measure of its duration and a list of Tasks (Ordered and/or Optional).

5.5 Auxiliary Modules

To represent extra information useful for the ontology and to ease reuse, some auxiliary modules were built: one for Kind of Dish; one for Units and Measures (used by Ingredients and Utensils);

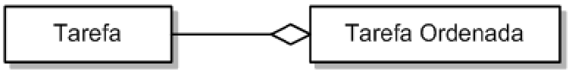

Figure 25: Ordered task concept.

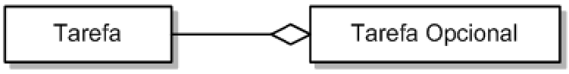

Figure 26: Optional task concept.

and other for Equivalences.

5.5.1 *Tipo de Prato* (Kind of Dish)

The *Tipo de Prato* (Kind of Dish) represents the moment of the meal when the described dish is suposed to be eaten. This is used to organize the recipes. The hierarchy of this is quite known as it corresponds to the criteria to organize the recipes in a great number of books. Being so, it was decided to represent it in a separated small module.

5.5.2 *Unidades* (units) and *Medidas* (measures)

The need of an measure module arose when defining the ingredients.

In the literature the units are often arranged according to the units system they belong to and according to what they measure (for instance, length). When using the system unit criteria all the measures ended mixed. When selecting an unit according to what was measured lots of different units for the some measure were found without indicating which was the standard. To solve this, a second hierachy was created. Units are the leaves of two conceptual trees and each one is connected to both. Figure 28 shows the two criterias and an unit, *Metro* (meter), as an example.

A Measure is an association between a Unit and a Value: a fixed one; or one in a range. This lead to the creation of two types of measures: *Medição Fixa* (Fixed Measure) and *Medição Delimitada* (Delimited Measure).

When measuring something the units that can be used are restricted to a set. Figure 30 shows how Units are related to Measures. An hierarchy similar to that of the Units exist in the Measures to represent the existing restrictions. With this, when measuring some *Comprimento* (length) only the correspondent units can be used.

5.5.3 *Equivalência* (equivalence)

Different ways to indicate the quantity of food needed for a recipe were found. Instead of converting them all to a stadard Unit when inserting the recipes, it was decided to add knowledge about them that would allow to infer the equivalences. This information can be very useful in the cooking context: for instance, when doing a cake without a scale, a coup or a spoon can be used to measure the needed flour.

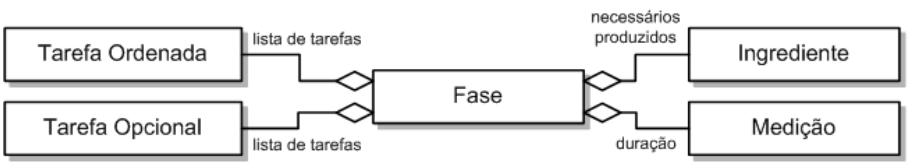

Figure 27: Phase concept.

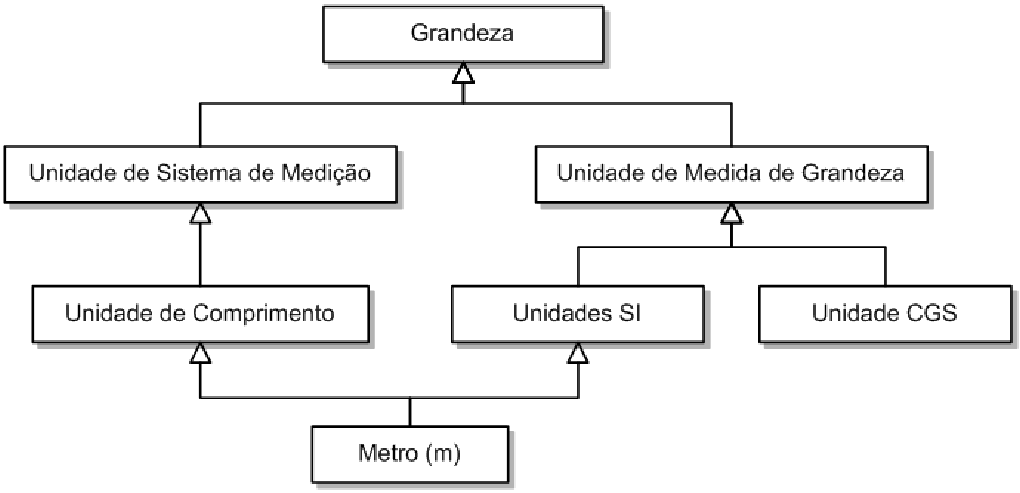

Figure 28: Unit concept.

