2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology

A cross-language approach for semantic recipe annotation and nutrition annotation augmentation based on extraction ontologies and culture-relevant policy reasoning

Zheng Meng Long and Yuri A. Tijerino
Department of Applied Informatics
Kwansei Gakuin University
Kobe-Sanda, Hyogo Prefecture, Japan
coolfighter@gmail.com

***Abstract*—This paper summarizes an approach for cross-language annotation and querying of semi-structured data with the use of cross-language extraction ontologies. The paper describes the capabilities of the approach in the context of annotation of nutritional components and their respective values for ingredients found in English, Chinese and Japanese recipes. To this end, a cross-language, data-extraction ontology is designed to annotate the recipes, and then augments the annotations with nutrition information on a per-ingredient basis using nutritional information obtained from nutritional databases found in China, Japan and the USA. The relevant nutritional information is used to augment ingredient annotations, which can be used to respond to specific queries related to nutritional goals or to meet specific dietary regimens. The approach meets cross-cultural constraints by incorporating a cultural policy service subsystem to enable responses based on cultural policy ontologies. The approach allows cross-language queries by annotating recipes and parsing queries with a cross-language extraction ontology. For example, responses to an English or Chinese query might be obtained from a Japanese recipe and vice versa.**

***Keywords*-Ontology engineering; cross-linguistic ontology; cross-cultural ontology; nutrion; recipes**

I. Introduction

As globalization continues at an increased pace and markets become even more interdependent across the globe, cross-lingual and cross-cultural interactions have become essential in our daily lives. Specially, as the Web crosses geopolitical barriers, and narrow the language and cultural gaps, it is paramount to design efficient approaches to facilitate cross-language and cross-cultural information queries. The Web, by design, was not meant to serve a single culture or language and despite the attempt to limit the access to resources across country boundaries either for political or business reasons, the fact is that the Web is one of the biggest driving factors of globalization. Although much is being done through technology to help alleviate cross-language interactions via readilly-available, handy tools such as Google Translate and Bing Translator, these tools only go so far and suffer from the same, if not worse, shortcomings of keyword-based search. As a matter of fact, in mission-critical situations such translation tools do not produce acceptable results. Thus semantic-based approaches are much needed to answer more complex queries. This paper describes the design of one such approach.

Some of the questions explored int this paper include the following. Is it possible to perform multi-domain reasoning across semantically annotated cross-language resources? What sort of cross-domain, cross-language and cross-cultural challenges exist and how can they be tackled? We have attempted to respond to some of these questions and their implications while designing a cross-cultural and cross-language recipe querying framework that augments semantic annotations of Chinese, English and Japanese recipe websites with information found on nutrition databases available on those languages and then applies cultural contraints through a policy service subsystem to answer queries.

II. Relationship of food and culture

Food and culture go hand in hand. The study of a particular culture is not complete if the study of the history and origin of its traditional foods is not methodically performed. The literature is full of manuscripts devoted to the study of culture through food. This section illustrates various aspects of how food is associated to culture, through selected contemporary literary works that inspired the authors of this paper to choose the narrow, but culturally-rich domains of recipes and nutrition to design the approach described herewith.

A. Symbolism of food in culture

Throughout the ages, man has associated cultural symbolism with food. Perhaps one of the most significant differentiating factors between humans and other species presents itself in the form of food. According to Montari [1], food becomes culture *when it is produced, prepared and eaten* because of the technology used to produce it, the practices used to prepare it and the socio-cultural symbolism represented in the choices we make as to what, when and how we eat. Though not always apparent, we must pay special attention to the symbolism of food, its ingredients and its preparation.

978-0-7695-4880-7/12 $26.00 © 2012 IEEE
DOI 10.1109/WI-IAT.2012.264

B. Health, nutrition and traditions surrounding food

Since 1826, when Brillat-Savarin coined the well-known phrase: *we are what we eat* [2], modern civilization has known that there exists a strong relationship between the nutritional value of food and the state of our health. For example, Kittler, Sucher and Nelms [3], propose that every registered dietician in the USA should have multicultural competence in order to perform their particular nutrition specialty efficiently and effectively. They devote their book, *Food and Culture* [3], almost in its entirety to help nutrion experts understand the cultural aspects of 1) food selection, preparation, and storage; 2) eating patterns including traditions and religion; 3) nutrition-related health problems associated with various cultures; and 4) the cultural values, health beliefs and nutrition practices of various cultures in the USA. It is important to consider these cultural characteristics, beliefs and tradition when annotating recipes and responding to cross-language user queries.

