

Constructing Cooking Ontology for Live Streams

Shanlin Chang

Department of Information Management
National Sun Yat-sen University
Kaohsiung, Taiwan
Email: slchang@mis.nsysu.edu.tw

San-Yih Hwang

Department of Information Management
National Sun Yat-sen University
Kaohsiung, Taiwan
Email: syhwang@mis.nsysu.edu.tw

Yu-Chen Yang

Department of Information Management
National Sun Yat-sen University
Kaohsiung, Taiwan
Email: ycyang@mis.nsysu.edu.tw

Abstract

We build a cooking domain knowledge by using an ontology schema that reflects natural language processing and enhances ontology instances with semantic query. Our research helps audiences to better understand live streaming, especially when they just switch to a show. The practical contribution of our research is to use cooking ontology, so we may map clips of cooking live stream video and instructions of recipes. The architecture of our study presents three sections: ontology construction, ontology enhancement, and mapping cooking video to cooking ontology. Also, our preliminary evaluations consist of three hierarchies—nodes, ordered-pairs, and 3-tuples—that we use to referee (1) ontology enhancement performance for our first experiment evaluation and (2) the accuracy ratio of mapping between video clips and cooking ontology for our second experiment evaluation. Our results indicate that ontology enhancement is effective and heightens accuracy ratios on matching pairs with cooking ontology and video clips.

Keywords

Natural language processing, semantic query, ontology construction, ontology enhancement, live stream

1 Introduction

With the emergence of grassroots internet journalists, the era of self-media is coming (Gillmor 2006). The term “self-media” refers to independently operated social media accounts, usually run by individual users on social networking platforms (e.g., Blog, Facebook, Twitch, Twitter, WeChat, Weibo, YouTube). These platforms share a similar feature: live streaming (also referred to as “streaming”). Streaming allows users to broadcast live video content with a camera and a computer/smartphone through the Internet. Hence, the broadcast industry has been undergoing a dramatic change recently (Doyle 2010). People can make and broadcast their own media content, including news, games, cooking, travel, sports, shopping, drama, and commentary, among others. Also, users can produce and broadcast live video at any time and any place, and viewers can watch these videos via the web.

Compared to traditional live TV shows, streaming makes broadcast preparation easier. For example, a traditional live show usually takes much effort to prepare prior and/or post production work. As a result, audiences often find it challenging to view live shows when they join a stream video (Buykx and Petrie 2011; Hayashi et al. 2013; Oh et al. 2016). However, new technologies such as image recognition (He et al. 2016) and speech recognition (Hannun et al. 2014) offer new opportunities to address these challenges. Among these stream videos, cooking-related videos are quite popular. However, semantic information about cooking is hardly used to facilitate viewers’ understanding of a live stream. Hence, we aim to construct cooking ontology by collecting a large amount of recipe data on the Internet, and we align these recipe instructions with live cooking videos. We believe that our study will help viewers better understand live streaming, especially when they just begin to watch a particular show. Viewers will then be able to quickly understand not only how to cook, but also how to mutually communicate with video providers and other viewers.

Oh et al. (2016) addresses the video content alignment problem using a simple information retrieval approach, based on transcripts. Following Oh et al. (2016), we employ the technique of natural language processing (NLP) to construct cooking ontologies from recipes. We then solve the video content alignment problem by comparing the transcripts of video clips with the cooking ontology. Our results show that our proposed technique helps construct high quality cooking ontology. Furthermore, this technique more accurately maps video clips, based on our proposed method.

The rest of our paper is organized as follows. In Section 2, we describe related works on constructing cooking ontology from recipe and videos. In Section 3, we present the architecture of our approach. In Section 4, we describe our cooking ontology schema and detail our methods of construction, enhancement, and mapping live streaming to cooking ontology. In Section 5, we discuss our preliminary evaluations, which show the performance of our proposed method: ontology construction/enhancement and mappings of video clips. Finally, we conclude this paper with Section 6.