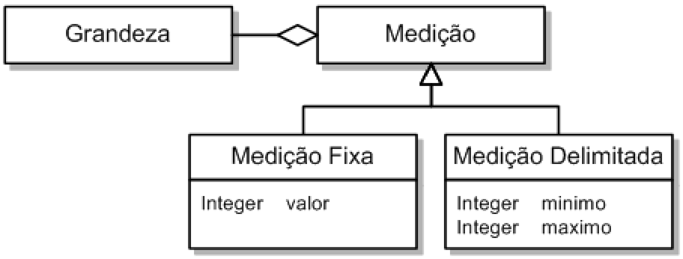

Figure 29: Measure concept.

This knowledge about conversions between different units was added using the concept equivalence that makes an association between to measures, stating they are equals (or much similar). A restriction should be verified: only related Units can be stated as equals. Only two units that have the same parent (what is measured) in the branch of the units tree organized can have equivalence relation.

Two exceptions are allowed: if a Food or an Utensil is associated to the Equivalence. In the first case, the equivalence uses the food density. In the second informal measures like "small" are associated with a numerical range to measure it.

This module also has the prefixes and suffixes used to name the different values of the units, such as "deca" for 10 times or "kilo" for 1000 times.

6 Future Work

Apart from sharing, reusing, maintaining and evolving, which are important issues for the future, the following subsections describe in detail the work to be performed in the scope of each module.

6.1 Food

In a module like this one, there is always room for improvement. At least, adding new food concepts is certainly a possibility.

Other, more interesting, possibility would be to add new classifications based on different criteria. It is not easy to classify all existing food concepts, but some work was done in this field [Villarías, 2004, Harvard School of Public Health, 2005] and focusing on different features it is

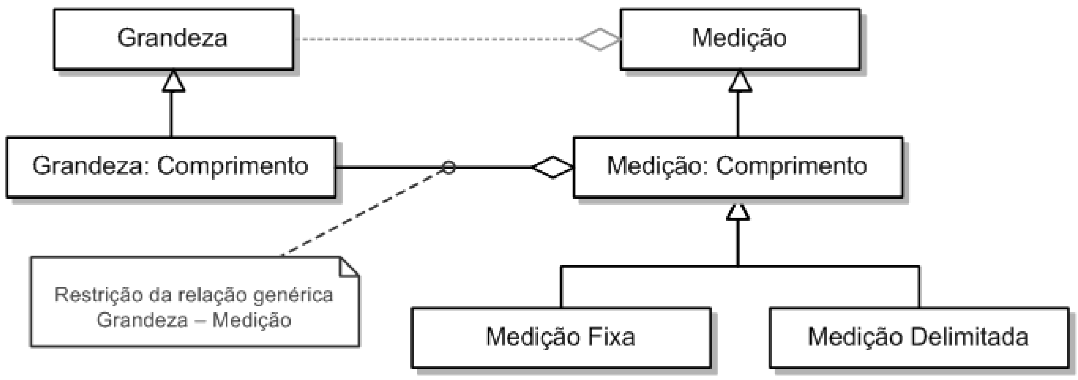

Figure 30: Relation between units and measures.

possible to complement the current conceptualization with new characterizations. For example, a flavor based classification or one based on the nutrition pyramid would increment the information level about food items.

6.2 Kitchen Utensils

Although most common kitchen utensils were considered in our ontology, there is always room for adding new concepts. Besides any possible missing concept, new inventions and developments are always contributing to the creating of new tools, brought to the market everyday. Concerning the utensils module, the maintenance of the information is always an important aspect to be taken into consideration in the future.