C. Cultural and religious acceptability of what, when, where and how we eat and how food is prepared

It is not uncommon to discover that not all foods eaten in some cultures are acceptable, desirable or even allowed to ingest in others. This is one of the most important aspects discussed by Ashley et al. in their book, *Food and Cultural Studies* [4], in which they explore various aspects of what, when, where and how it is or is not acceptable to eat and how food is prepared across various cultures. What makes their work of relevant to this paper, is how culture refers not only to place of origin, but also social status. For example, certain foods are culturally acceptable to eat at a bench at a park BBQ using our hands, while one must use utensils when seating at a formal dinner. They also point out other interesting aspects of food and culture, such as table manners, which are not as relevant to this paper, but are worth keeping in mind as they indicate food and culture are intricately interrelated. However, some of the other aspects need to be considered when designing a cross-culture and cross-language recipe recommendation system, *vis-à-vis* acceptability of foods in responding to cross-language queries.

III. Cross-language semantic annotation of recipes augmented with nutrition information

The Merriam-Webster dictionary defines recipe as: *a set of instructions for making something from various ingredients* [5]. The purpose of cross-language semantic annotations is to construct replies to queries with information that might or not be in the original query language. By augmenting semantic annotations with additional nutrition information it is possible to construct replies to more complex queries. For example, to queries regarding specific dietary constraints. The architecture depicted in Figure 1 illustrates a high level design view of the approach. The general flow of the approach is as follows:

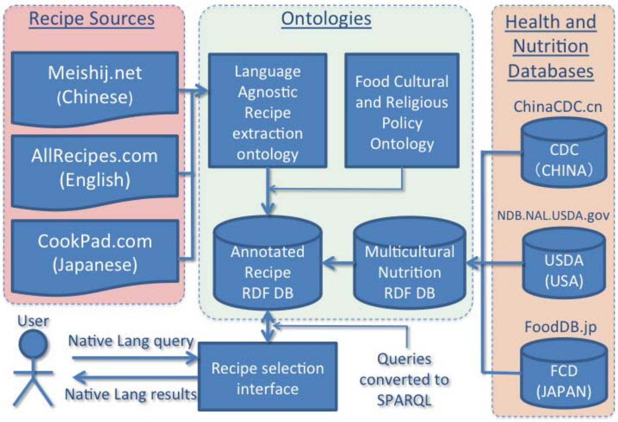


Figure 1. Experimental cross-domain, cross-culture and cross-language recipe extraction, annotation and querying system.

1) Recipes are extracted and annotated from various language-specific resources. However, only ingredients, and not the instructions are the current scope of the approach.
2) A language-agnostic recipe extraction ontology is used to analyze the ingredients from the linguistic perspective.
3) Ingredients are annotated with nutritional information extracted from various nutrition databases, giving preference to the database in the language of the recipe.
4) An annotated recipe database, in RDF format, is generated from the various annotation sources.
5) User native language queries are processed with the language-agnostic ontology and converted to language-neutral SPARQL queries, to query the annotated recipe database.
6) Depending on the ingredients and their nutrients, a cultural and religious policy service subsystem is used to analyze the ingredients with known cultural and religious information.
7) Using the language-agnostic ontology, the results are then converted to the original language of the query.