2 Related Work

2.1 Cooking Domain Knowledge Building using Ontology

A number of studies have investigated the construction of cooking ontology. Batista et al. (2006) describe a cooking ontology that includes four concepts: actions, food, recipes, and utensils. In addition to modules for constructing these four concepts, there are three auxiliary modules that construct auxiliary concepts, such as units and measures, equivalencies, and plate types. The resultant ontology is used subsequently to facilitate a generic dialogue system in home domains. The JColibriCook system (DeMiguel et al., 2008) is based on ontology and extends Case-Based Reasoning (CBR) for ColibriCook computer cooking contests. Its ontology contains not only ingredients, but also dish features, such as formal type, cuisine type, and dietary type. JColibriCook is built around four components in the CRB cycle: retrieval, reuse, revision, and retainment. Also, to test proposed systems, the Computer Cooking Contest provides five exercise queries to search for recipes and create new dishes. JColibriCook also maps ingredients and creates recipes. They propose to construct a domain ontology and extend query for different degrees of similarity. Several studies (Cordier et al. 2012; Cordier et al. 2009; Gaillard et al. 2014) have extended this JColibriCook system. A notable example is the TAAABLE system, which evolved into wikiTAAABLE by Cordier et al. (2009) and includes a semantic wiki for TAAABLE. Cordier et al. (2012) then use WikiTaaable to include user feedback and incorporate another adaptation knowledge (AK) resource. Collectively, these studies help construct domain knowledge to promote a better understanding of cooking ontology. Such concepts could be linked, extended, and reused.

2.2 Interaction on Cooking Video

Large numbers of cooking stream videos are available on the Internet, and it is imperative to identify knowledge from these video clips. Doman et al. (2011) apply image recognition techniques to identify cooking recipes from cooking videos, but they do not use any domain ontology to facilitate recipe identification. Their proposed method will generate recipe video tagging. Oh et al. (2016) establish Cooking Video Annotation (CVA) ontology for aligning videos and recipes. They apply information extraction techniques with lexicon-syntactic patterns to identify chef, ingredient, ingredient portion, and cooking tool from video captions by using Named Entity Recognition (NER). Doman et al. (2011) and Oh et al. (2016) both study tagging videos, but their approaches are not intended to be applied to streaming video.

3 Research Framework

In this section, we present the design of a cooking ontology and show how it can be used to annotate a video stream. As we show in Figure 1, our approach involves three steps: cooking ontology construction, ontology enhancement, and mapping between cooking live streams and recipe instructions in the cooking ontology.

With respect to the ontology construction step, we define the schema of a cooking ontology and then populate the ontology from a large corpus that contain the text data of recipes. We use Natural Language Processing (NLP) techniques (e.g., part-of-speech tagging, stemming, dependency parsing, n-gram search, NER) to identify entities and derive the various types of concepts in the ontology. We then employ semantic web queries to further instantiate concepts in the ontology. Finally, we map stream video clips to recipe instructions within the cooking ontology.

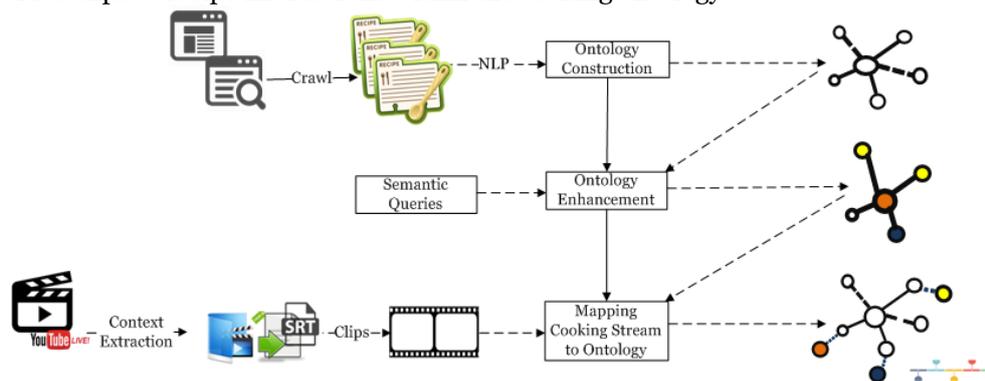


Figure 1: Skeleton of our approach

4 Building Cooking Ontology

4.1 Ontology Schema

Cooking ontology involves many concepts related to the cooking process, including culinary materials, the amount of units, utensils used, cooking instructions, and so forth. Our cooking ontology has six concepts and six properties, as we show in Figure 2. Batista et al. (2006) define four concepts related to cooking: food, actions, utensils, and recipes. Following the work of Batista et al. (2006), we separate the food concept into ingredients and raw food, since some ingredients are processed from raw food (e.g., cheese is defined as a *raw food* since it is an *ingredient* processed from “milk”). The following example shows how we construct the cooking ontology. A statement in a recipe is “60g/2¼oz shiitake mushrooms, sliced.” This statement can be divided into two parts: raw food (e.g., mushrooms) and processes to transform food into an ingredient (e.g., shiitake mushrooms, sliced). A recipe includes step-by-step *instructions*. Each instruction involves some action that incorporates ingredients and kitchen utensils. Consider the following instruction: “Place the chicken in a frying pan.” This instruction involves an action “place”, a utensil “frying pan”, and an ingredient “chicken”. Finally, a recipe may be followed by the cooking processes of several live video streams.