Additional information could also be appended to each concept, either in the scope of the documentation, or in what concerns to the set of defined properties.

6.3 Actions

The actions module can be completed. We just inserted the actions we needed to define a small number of recipes. So, a task that can be done to improve is to complete the actions module from the books.

Another improvement could be to integrate the process ontology to define the complex actions and the recipes. These would cause restructuring of the actions module and may not be worthwhile.

6.4 Recipes

This module is never finished. More and more recipes can be added. As the recipes will be stored (in the final application) in a database, only the relevant cases were introduced in the ontology as instances. The idea was to show how the relevant cases could be described.

In the future, when using the ontology in a Dialogue System the application could include a personalized configuration to specify the real utensils that the user has at home as Utensils instances. In that case, when referring to the objects the system could even refer the drawer were it is stored.

Kind of dish could be extended to take some cultural differences into consideration. For example, *Pasta* is eaten before the main dish (as an appetizer) by Italians while Portuguese people eats it as a main dish (as companion for the meat or fish).

The origin of plates could be connected to some Geographical Ontology in order to allow inferece on geographical proximity, for example.

Also the Season of the year when a meal is more adequate could lead to a new module to be used replacing the current discrete values.

6.5 Auxiliary Modules

The current auxiliary modules could be extended. In Units and Measures module all the International System of Units (SI) are represented. Even those not related to cooking. More units systems can be added. For example the old English units system, the CGS (Centimeter-Gram-Second system).

Equivalences between measures commonly used by cookers (such as "cup" or "spoon") are equivalences between volume and weight. These measures depend on the density of the used food. Some were added as example, more could be added.

6.6 Inference

The Ingredients used (or produced) by a Recipe can be deduced by the ones used by its Phases. The same can be said between the Ingredients needed (or produced) by a Phase and its Tasks. Currently, the Ingredientson are descriped in slots and are explicitly declared refering the same instances when stating the same Igredients. Nevertheless, it could be discovered in an automatic way by stating the needed restrictions and letting Protégé do that repetitive job.

If more information was added, for instance, about the usage of some dishes with others, help could be provided to the user to choose the best combination of dishes for a meal. If beverage was added, this help could include drinks.

Equivalences between volumes and weights could be inferred if the Food description was enriched with density information.

6.7 Integration in a dialogue system

As the motivation for this project was extending an existing dialogue system [Mourão et al., 2003, Mourão, 2005, das Comunicações, 2003], the next step is the use of the resulting ontology to enrich it.

The current version of the ontology is useful and usable in that context. Some preliminar tests have been made to allow the autonomous agent, a butler named "Ambrósio" to start helping in kitchen tasks as soon as possible. The current version of the system takes a list of recipes and reads them to the user.

The instances of Recipe that should be in the database are not coded yet. That task will request some extra work before the integration of the ontology in the dialogue system is possible.

When performing the merge, some methodological knowledge will be gathered to identify the main steps when joining Dialogue Systems with Ontologies. This work will be part of a PhD thesis project.

The existing dialogue system has an architecture similar to that used by TRIPS [Allen et al., 2005] (figure 31). It will be interesting to explore how the knowledge stored in an ontology that can be used automatically and dinamically by a dialogue system. For example, the words that name the concepts can be used to collect the related vocabulary.

If this method proves to be efficient, the enrichment of a dialogue system with a new domain can be reduced to the construction of an ontology. Here the ontology will gather all the knowledge that currently has to be spread by the modules of the system. This can be an advantage as all the knowledge information will be concentrated in the same module, the ontology. Also, because, presently, is already necessary to collect the knowledge when integrating a new domain. Here instead of splitting that knowledge into the relevant modules, it is isolated, which turns it pluggable (plug-and-play). Some initial work has been done by Milward and Beveridge [2003] and Flycht-Eriksson [2004]. Both proved that the integration of ontologies can bring several advantages to dialog systems.