A. Semantic recipe annotation across multiple languages

The language-agnostic, recipe extraction ontology is a direct application of earlier work on language-agnostic, data extraction ontologies [6], which is defined as follows:

A *cross-language ontology* localized to n contexts $\{\mathcal{C}_1, ..., \mathcal{C}_n\}$ is an n+1-tuple $\mathcal{O} = (\mathcal{I}, \mathcal{L}_1, ..., \mathcal{L}_n)$, where $\mathcal{I}$ is an interlingua representing concepts and facts in the ontology from a language-agnostic perspective, and each $\mathcal{L}_i$, $1 \le i \le n$, is a localization of $\mathcal{I}$ to one of the n contexts. The interlingua

$\mathcal{I}$ is an extraction ontology that consists of a set of structural concepts (e.g., object sets, relationship sets, data frames) and facts (e.g., objects, relationships) that describe a domain of interest in a language-agnostic way. Each localization is a 4-tuple $\mathcal{L}_i$ = ($\mathcal{C}_i$, $\mathcal{O}_i$, $\mathcal{M}_{\mathcal{I}\mapsto\mathcal{L}_i}$, $\mathcal{M}_{\mathcal{L}_i\mapsto\mathcal{I}}$), where $\mathcal{C}_i$ is a local context label, $\mathcal{O}_i$ is an extraction ontology, $\mathcal{M}_{\mathcal{I}\mapsto\mathcal{L}_i}$ is a set of mappings from $\mathcal{I}$ to $\mathcal{L}_i$, and $\mathcal{M}_{\mathcal{L}_i\mapsto\mathcal{I}}$ is a set of mappings from $\mathcal{L}_i$ to $\mathcal{I}$. Each concept in $\mathcal{O}_i$ must map to a single concept in $\mathcal{O}$, but concepts in $\mathcal{O}$ may map only partially to concepts in $\mathcal{O}_i$.

The key idea of our multilingual ontology architecture is that each localized ontology maps to a central interlingua and vice versa. This "star architecture" avoids the n^2 complexity of mapping each localized ontology to all other localizations, and instead provides a nearly linear scaling. Adding another localization involves constructing the localized extraction ontology ($\mathcal{O}_i$) along with mappings ($\mathcal{M}_{\mathcal{L}_i\mapsto\mathcal{I}}$ and $\mathcal{M}_{\mathcal{I}\mapsto\mathcal{L}_i}$) to and from the interlingua. In the process, it may be necessary to adjust the interlingua so that all concepts in $\mathcal{O}_i$ are represented directly in $\mathcal{I}$, and this may in turn require adjusting some of the mappings for other localizations. But since most mappings are trivial, the expected case is a linear effort required to add an additional localization to $\mathcal{O}$.

It is customary to identify language and culture contexts by spoken language and country name such as Mandarin/China (zh_CN) or English/USA (en_US). But in general there could be many contexts associated with a given language/country pair, such as Mandarin/Taiwan in contrast to Cantonese/China, or even tourist Spanish/Mexico versus business Spanish/Mexico. The concepts chosen for a particular localization may vary for many reasons. Ultimately, the precise meaning of "context" is defined by the author of the localization who expresses a selected set of ideas in a particular language. In our definition, we only need to note that a context has a chosen label, $\mathcal{C}_i$ (though conventional locale labels such as "en_US" or "zh_CN" could easily be used where appropriate). As a convention, we may replace i with $\mathcal{C}_i$ when referring to elements of $\mathcal{O}$. For example, the English/U.S. localization $\mathcal{L}_i$ could be designated $\mathcal{L}_{en\_US}$ = ("en_US", $\mathcal{O}_{en\_US}$,$\mathcal{M}_{\mathcal{I}\mapsto\mathcal{L}_{en\_US}}$, $\mathcal{M}_{\mathcal{L}_{en\_US}\mapsto\mathcal{I}}$).

Figures 2 and 3, illustrate the language agnostic ontology from the localizations $\mathcal{C}_i$ for $\mathcal{L}_{zh\_CN}$ (Mandarin Chinese, China), and $\mathcal{L}_{en\_US}$. Notice that only the ingredients, their amounts and units are annotated. An annotation at this point consists of the following language-agnostic components, surrounding the ingredients, the amounts and the units used.

1) a canonical label of the ingredient,
2) a canonical label for the amount,
3) a canonical label for the units used, and
4) canonical labels to qualify ingredient other descriptive features of the ingredient. In Figure 3 the terms "all-purpose"', "brown", "beaten", "overripe", and "mashed", are examples of descriptive features.