Each concept can be described by attributes. The *Recipe* concept records include five attributes: recipe title, serves, recipe description, and recipe URL. The *Instruction* concept illustrates recipe cooking processes and operational actions. The *Ingredient* concept describes culinary material and how this material is transformed from a raw product; this concept also includes any restrictions. The *Raw food*

concept includes various aspects about food, such as name, category, and nutrition. The *Kitchen utensil* concept includes attributes that describe cookers, kitchenware, and food preparation appliances. Finally, the *Video* concept, retrieved from either a video-sharing website or generated by users, describes video content, captions, and so forth.

Property specifies the directional relation between concepts. The Recipe concept has three properties--Ingredient, Instruction, and Video—that we name *hasIngredient*, *hasSteps*, and *playedBy*, respectively. As we previously mention, an ingredient may be processed by some raw food, signified by the property *processedFrom*. The *useIngredient* property specifies that an instruction uses a certain ingredient. Finally, the *useUtensil* property reflects that an instruction uses some utensils for processing food.

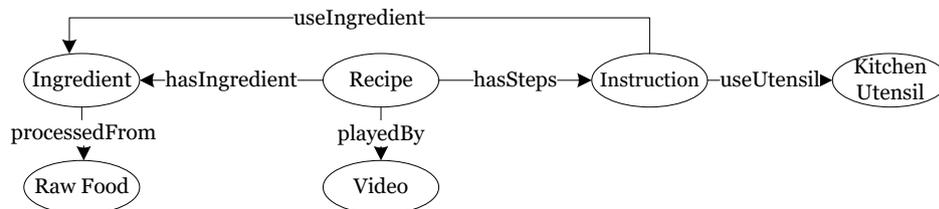


Figure 2: Cooking ontology

4.2 Cooking Ontology Construction

Prior studies (Buitelaar et al. 2005; IJntema et al. 2012) use NLP techniques to construct ontologies in various domains. NLP involves several techniques, including tokenization, part-of-speech (POS) tagging, sentence parsing, word dependencies, named entity recognition (NER), and so forth. Buitelaar et al. (2005) describe a six-layer approach for ontology learning, such as terms, synonyms, concepts, concept hierarchies, relations, and rules. Our study adopts a similar approach by using the pseudo-code that we list in Figure 3. We first collect recipes from some recipe websites. In Figure 4, we show one example of a BBC recipe, which has notable sections for ingredients and instructions. We first try to identify ingredient instances from the ingredient section by using NLP techniques. To do so, we consider two examples: “60g/2¼oz shiitake mushrooms, and sliced” and “2 x 7 g sachet of dried yeast.” We first apply a lexicon-based method to extract the names of ingredients and then use NER to identify measurements (e.g., 60g (grams), 2¼ oz (ounces)). However, in the second example, NER fails to identify the multiplication symbol, which means two packs of yeast. We thus prepare a set of rules to facilitate the identification of ingredients. In addition, we apply POS tagging to find verbs, nouns, adjectives, or quantifiers, which are part of ingredient attributes. In particular, a verb will serve as a label for the *processedFrom* property; for example, “sliced” is a label between the ingredient (mushroom) and the raw food (mushroom).

We then identify the other concepts from the method section, as we show in Figure 4. First, we extract ingredients described in the instruction by using *n*-gram ($n=1, 2, \text{ and } 3$) and then compare them with the ingredient instances we just extracted. Thus, ingredients mentioned in the method section (e.g., “white wine vinegar”) can be preserved. We next identify kitchen utensils mentioned in the instruction by using some predefined rules as defined in (Oh et al. 2016). One such rule is “IN {DT} {JJ} {NN}”. If we apply this rule to “in a large pot”, then the noun (pot) will be identified as a utensil. In addition, we extract nouns with the suffix “-ware”(e.g., glassware, silverware), as demonstrated on the Wikipedia kitchen utensil webpage. Finally, we confirm the retrieved nouns using lexicon from several Wikipedia categories (e.g., food preparation appliances, cooking appliances, cooking vessels). Finally, we extract actions from instruction sentences to form 3-tuples for instructions in our ontology: action, ingredient, and kitchen utensil. We parse an instruction sentence into a dependency tree via a Stanford dependency parser. If the root of the dependency is a verb (which thereby reflects an action) and the dependency tree involves some ingredient terms and kitchen utensil terms, we can subsequently form the 3-tuples. For example, consider the following sentence: “For the pizza base, place the flour, oil, water, and salt into a food processor and blend together until a dough is formed.” After parsing the sentence, we obtain a dependency tree with the root “place” in a conjunction with “blend”:

root (ROOT-o, place-6)
conj:and (place-6, blend-21)

As a result, we can form the following instruction tuples:

(place, flour, food processor)
(place, oil, food processor)
(place, water, food processor)

(place, salt, food processor)
(blend, flour, food processor)
(blend, oil, food processor)
(blend, water, food processor)
(blend, salt, food processor)

```

Let  $R = \{r_1, r_2, \dots, r_n\}$  // recipes
Let  $r_n = \{r_{ing}, r_{ins}\}$  // ingredient section and instruction section of recipes
Let  $V = \{v_1, v_2, \dots, v_m\}$  // video
Let  $c_r, c_{ing}, c_{food}, c_{ku}, c_{ins}, c_v$  be cooking ontology concepts
Let  $a_{ing\_measurement}, a_{ing\_multiplication}$  be attributes of ingredient concept
Let  $t(w_{action}, w_{ing}, w_{ku})$  be attributes of instruction concept
Let  $p_{hasIngredient}, p_{processedFrom}, p_{useUtensil}, p_{useIngredient}, p_{hasSteps}, p_{playBy}$  be cooking ontology
properties
Let  $D_{ing}, D_{ku}$  be list // ingredient dictionary and kitchen utensil dictionary

Input:  $r_{ing}, c_{ing}, c_{ing}$ 
Output:  $a_{ing\_measurement}, a_{ing\_multiplication}, p_{hasIngredient}, p_{processedFrom}, D_{ing}$ 
1 for each ingredient  $r_{ing}$  in recipe  $r_n$ 
2   A measurement of  $c_{ing}$  ingredients  $a_{ing\_measurement} = NER(r_{ing})$ 
3   A multiplication of  $c_{ing}$  ingredients  $a_{ing\_multiplication} = Rule(r_{ing})$ 
4   Property  $p_{processedFrom} = part-of-speech(r_{ing}) \cap verb\ of\ part-of-speech$ 
5   Build a dictionary of ingredient  $D_{ing}$  from  $c_{ing}$ 
6   Build a property  $p_{hasIngredient}$  between concept recipe  $r_n$  and ingredient  $c_{ing}$ 
7   Build a property  $p_{processedFrom}$  between concept ingredient  $c_{ing}$  and raw food  $c_{food}$ 

Input:  $r_{ins}, c_{ins}$ 
Output:  $c_{ku}, D_{ku}, p_{useUtensil}$ 
1 for each instruction  $r_{ins}$  in recipe  $r_n$ 
2    $c_{ku\_temp} = Pattern(r_{ins})$ 
3   Concept kitchen utensil  $c_{ku} = c_{ku\_temp}$  in Wikipedia | suffix check ( $c_{ku\_temp}$ ) = true
4   Build a dictionary of kitchen utensil  $D_{ku}$  from  $c_{ku}$ 
5   Build a property  $p_{useUtensil}$  between concept instruction  $c_{ins}$  and kitchen utensil  $c_{ku}$ 

Input:  $D_{ku}, D_{ing}, r_{ins}, c_{ins}$ 
Output:  $p_{hasSteps}, p_{useIngredient}, T(w_{action}, w_{ing}, w_{ku})$ 
1 for each instruction  $r_{ins}$  in recipe  $r_n$ 
2    $w_{action} = \{w_{action\_root}, w_{action\_verb}\}$ 
3    $w_{action\_root} = part-of-speech(r_{ins}) \cap verb\ of\ part-of-speech \mid root\ of\ dependency$ 
   parser
4    $w_{action\_verb} = conj\ with\ w_{action\_root}\ in\ parser\ tree$ 
5    $w_{ing} = max(n-gram(r_{ins}) \cap D_{ing})$ 
6    $w_{ku} = Pattern(r_{ins}) \cap D_{ku}$ 
7 for each parser tree( $r_{ins}$ )
8   Build attributes of  $c_{ins}$  instruction  $t(w_{action}, w_{ing}, w_{ku})$  by parser tree relation
9   Build a property  $p_{hasSteps}$  between concept recipe  $c_r$  and concept instruction  $c_{ins}$ 
10  Build a property  $p_{useIngredient}$  between concept instruction  $c_{ins}$  and ingredient  $c_{ing}$ 

Input:  $R, V$ 
Output:  $c_r, c_v, p_{playBy}$ 
1 for each video  $v_m$  in  $V$ 
2   Stream video  $v_m$  has a recipe  $r_n$ 
3   Build a property  $p_{playBy}$  between concept recipe  $c_r$  and video  $c_v$ 