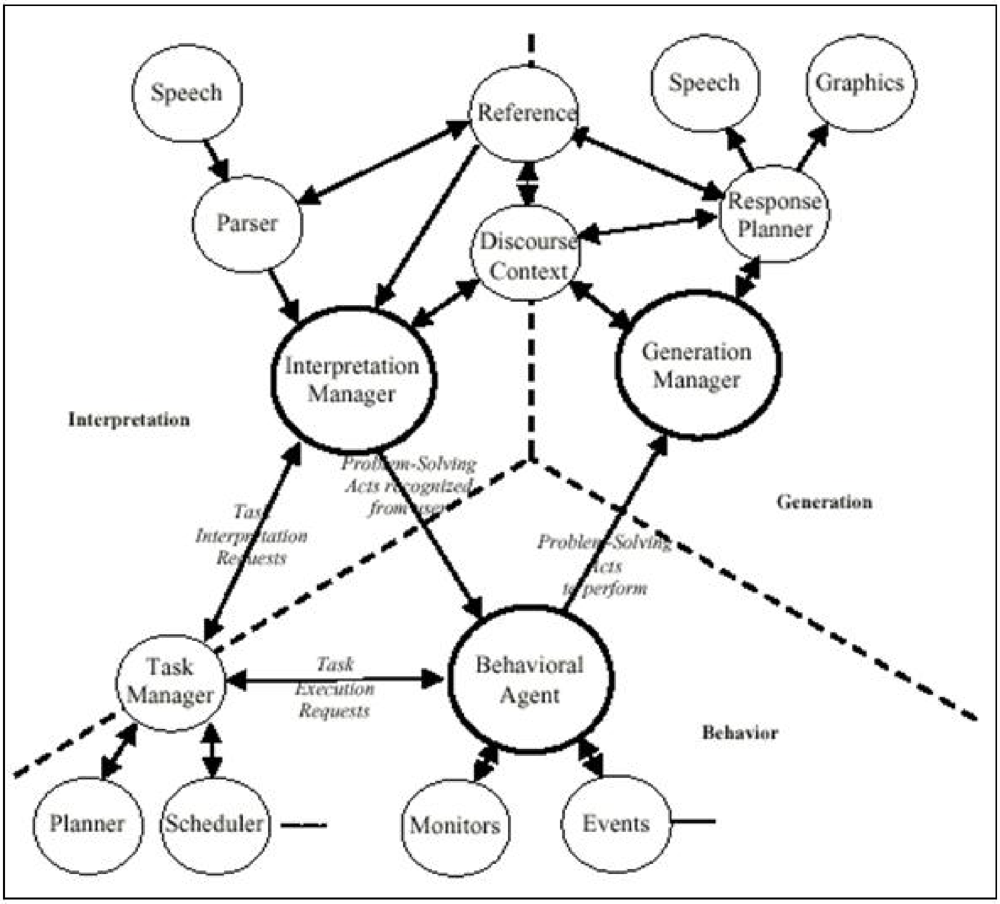

Figure 31: TRIPS architecture.

7 Conclusions

The aim for this work consisted on developing an ontology on the cooking domain, in order to be integrated in a dialog system. The resulting ontology covers four main areas of the domain knowledge: food, kitchen utensils, actions and recipes. Food, utensils and actions areas of knowledge are formalized as class hierarchies with instances (in what concerns actions), covering in a considerable extent – at least, accordingly to the used information sources – the target domain. Recipes concepts interconnect concepts from all the other areas, in order to define an adequate model of the cooking domain. Two instances of Recipe were created to demonstrate the usability of the developed specification.

The ontology building process was strongly influenced by the methodology proposed by López et al. [1999] and the phases of specification, knowledge acquisition, conceptualization, implementation and evaluation were essential to achieve the intended result. The sources of information consisted mainly of books and the internet, and informal and formal text analysis techniques, as well as brainstorming and cross-validation sessions, were used in the knowledge acquisition phase. Conceptualization, essentially, relied on the identification of concepts and groups of concepts and in building classification trees. The knowledge model was then formalized using Protégé, which can also be used to automatically generate the ontology code. Other relevant issue was integration: the built ontology is composed by four main modules – food, kitchen utensils, actions, and recipes – and two auxiliary modules – units and measures, and equivalencies. Although not using a standard evaluation methodology, all modules were cross-validated in several meetings and informal competency questions were used to check if the ontology satisfied the client needs.

