Digitizing National Cuisines: Cooking Recipes as Conceptual Graphs

Dmitri Dmitriev
Institute of Linguistic Studies, Russian Academy of Sciences
St Petersburg, Russia
dmitri@globbie.net

Abstract

The use-case of digitizing national cooking recipes through consistent use of a foundational semantic ontology is considered. GSL-based formal representations of recipes are stored in Knowdy, an open-source graph database. Cross-cultural sharing of knowledge of the underresourced languages via intermediate semantic framework with NLP adapters appears to be the most efficient strategy to preserve the unique cultural background of different geographies.

Keywords: conceptual graph, cooking recipe, Knowdy graph DB, GSL, NLP

1. National cuisine as cultural phenomenon

Preserving ethnic and cultural diversity in general becomes a great challenge nowadays similar to the task of preserving the biological diversity of our planet. The endangering factors include the ubiqitous spread of globalized digital media that are unlikely to make allowances for local national traditions. National cuisine is widely agreed to be a key cultural phenomenon that forms national identity.

The millions of combinations of tastes, aromas, techniques, as well as historical, religious, and cultural allusions incrementally form the heritage with references in literature, folklore, music - all that makes national cuisines unique and interconnected with other areas of human activities.

Fabio Parasecoli rightly applies a concept of “signifying networks” to national cuisines: “Each element in a culinary tradition is thus also part of several interconnected networks of meaning, practices, concepts and ideals; the full extent of its meaning and value cannot be grasped without analysing its interaction with other apparently unrelated domains. We can define these networks as "signifying" because they help us make sense of reality, allowing us to comprehend our cultural environment and to act within its rules and boundaries” (Parasecoli, 2005).

The notion of signifying networks allows us to model a graph of numerous components that could help us keep track of unique and shared features of national cuisines: “These signifiers therefore define local, regional or even national identities, and include ingredients, techniques, trade, location, time and media, all of which give rise to variations and, eventually, differences that are interpreted as national. Yet national cuisines remain a complicated part of the world of globalization (and, in the European context, of pan-European administration). Russia is one country where the broad array of influences on national cuisines is evident” (Smith, 2012).

As the world becomes smaller in terms of travelling and communication, we get a lot of opportunities to discover new cultural dimensions to ourselves. One can tell quite a lot about a national group just by trying its famous dishes. Foreign visitors are often quite keen to try local cuisine but might find it risky unless enough explanatory information is provided.

Sharing cooking recipes not only encompasses a list of plain ingredients and cooking directions but also the environment where food products grow. National cuisines involve a big number of factors that make the dishes special, including specific ways of whole food processing, the use of utensils, applying cooking techniques etc.

Thus the use case of digitizing cuisine is quite instructive for understanding the principles of present day cross-cultural knowledge exchange.

Today’s big challenge is to encode this information into a proper digital form so that the data exchange can open doors to foreign tourists, boost economic ties, and bring cross-cultural communication to a much higher level.

2. Foundational ontologies and digital formalisms

Existing technologies of formalized knowledge representation fall into several groups of frameworks. These include Semantic Web approach in its original form of OWL and RDF. Online collections of interrelated datasets using Semantic Web instruments are known as Linked Data. Many digitizing projects are built around an idea of using some kind of foundational ontology that can be extended by knowledge engineers in a particular specific field of expertises. We shall consider the use-case of applying these tools to digitizing the national cuisines. A comprehensive overview of approaches to developing ontologies related to culinary area is given in (Ribeiro et al., 2006).

2.1 RDF

The dominant rationale of RDF is that “the Web is moving from having just human-readable information to being a world-wide network of cooperating processes. RDF provides a world-wide lingua franca for these processes”[1]. As its name suggests, RDF is a framework for expressing information about *resources* – primarily web documents and various entities. Its formalism is based around an idea of static classes and properties. The question naturally arises as to what extent it is sensible to treat a recipe as an entity rather than as a complex process with arguments, timings, nested complexity etc.

[1] https://www.w3.org/TR/rdf-concepts/

2.2 Schema.org Initiative

Schema.org is a collaborative community activity with a mission to create, maintain, and promote schemas for structured data on the Internet. This initiative aims at providing a standardized vocabulary for shared metadata of published web resources.