```

Figure 3: Pseudo code for ontology construction

Cooking ontology has six concepts and six properties, as we mention in the previous paragraph. To store the large number of ontology instances that we derived using NLP techniques, we choose Neo4j, a popular DBMS for storing graph data, as the storage server. Concepts correspond to nodes in Neo4j, and attributes are labels of Neo4j nodes. Properties are stored as relations in Neo4j.

4.3 Enhancing Cooking Ontology

We note that instruction sentences may mention some ingredients that are not exactly the same as the ingredients described in the ingredient section. Take a pizza making recipe, for example. Although the ingredient section may include “olive oil,” the instruction section may only use “oil” instead, as we show

in our previous example. In fact, both “plain flour” and “flour” refer to the same thing. In this subsection, we describe our method for both identifying these ingredients and enhancing the instruction 3-tuples.

Our goal is to determine if two terms are semantically similar. To do so, we utilize Sematch API (Zhu and Iglesias 2017), which is a knowledge graph search system that allows users to compare semantic similarity. Sematch API maintains a knowledge base derived from WordNet, DBpedia, and YAGO. We follow the work of Sematch API and set the similarity threshold at 0.5. For example, the similarity between oil and olive oil when we use Sematch API is 0.8187 (>0.5). Therefore, we add the following 3-tuples to the recipe instruction ontology:

(place, plain flour, food processor)
(blend, plain flour, food processor)
(place, olive oil, food processor)
(blend, olive oil, food processor)

4.4 Mapping Live Stream to Cooking Ontology

In this subsection, we describe how we incorporate live streaming video into our ontology. To do so, we partition a video into several short video clips. Our goal is to establish a mapping between the instruction 3-tuples and the video clips. More formally, a recipe has a list of instructions, and each instruction is characterized by a set of 3-tuples (a, i, u) . Each video clip can be described as a transcript, derived from some audio-to-speech software. Our goal is to decide upon the closest instruction(s) for each video clip.

Given a set of 3-tuples $AIU_t = \{(a_1, i_1, u_1), (a_2, i_2, u_2), \dots, (a_k, i_k, u_k)\}$ for an instruction t , we first form its action union, ingredient union, and utensil union, as denoted by A , I , and U , respectively. Specifically,

$$A_t = \bigcup_{1 \leq j \leq k} a_j,$$

$$I_t = \bigcup_{1 \leq j \leq k} i_j,$$

$$U_t = \bigcup_{1 \leq j \leq k} u_j.$$

We then form three possible pairs:

$$AI_t = \bigcup_{1 \leq j \leq k} (a_j, i_j),$$

$$IU_t = \bigcup_{1 \leq j \leq k} (i_j, u_j),$$

$$AU_t = \bigcup_{1 \leq j \leq k} (a_j, u_j),$$

For a video clip transcript r , we first use NLP techniques to retrieve all tokens that are tagged as a verb, noun, or adjective. We then identify the actions, ingredients, and utensils, as denoted by A_r , I_r , and U_r , respectively. To compare the similarities between a video clip transcript r and an instruction t , we propose three measures, S^1 , S^2 , and S^3 , as defined below:

$$S_{r,t}^1 = \frac{|A_t \cap A_r|}{|A_t|} + \frac{|I_t \cap I_r|}{|I_t|} + \frac{|U_t \cap U_r|}{|U_t|},$$

$$S_{r,t}^2 = \frac{|AI_t \cap (A_r \times I_r)|}{|AI_t|} + \frac{|IU_t \cap (I_r \times U_r)|}{|IU_t|} + \frac{|AU_t \cap (A_r \times U_r)|}{|AU_t|},$$

$$S_{r,t}^3 = \frac{|AIU_t \cap (A_r \times I_r \times U_r)|}{|AIU_t|}.$$

Note that $S_{r,t}^1$ measures the overlap between the actions, ingredients, and utensils mentioned in the video clip r and the instruction t , and $S_{r,t}^2$ measures the overlap between the pairs of actions, ingredients, and utensils mentioned in the video clip r and the instruction t . Finally, $S_{r,t}^3$ shows how well the video clip r matches the complete 3-tuples in the instruction t .