The following lists the problems that we considered important:

- The cooking domain was too broad for the available time;
- Small team interaction difficulties;
- Loss of data caused by bugs in the IDE;
- Collaborative implementation difficulties.

Despite the problems found, the ontology reached a usable state. All concepts were structured and well documented. The built ontology is not currently used by the dialogue system, since that work will be part of a PhD thesis project.

References

Biblioteca Activa. *Doces Conventuais.* Edimprensa Editora, Lda., 1st edition, 2004.

James Allen, George Ferguson, Mary Swift, Amanda Stent, Scott Stoness, Lucian Galescu, Nathanael Chambers, Ellen Campana, and Gregory Aist. Two diverse systems built using generic components for spoken dialogue (recent progress on trips). In Ann Arbor, editor, *Proceedings of the Interactive Poster and Demonstration Sessions at the 43rd Annual Meeting of the Association for Computational Linguistics (ACL-2005)*, pages 85–88, Michigan, USA, 2005. Ver: `http://www.cs.rochester.edu/research/cisd/projects/trips`.

Allrecipes.com. `http://allrecipes.com`, n.a.

Arte Digital Design e Publicidade. Roteiro gastronómico e portugal, 1997. URL `\url{http://www.gastronomias.com/}`.

Françoise Bernard. *Super cocotte.* ILTE Torino, 1st edition, 1964.

CondéNet. `http://www.epicurious.com/`, n.a.

Fundação Portuguesa das Comunicações. Casa do futuro, 2003. `http://www.fpc.pt`, `http://www.casadofuturo.org`.

Associação Guias de Portugal. *O grande livro ilustrado da Culinária.* 1st edition, 1987.

Jorge Brum do Canto. *O Livro do Pantagruel.* Temas e Debates, 43 edition, 1997.

FAO. AGROVOC. `http://www.fao.org/agrovoc/`, 2004.

Mariano Fernandez, Asuncion Gomez-Perez, and Natalia Juristo. METHONTOLOGY: from Ontological Art towards Ontological Engineering. In *Proceedings of the AAAI97 Spring Symposium Series on Ontological Engineering*, pages 33–40, Stanford, USA, March 1997.

Annika Flycht-Eriksson. *Design and Use of Ontologies in Information-providing Dialogue Systems.* PhD thesis, School of Engineering at Linköping University, 2004.

Martin Fowler. *UML Distilled: A Brief Guide to the Standard Object Modeling Language.* Addison-Wesley Professional, 3rd edition, 2003. ISBN 0321193687.

General Conference on Weights and Measures. The International System of Units, 7th edition. `http://www.bipm.org/utils/en/pdf/si-brochure.pdf`, 1998. General Conference on Weights and Measures (Conférence Générale des Poids et Mesures, CGPM), the International Committee for Weights and Measures (Comité International des Poids et Mesures, CIPM) and the International Bureau of Weights and Measures (Bureau International des Poids et Mesures, BIPM).

General Conference on Weights and Measures. Supplement 2000: addenda and corrigenda to the 7th edition of The International System of Units. `http://www.bipm.org/utils/common/pdf/si-supplement2000.pdf`, 2000. General Conference on Weights and Measures (Conférence Générale des Poids et Mesures, CGPM), the International Committee for Weights and Measures (Comité International des Poids et Mesures, CIPM) and the International Bureau of Weights and Measures (Bureau International des Poids et Mesures, BIPM).

John H. Gennari, Mark A. Musen, Ray W. Fergerson, William E. Grosso, Monica Crubézy, Henrik Eriksson, Natalya F. Noy, and Samson W. Tu. The evolution of protégé: an environment for knowledge-based systems development. *International Journal Human-Computer Studies*, 58(1): 89–123, 2003. ISSN 1071-5819.

Google. `http://www.google.com`, 1995.

João Graça, Márcio Mourão, Orlando Anunciação, Pedro Monteiro, H. Sofia Pinto, and Virgílio Loureiro. Ontology building process: the wine domain. In *Proceedings of the 5th Conference of the European Federation for Information Technology in Agriculture, Food and Environmen*, Vila Real, Portugal, 2005.