Web resources related to cooking can use the metadata fields of a Recipe class, maintained here: http://schema.org/Recipe. Let's consider an example of a recipe markup provided by Google.[2]

```
<script type="application/ld+json">
    {
"@context": "http://schema.org/",
"@type": "Recipe",
"name": "Strawberry-Mango Mesclun Recipe",
"image": [
"https://example.com/photos/1x1/photo.jpg"
  ],
"author": {
"@type": "Person",
"name": "scoopnana"
},
"datePublished": "2008-03-03",
 "description": "Mango, strawberries, and sweetened dried cranberries are a vibrant addition to mixed greens tossed with an oil and balsamic vinegar dressing.",
"aggregateRating": {
  "@type": "AggregateRating",
  "ratingValue": "5",
  "reviewCount": "52"
},
    "prepTime": "PT15M",
    "totalTime": "PT14M",
    "recipeYield": "12 servings",
    "nutrition": {
 "@type": "NutritionInformation",
 "servingSize": "1 bowl",
 "calories": "319 cal",
 "fatContent": "20.2 g"
    },
"recipeIngredient": [
  "1/2 cup sugar",
  "3/4 cup canola oil",
  "1 teaspoon salt",
  "1/4 cup balsamic vinegar",
  "8 cups mixed salad greens",
  "2 cups sweetened dried cranberries",
"1/2 pound fresh strawberries, quartered",
  "1 mango - peeled, seeded, and cubed",
  "1/2 cup chopped onion",
  "1 cup slivered almonds"
 ],
"recipeInstructions": "\n1. Place the sugar, oil, salt, and vinegar in a jar with a lid. Seal jar, and shake vigorously to mix.\n2. In a large bowl, mix salad greens, sweetened dried cranberries, strawberries, mango, and onion. To serve, toss with dressing and sprinkle with almonds."
     }
   </script>
```

Code listing 1: Linked Data Recipe representation using JSON format and Schema.org's vocabulary.

The vocabulary and format adopted by Schema.org is primarily oriented towards representing high level metadata of web documents. In our view, a properly normalized semantic graph requires much more explicit representation of concepts. Most string values in Schema.org's fields are natural texts requiring human cognitive interpretation. Such a text can not be directly read by digital systems without special NLP tools that for the most part are fairly error-prone. As a funny example, we tried to apply Google own's machine translation to a single string item from a list of ingredients of their example recipe:

'*fresh strawberries, quartered*'

English-to-Russian translation produces an utterly incorrect sense disambiguation:

'*свежая клубника, расквартированная*'

with 'quartered' used in a sense of 'lodging' as in "Our troops were quartered in Boston" instead of `cut in four.'

The lack of clearly defined roles of objects and explicit identification of methods makes the task of machine translation of a recipe a lot more difficult.

Another futuristic challenge for digital recipes is to foresee a scenario where a robotic machine could follow the instructions to cook the dish. By no means could a raw string text representation assist in this task.

Let's consider a typical description of an ingredient: '*8 Granny Smith apples - peeled, cored and sliced*' It is pretty obvious that this text string contains a lot of classifying information: a raw food ingredient as a class, a specific cultivar of apples, quantity in pieces, a list of methods to be applied to each piece in order to actually use the ingredient – removing the seed center, the skin, cutting an object into parts of a specific shape.

The formalism of Schema.org is not expressive enough to help us encode the cooking process with some sort of programming code. To properly digitize recipes for robotic cooking, we need to separate ingredient descriptions from logic.

2.3 Cognitive Modeling of a Recipe

The traditional way to give directions of a recipe is to start with things and operations that are needed to be executed first. This order and style of description is known as imperative or procedural. The declarative or functional

[2] https://developers.google.com/search/docs/data-types/recipe

style of describing the logic of a process usually starts from the top of the execution pyramid, the expected useful result that we wish to achieve.

Consider the following simplified representation of a classical Russian recipe of a mushroom soup. Arrows denote subprocesses (sometimes alternative) needed by the parent process.