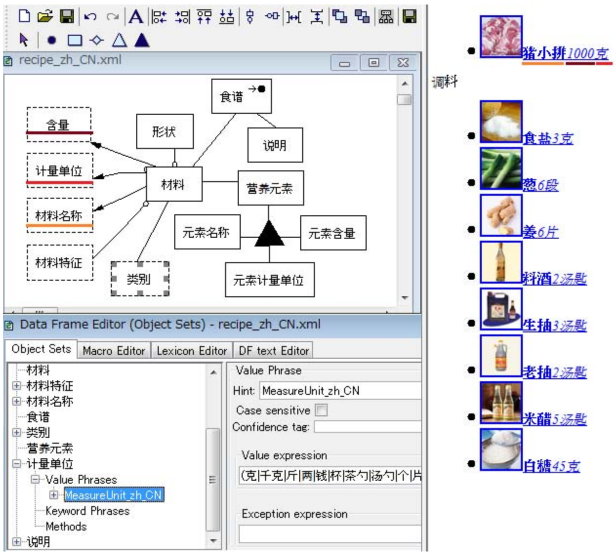


Figure 2. Using data extraction ontologies to annotate ingredients in Chinese language recipes.

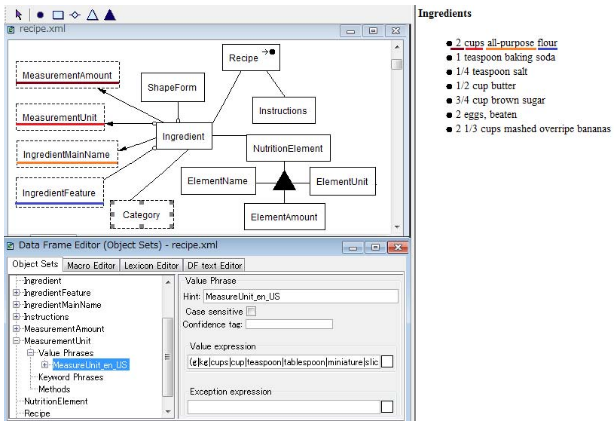
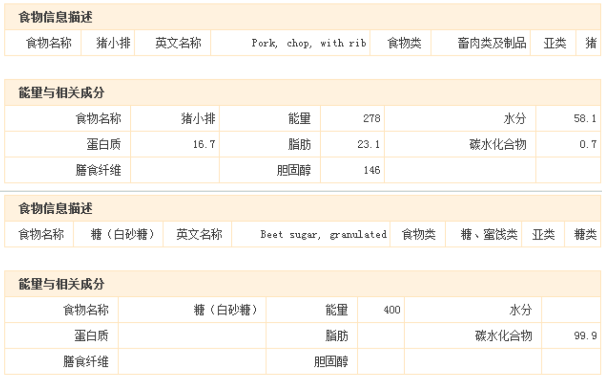


Figure 3. Using data extraction ontologies to annotate ingredients in English language recipes.

Once the recipes have been annotated with this language-agnostic mechanism, we can enrich the resulting RDF annotation with other relevant nutrion information and apply culture-relevant policies to construct replies to queries. These processes are further described in Sections III-B and III-C.

B. Nutrition-relevant augmentation

In order to augment semantic annotations associated with recipe ingredients described in Sub-Section III-A with nutrition information, the ingredients are matched against a multilingual nutrition database. As illustrated in Figure 1, currently we are mapping three databases obtained from



食物信息描述							
食物名称	猪小排	英文名称	Pork, chop, with rib	食物类	畜肉类及制品	亚类	猪

能量与相关成分					
食物名称	猪小排	能量	278	水分	58.1
蛋白质	16.7	脂肪	23.1	碳水化合物	0.7
膳食纤维		胆固醇	146		

食物信息描述							
食物名称	糖（白砂糖）	英文名称	Beet sugar, granulated	食物类	糖、蜜饯类	亚类	糖类

能量与相关成分					
食物名称	糖（白砂糖）	能量	400	水分	
蛋白质		脂肪		碳水化合物	99.9
膳食纤维		胆固醇			

Figure 4. A sample table extracted from the Chinese nutrition database [7].