For example, for an instruction “place the chicken in a frying pan” and a clip of “put garlic and chicken on a pan”, a cooking ontology tuple is (place, chicken, frying pan), and video content extractions are put, garlic, chicken, and pan. The instruction nodes include place, chicken, and frying pan. The instruction edges include (place and chicken), (place and frying pan), and (chicken and frying pan). Also, the instruction subgraph is (place, chicken, frying pan). Interconnection marks between cooking ontology and live streams reflect accurate rates more than 0.25. For this particular example, the accurate rate by nodes is $2/3$ ($0.67 > 0.25$), which namely reflects chicken and frying pan. We thus mark a relevance between the instruction and the clip. The next example for accurate rate by edge is $1/3$ ($0.33 > 0.25$).

5 Preliminary Evaluation

5.1 Dataset

We collect the recipes from two sources: a BBC Recipe website (<https://www.bbc.com/food/recipes>) and streamers' individual websites. Streaming videos are downloaded from YouTube live. The BBC recipe website includes more than 10,000 recipes. Some streamers provide cooking recipes on either their websites or YouTube description box. Recipes are stored in a Neo4j graph database. In Figure 4, we provide an example from a BBC recipe website.

With our proposed approach, we randomly extract 15 cooking videos from YouTube, and their recipes are from either the BBC Recipe site or streamers' websites. We obtain the transcript of a cooking live stream by using YouTube's auto-caption function. Because these transcripts may contain noisy data, we conduct a spelling check, POS, and stemming. Finally, we divide each streaming video transcript into several one-minute clip transcripts, and we show this video process in Figure 5. Meanwhile, we use Figure 6 to show the screenshot of a live streaming video example for roast chicken.

The screenshot shows a recipe page for 'Ten-minute pizza'. It includes preparation and cooking times (less than 30 mins and less than 10 mins), serving size (Makes 1), and the author (James Tanner). Below are interactive buttons for Share, Add to favourites, Shopping list, and Print recipe. The ingredients section is divided into 'For the pizza base' and 'For the topping'. The method section contains seven numbered steps for preparing the pizza.

Ten-minute pizza

Preparation time
less than 30 mins

Cooking time
less than 10 mins

Serves
Makes 1

By James Tanner
From Ready Steady Cook

Share Add to favourites

Shopping list Print recipe

Ingredients

For the pizza base

250g/8³/₄oz plain flour, plus extra for dusting
8 tbsp olive oil
2 tsp water
1 tsp sea salt

For the topping

4 tbsp tomato purée
60g/2¹/₄oz shiitake mushrooms, sliced
4 slices prosciutto
100g/3¹/₂oz gorgonzola cheese
1 free-range egg

How-to-videos

Method

1. Preheat the oven to 200C/400F/Gas 6.
2. For the pizza base, place the flour, oil, water and salt into a food processor and blend together until a dough is formed. Tip out onto a floured work surface and knead. Shape into a round base about 20cm/8in wide.
3. Place into a frying pan over a high heat and brown the base, then using a mini-blowtorch, crisp the top of the pizza. (Alternatively you can do this under the grill.)
4. For the topping, spread tomato purée over the top of the base.
5. Fry the mushrooms in a dry frying pan then scatter over the tomato purée. Arrange the prosciutto and cheese on top.
6. Crack an egg into the middle, then place into the oven for five minutes to finish cooking.
7. Serve on a large plate, and slice into wedges to serve.