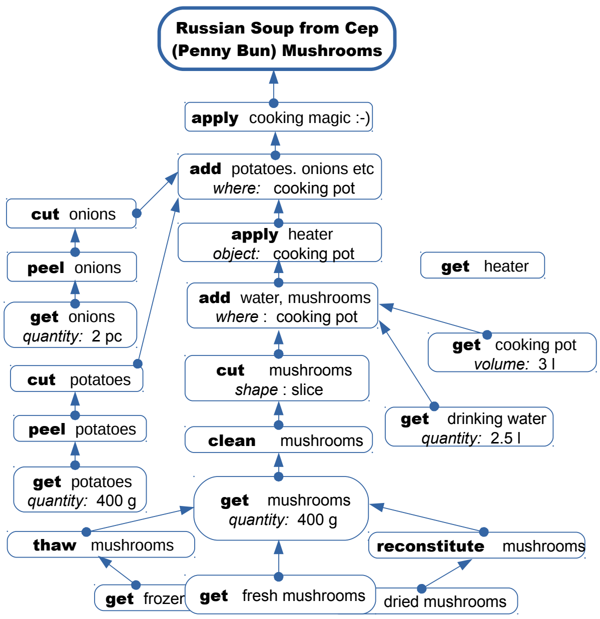

Figure 1: Representing a cooking recipe as a hierarchy of processes.

An instructive classroom experiment described in (Sam et al., 2014) aimed at the intuitive construction of an ontology for cooking recipes. The cognitive challenge of this task looks hard straight from the outset when we even try to define what the 'recipe' is. "The term *recipe* has several contextual meanings. It can be defined in a general sense as a method to obtain a desired end. When used in the context of cooking, it is generally considered to be a set of instructions on preparing a culinary dish. As such, it could be viewed as an object with properties such as ingredients and time needed. Alternatively, it could be viewed as a *process*, which takes in some input, has a series of steps to be executed, and produces some output. The time taken to execute the steps and the utensils needed also help describe the recipe" (Sam et al. 2014).

Once faced with the task of explicit explanation of concepts that are mostly known to us from our daily life experience, we as human beings tend to come up with different semantic segmentations of the shared reality. To some people the answer to this problem lies in imposing as many global standards as possible. However, in our view it's pointless to try and enforce any homogeneity in our cognitive shaping of foundational ontologies. The possible way of cross-cultural semantic integration in our view is to promote the use of ontologies that are tightly coupled with natural language processing.

3. Knowdy Project

Since 2006, in order to develop an optimal formalism for expressing complex semantics of natural language, we've been conducting research advocating for the interlingua approach to cognitive analysis of data (Dmitriev, 2006). Our research program was densely coupled with production-ready software development. One of our latest projects is focused on graph data management.

3.1 Open source software tools

Knowdy[3] is an open source software project of our research and development group in St Petersburg aiming at developing an ultra-fast graph database that allows direct and efficient manipulation with conceptual graphs, bypassing any intermediate representations like SQL tables. The database engine is implemented in plain C and can be used both as a network service as well as a standalone library for the embedded environment.

After several years of research and development our team of data scientists came up with a custom data format for Knowdy DB called GSL (an acronym for Globbie Semantic Language). GSL is optimized for compact storage of conceptual graphs. It is used for data storage, message passing and information exchange. This format is not so excessively verbose as XML and even slightly more compact than JSON. The language takes some features of S-expressions of Lisp, but with major modification of semantics since one should keep in mind that graphs are not lists! Bracket notation in GSL has a special meaning, allowing users to express not only the multilevel grouping but also the CRUD operations within a database storage system.

3.2 Coding recipes as GSL graphs

In GSL notation a process is coded as a first class function that can be named or anonymous, supports inheritance from a base function, has arguments and subprocesses that can run in parallel. For the sake of saving space we'll limit ourselves to a couple of examples. The processes below describe some of the logic behind the above mentioned Russian mushroom soup recipe.

```
(proc prepare mushroom soup mix
     [_gloss {ru подготовка заправки
               грибного супа}]
     (base cooking by boiling)
   (arg cut-mushrooms
        {run prepare mushroom mix})
   (arg cut-potatoes
        {run prepare potato mix})
   (arg cut-onions
        {run prepare onion mix})
   {run _put
      [_gloss {ru Все ингредиенты выложить
                 в контейнер и перемешать.}]
      {obj _all}
      {into_loc container}})
```

[3] https://github.com/globbie/knowdy

```
(proc prepare mushroom mix
    [_gloss {ru подготовка
                грибной заправки}]
    (arg clean-mushrooms
        {run clean mushrooms})
    {run _cut
      [_gloss {ru Нарезать грибы
                  мелкими кубиками.}]
      {obj clean-mushrooms}
      {form slice {size 1.5 {unit cm}}}})
```

Code listing 2: GSL declarative descriptions of cooking processes.