government sources of China, USA and Japan [7], [8], [9][1]. Of these three databases, the USA nutrition database provided by the US Department of Agriculture (USDA) is the most extensive in terms of nutrients and variety of ingredients. On the other hand, the Chinese and Japanese databases, though less extensive with respect to variety of ingredients and nutritional elements, contain information for other traditional ingredients not found in their USA cousin nutrition database. Figures 4 and 5, provide samples of nutritional elements and their values per serving, for the Chinese and USA nutrition databases respectively[2]. In combining these databases, an effort is made not to duplicate ingredients in the three databases by creating a canonical label for the ingredients. However, the nutritional values of these ingredients tend to vary because they are calculated by different organizations, under different standards and methodologies. To tackle this problem, nutritional values are tagged with provenance information. This provenance annotation is then used during reasoning to produce one, two or three different values depending on the availability of the ingredient in the respective databases. When a response is generated to satisfy a user query, provenance information is also provided, when available and as needed.

C. Culture-relevant augmentation

In order to tackle culture-relevant aspects in replying to queries about recipe and their ingredients, a set of policy service subsystem based on the KAoS framework [10], is introduced. In essence, this subsystem is an extension of the KAoS core policy ontology using a set of OWL policies for each particular culture that encapsulate cultural and religious

[1]It should be noted that combination of these database, though necessary to capture the various cultural idiosyncrasies, intrinsically embedded, inevitably, in those sources, is not a trivial task. However, due to space limitations the authors leave a more detailed discussion for a future paper.

[2]To save valuable space, we have not included an example of the Japanese nutrition database on purpose.

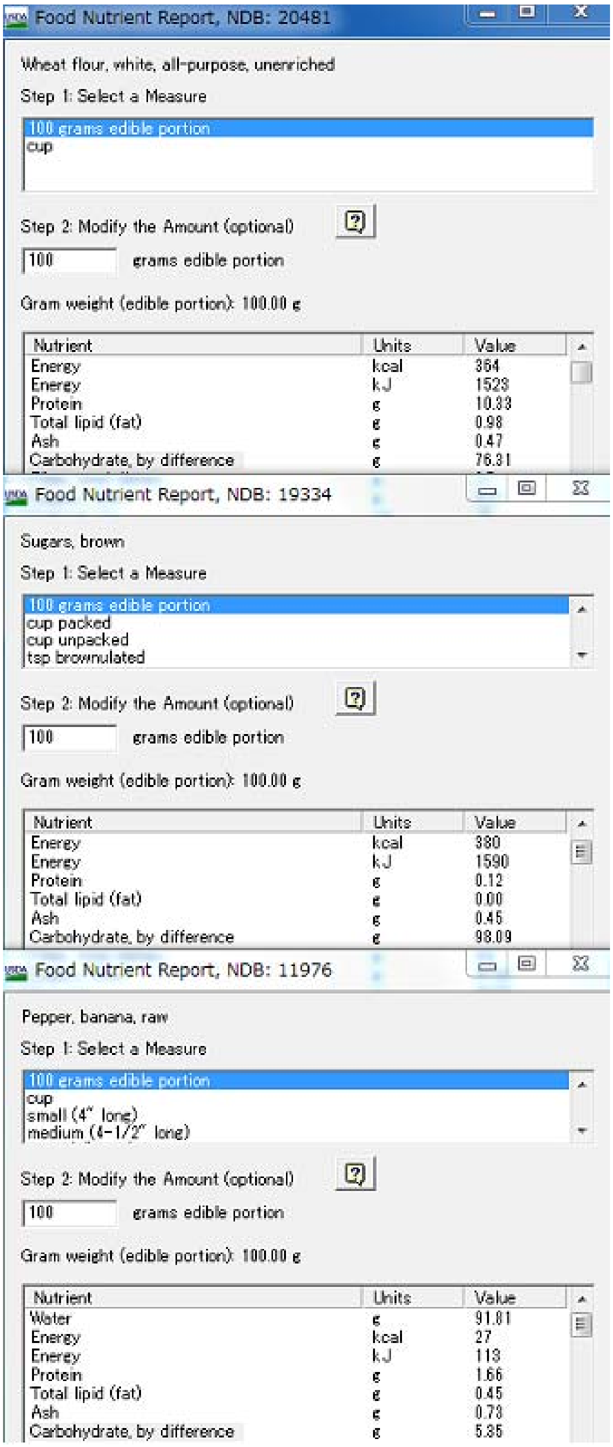


Figure 5. Sample nutrients for various ingredients from the USDA nutrients database [8].