Figure 4. A recipe example from a BBC recipe website

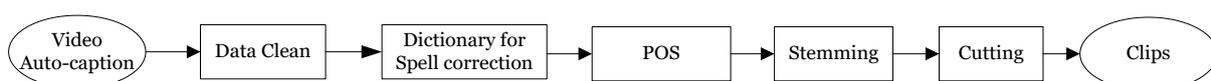
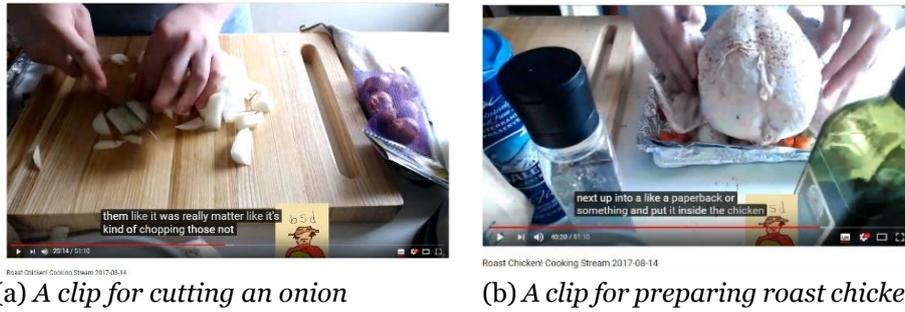


Figure 5. Stream video data process



(a) A clip for cutting an onion

(b) A clip for preparing roast chicken

Figure 6. Clips example from YouTube stream video

5.2 Ontology Evaluation

Our first evaluation concerns the recipes features from our constructed ontology. Specifically, we evaluate the ingredients and instructions in our ontology. To do so, we choose 9 recipes with a total of 65 instructions for evaluating instructions. We then manually tag ingredients sentences, which results in a set of ingredients for each recipe. We also manually inspect the instruction sentences and create 3-tuples, which involve actions, ingredients, and kitchen utensils.

To evaluate the instructions, we compare the 3-tuples manually notated by expert with those derived by our approach; we present these results in Table 2. We measure the hit ratio of the proposed methods, which is defined as the ratio of the corrected identified objects to the total number of correct objects. In addition to the hit ratio of 3-tuples, we also measure that of the ordered pairs (i.e., (action, ingredient), (ingredient, utensil), and (action, utensil)), as well as that of the single nodes (i.e., action, ingredient, and utensil). We then compare the performance of cooking ontology construction with and without enhancement. When we do so, we find that cooking ontology construction with (semantic) enhancement achieves a higher hit ratio across all three measures.

	Nodes(S^1)	Ordered-pair(S^2)	3-tuple(S^3)
Cooking ontology construction	78%	58%	21%
Cooking ontology construction and enhancement	86%	63%	34%

Table 2. Average hit ratio of instructions tuples

5.3 Video Clip and Ontology Mapping Evaluation

The second evaluation is a mapping between video clips and ontology. To begin, we manually tag 15 cooking streams, and a total of 459 one-minute video clips. When we map pairs between clips and instructions, we obtain 151 observations. When we tag a true positive (TP) of a clip with an instruction by extracting a number of three methods (i.e., nodes, pairs or tuples), we find more than one quarter of our total amount of cooking ontology. For each method of nodes, pairs, or tuples, we subtract repetitiveness. On the contrary, we do find that others are true negative (FN) tagging. Also, we manually compare the False Positive (FP) and True Negative (TN) with respect to tagging and unequal system tagging, and we find that the number of true positives is less than false negatives. Finally, to evaluate mapping pairs, we use two implements: precision = $TP / (TP + FP)$ and recall = $TP / (TP + FN)$.

Oh et al. (2016) examine a generation of interactive cooking videos by using semantic annotation by Wikipedia and DBpedia. In their study, they focus on creating semantic queries on ingredient and kitchen utensil terms. Their algorithm solves the mapping problem by means of LCS and sentence similarity over a threshold value. However, since closed cooking videos might have included post-production work and less video noise, their method is not suitable for analyzing the modern live streaming technique.

The result of our second evaluation, which we present in Table 3, may help account for ordered-pairs and tuples. Mapping video clips and cooking ontology results in ordered-pairs that are higher than nodes and close to 3-tuples. Also, the research methods that Oh et al. (2016) use are less effective for analyzing live video. Future work could consider memory benefits for identification antecedents and consequences. Because the video is continuous, the streamer must repeat the action.