4. Knowledge sharing via crowd-sourcing

4.1 Web technologies

What we'd like to offer to the community of linguists, anthropologists, knowledge engineers, and all other interested parties is a set of methodologies, digital formats, and software tools to help establish a collaborative platform for knowledge sharing. Surely, we are far from thinking that such complex formalisms like GSL semantic graphs should be directly used by local communities that wish to share their cultural heritage with the rest of the world. For this purpose a different class of web based authoring tools can be applied (Dmitri Dmitriev, 2014).

4.2 Natural language processing

We advocate for the wider use of natural language processing to convert the free text input provided by a user, eg. the cooking directions from the original language into a set of interrelated semantic propositions that can be generated on the fly and presented in a user-friendly graphical interface. The propositions can be restated in another natural language or the same original language but in a more generic way. If any of the automatically parsed propositions seem incorrect or ultimately wrong, one can either try to restate the original instruction or redirect this issue to a support team.

One of our current projects aims at producing a semantic transcription of the traditional cuisine of the Mari.[4] It is a semi-automated process in which the unique lexical items of Mari are mapped by the language experts to a semantic ontology, eg. an atomic lexical item **тӱрлаш** means 'crimping the edges of an unleavened dough to seal a pie with decorative patterns' (Yuadarov, 2009). All this information must be explicitly linked with the interlingual representation of the ontology concepts:

```
(proc sealing a pie with decorative patterns
   [_gloss {mari тӱрлаш}
           {ru защипывать узоры на пирогах
               из пресного теста}]
   (base sealing a pie
       (arg obj (class Unleavened rolled up
                    dough))
       (arg method (class  Decorative Pattern))))
```

Code listing 3: GSL mapping between lexical items and a foundational ontology

[4] https://en.wikipedia.org/wiki/Mari_people

5. Use cases and practical applications

Nowadays, the culinary topics are of vivid interest in social networks and mobile applications. Our GSL-based formalism was tested in various commercial and non-profit projects. In 2017, a database of more than 3,000 recipes for multicooking devices was compiled with Russian as a primary language, and 5-6 other local national languages (eg. Mari) being added currently.

6. Conclusion

The process of digitizing national cuisines remains an important and challenging task of modern civilization. Cross-cultural sharing of knowledge via intermediate semantic framework using NLP adapters appears to be the most efficient strategy to preserve the unique cultural background of different geographies.

7. Bibliographical References

Dmitriev, Dmitri. Semantic Interlingua for the Knowledge Base Creation // Electronic Government - Workshop and Poster Proceedings of the Fourth International EGOV Conference 2005, August 22-26, 2005, Copenhagen, Denmark. Schriftenreihe Informatik 13 Universitätsverlag Rudolf Trauner, Linz, Austria 2005, ISBN: 3-85487-830-3. P. 11—18.

Dmitriev, Dmitri. Web lexicography for and by non-tech-people / Proceedings of the 4th International Workshop on Spoken Language Technologies for Under-resourced Languages SLTU-2014. St. Petersburg, Russia, 2014. P. 247-251. ISBN: 978-5-8088-0908-6.

Parasecoli, Fabio. Introduction to : Culinary cultures of Europe: identity, diversity and dialogue. Ed. Merkle, Kathrin; Goldstein, Darra; Mennell, Stephen; Council of Europe. Committee for Educational Research. Strasbourg: Council of Europe, 2005. ISBN: 9287157448.

Ribeiro R., Batista F., Pardal J.P., Mamede N.J., Pinto H.S. Cooking an Ontology. In: Euzenat J., Domingue J. (eds) Artificial Intelligence: Methodology, Systems, and Applications. AIMSA 2006. Lecture Notes in Computer Science, vol 4183. Springer, Berlin, Heidelberg. ISBN: 978-3-540-40930-4. P. 213-221.

Sam, Monica; Krisnadhi, Adila Alfa; Wang, Cong; Gallagher, John; Hitzler, Pascal. An Ontology Design Pattern for Cooking Recipes: Classroom Created // Proceedings of the 5th International Workshop on Ontology and Semantic Web Patterns (WOP 2014), CEUR Workshop Proceedings 1302. Aachen, Germany: CEUR-WS.org. P. 49-60.

Smith, Alison K. National Cuisines // *The Oxford Handbook of Food History.* Edited by Jeffrey M. Pilcher. Oxford, 2012. ISBN: 9780199729937.

Yuadarov, K. Mari Peasant Food. Yoshkar-Ola, 2009. [in Russian: Марийская крестьянская кухня]