constraints about ingredients[3] during reasoning. We extend the core policy ontologies as follows:

1) CultureAction.owl defines culture-specific action classes and their properties. New Action classes are subclassed from the core class http://ontology.ihmc.us/Action.owl#Action. Examples of culture-specific action classes might include actions describing culture-related acceptance of various ingredient groups.
2) CultureActor.owl - defines culture actor classes (or roles) and their properties. New Actor classes are subclassed from http://ontology.ihmc.us/Actor.owl#Actor. Examples of culture-specific actor classes might include consumer recipe sites, health and nutrion organizations, or nutrion and health practitioners who issue health warnings or nutrition reports.
3) CultureEntity.owl defines culture specific entities and their properties. These are used to define contexts for culture actions. New Entity classes are subclassed from http://ontology.ihmc.us/Entity.owl#Entity. Examples of culture-specific entity classes might include recipe, ingredient, religion, traditions, health warnings and nutrion reports.

IV. Conclusion

This paper described the preliminary design of an approach for cross-language annotation of recipes and augmentation of those annotations with information found in public nutrition databases. The approach enables the construction of replies to complex dietary queries about recipes from multilingual sources using nutrition information found in multilingual nutrition databases. In aswering queries, other culture-relevant constraints can also be taken into consideration through the use of a cultural policy service subsystem based on the KAoS framework. Although this is still work in progress, previous results in the domain of restaurant recommendations [6] indicate that the approach will be viable with respect to cross-language annotations or recipes. In this paper, we extend the approach further to incorporate additional annotations relevant to nutrition, which was not described in that approach. In addition, semantic reasoning is further enhanced through the use of cultural policies.

References

[1] M. Montari, *Food is Culture*. Columbia University Press, 2006.

[2] J. Brillat-Savarin, *Physiologie du Gout, ou Meditations de Gastronomie Transcendante*. Larousse, 1826.

[3] P. G. Kittler, K. P. Sucher, and M. N. Nelms, *Food and Culture*, 6th ed. Wadsworth Cengage Learning, 2006.

[4] B. Ashley, J. Hollows, S. Jones, and B. Taylor, *Food and Cultural Studies – Studies in Consumption and Markets*. Routledge, 2004.

[5] Merriam-Webster, *Merriam-Webster's Collegiate Dictionary, 11th Edition thumb-notched with Win/Mac CD-ROM and Online Subscription*. Merriam-Webster, Jul. 2003. [Online]. Available: http://www.worldcat.org/isbn/0877798095

[6] D. W. Embley, S. W. Liddle, D. W. Lonsdale, and Y. A. Tijerino, "Multilingual ontologies for cross-language information extraction and semantic search," in *ER'11*, 2011, pp. 147–160.

[7] C. C. for Disease Control and Prevention. (2012, Jul.) Nutrion database. [Online]. Available: http://chinacdc.cn

[8] U. D. of Agriculture. (2012, Jul.) Agricultural research service. 2011. usda national nutrient database for standard reference, release 24. nutrient data laboratory home page. [Online]. Available: http://www.ars.usda.gov/ba/bhnrc/ndl

[9] J. M. of Education Culture Sports Science and T. (MIXT). (2012, Jul.) Food composition database. [Online]. Available: http://www.fooddb.jp/

[10] A. Uszok, J. Bradshaw, J. Lott, M. Breedy, L. Bunch, P. Feltovich, M. Johnson, and J. Hyuckchul, "New developments in ontology-based policy management: Increasing the practicality and comprehensiveness of kaos," in *In proceedings of IEEE Workshop on Policies for Distributed Systems and Networks, 2008. POLICY 2008*, 2008, pp. 145–152.

[3]An attempt is made to keep cultural policies as simple as possible by developing policies that pertain exclusively to the ingredients and not to their place of origin or mode of preparation. Although this might not be sufficient to answer strict religious constraints, such as Halal or Kosher rules of preparation, this approach is sufficient to answer "what?", as opposed to "how", type of queries.