Thomas R. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, 1993. ISSN 1042-8143.

M. Grüninger and M. Fox. Methodology for the design and evaluation of ontologies. In *IJCAI'95, Workshop on Basic Ontological Issues in Knowledge Sharing, April 13, 1995*, 1995. URL `\url{http://citeseer.ist.psu.edu/grninger95methodology.html}`.

Nicola Guarino and Christopher Welty. An overview of ontoclean. In *Handbook on Ontologies*, pages 151–159, Germany, 2004. Springer Verlag.

Harvard School of Public Health. Food Pyramids. `http://www.hsph.harvard.edu/nutritionsource/pyramids.html`, 2005.

Mariano Fernández López, Asunción Gómez-Pérez, Juan Pazos Sierra, and Alejandro Pazos Sierra. Building a chemical ontology using methontology and the ontology design environment. *IEEE Intelligent Systems*, 14(1):37–46, 1999. ISSN 1094-7167.

Deborah L. McGuinness. Ontologies come of age. *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*, 2003.

David Milward and Martin Beveridge. Ontology-based dialogue systems. In *3rd Workshop on Knowledge and Reasoning in Practical Dialogue Systems – 18th International Joint Conference on Artificial Intelligence (IJCAI03)*, 2003.

Márcio Mourão, Pedro Madeira, and Nuno Mamede. Interpretations and discourse obligations in a dialog system. In Nuno J. Mamede, Jorge Baptista, Isabel Trancoso, and Maria das Gracas Volpe Nunes, editors, *Computational Processing of the Portuguese Language – Proc. of the 6th Intl. Workshop, PROPOR 2003*, number 2721 in Lecture Notes in Artificial Inteligence, pages 197–200, Faro, Portugal, June 26–27 2003. Springer-Verlag, Heidelberg. ISBN 3-540-40436-8.

Márcio Duarte Albasini Mourão. Gestão e representação de domínios em sistemas de diálogo. Master's thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa, Portugal, 2005.

Nestscape Communications Corporation. DMOZ – Open Directory Project. `http://dmoz.org`, n.a.

Natalya F. Noy and Deborah L. McGuinness. Ontology development 101: A guide to creating your first ontology. Technical Report KSL-01-05/SMI-2001-0880, Stanford Knowledge Systems Laboratory/Stanford Medical Informatics, 2001.

Oregon State University. Food Resource. http://food.oregonstate.edu/, n.a.

H. Sofia Pinto and João Pavão Martins. Ontologies: How can they be built? *Knowl. Inf. Syst.*, 6 (4):441–464, 2004.

York Sure, Michael Erdmann, Juergen Angele, Steffen Staab, Rudi Studer, and Dirk Wenke. OntoEdit: Collaborative ontology development for the semantic web. In *Proceedings of the first International Semantic Web Conference 2002, ISWC 2002*, Sardinia, Italia, June 9-12 2002. Springer, LNCS 2342. URL \url{http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/2002_iswc_on%toedit.pdf}.

Terence and Caroline Couran. *O grande livro ilustrado da Culinária*. Circulo de Leitores, 1st edition, 1982.

United Stated Department of Agriculture. USDA National Nutrient Database for Standard Reference. http://www.nal.usda.gov/fnic/foodcomp/Data/, 2005.

M. Uschold, M. King, S. Moralee, and Y. Zorgios. The enterprise ontology, 1995. URL \url{http://citeseer.ist.psu.edu/uschold95enterprise.html}.

Leticia Gutiérrez Villarías. Ontology-based semantic querying of the web with respect to food recipes. Master's thesis, Technical University of Denmark, 2004.

Wikipedia. SI units. http://en.wikipedia.org/wiki/SI_units, 2001a.

Wikipedia. Sistema Internacional de Unidades. http://pt.wikipedia.org/wiki/Sistema_Internacional_de_Unidades, 2001b.

Wikipedia. Wikipedia, The Free Encyclopedia. http://www.wikipedia.org/, 2001c.