Precision	Recall
-----------	--------

Cooking ontology construction(S^1)	0.372	0.539
Cooking ontology construction and enhancement(S^1)	0.488	0.629
Cooking ontology construction(S^2)	0.571	0.265
Cooking ontology construction and enhancement(S^2)	0.690	0.589
Cooking ontology construction(S^3)	0.545	0.199
Cooking ontology construction and enhancement(S^3)	0.648	0.550
Oh et al. 2016	0.167	0.066

Table 3. Mapping pairs evaluation of three methods

6 Conclusion and Future Work

The interconnection function must discern cooking processes from video auto-captions and cooking features. This study builds a domain knowledge for recipes, thereby extending our knowledge about semantic webs. Specifically, the results of our could underscore the usefulness of this idea. Reviewing our method could strengthen the establishment of property in our research methods, such as building sentences between tuples or the benefits of memory. Complementary, knowledge expansion of this study influences instances of concept (e.g., node data). Among topics that could be explored in future research, some important ones include translation and terminology. The translation of a dish or ingredient title may, for example, include semantic translation, free translation, word-for-word translation, or faithful translation. As just one example, Japanese cuisine consists of batter, cabbage, anything from meat, seafood, and so forth, all ingredients that are pan fried. The name for this particular Japanese cuisine is okonomiyaki, also called a "Japanese pancake" or "as-you-like-it pancake." Aside from translation, the other area that future studies could explore is terminology issues. When cooking in live stream videos, a chef may clearly and faithfully present cuisine skills and use professional language. However, the terminology that chef uses may be too complex for novice chefs or viewers to understand. For example, the term "Julienne" refers to a kind of cutting or shredding of food into 2-3 inch long, thin strips, and this term may not be fully understood by novice chefs or viewers. Terminology issues also meant that we were unable to use our algorithm to match with recipe. Perhaps, then, future work could consider related issues, such as referencing multi-language translations and broadening professional language.

7 References

- Batista, F., Pardal, J. P., Mamede, P. V. N., and Ribeiro, R. 2006. "Ontology Construction: Cooking Domain," *Artificial Intelligence Methodology Systems and Applications* (4183), pp. 213-221.
- Buitelaar, P., Cimiano, P., and Magnini, B. 2005. "Ontology Learning from Text: An Overview," *Ontology learning from text: Methods, evaluation and applications* (123), pp. 3-12.
- Buykx, L., and Petrie, H. 2011. "What Cooks Needs from Multimedia and Textually Enhanced Recipes," *IEEE International Symposium on Multimedia, California, USA: IEEE*, pp. 387-392.
- Cordier, A., Gaillard, E., and Nauer, E. 2012. "Man-Machine Collaboration to Acquire Cooking Adaptation Knowledge for the Taaable Case-Based Reasoning System," *Proceedings of the 21st International Conference on World Wide Web: ACM*, pp. 1113-1120.
- Cordier, A., Lieber, J., Molli, P., Nauer, E., Skaf-Molli, H., and Toussaint, Y. 2009. "Wiki-Taaable: A Semantic Wiki as a Blackboard for a Textual Case-Based Reasoning System," *6th European Semantic Web Conference, Heraklion, Greece*, pp. 88-101.
- DeMiguel, J., Plaza, L., and Díaz-Agudo, B. 2008. "Colibricook: A Cbr System for Ontology-Based Recipe Retrieval and Adaptation," *The 9th European Conference on Case-Based Reasoning, K.-D. Althoff, Bergmann, R., Minor, M., Hanft, A. (ed.), Trier, Germany*, pp. 199-208.
- Doman, K., Kuai, C. Y., Takahashi, T., Ide, I., and Murase, H. 2011. "Video Cooking: Towards the Synthesis of Multimedia Cooking Recipes," *International Conference on Multimedia Modeling, Taipei, Taiwan: Springer*, pp. 135-145.
- Doyle, G. 2010. "From Television to Multi-Platform: Less from More or More for Less?," *Convergence* (16:4), pp. 431-449.

- Fabian, M., Gjergji, K., and Gerhard, W. 2007. "Yago: A Core of Semantic Knowledge Unifying Wordnet and Wikipedia," Proceedings of the 16th international conference on World Wide Web, Alberta, Canada, pp. 697-706.
- Gaillard, E., Lieber, J., and Nauer, E. 2014. "Case-Based Cooking with Generic Computer Utensils: Taaable Next Generation," 22nd International Conference on Case-Based Reasoning, Cork, Ireland, pp. 254.
- Gillmor, D. 2006. *We the Media: Grassroots Journalism by the People, for the People.* " O'Reilly Media, Inc."
- Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., Prenger, R., Satheesh, S., Sengupta, S., and Coates, A. 2014. "Deep Speech: Scaling up End-to-End Speech Recognition," arXiv preprint arXiv:1412.5567).
- Hayashi, Y., Doman, K., Ide, I., Deguchi, D., and Murase, H. 2013. "Automatic Authoring of a Domestic Cooking Video Based on the Description of Cooking Instructions," Proceedings of the 5th international workshop on Multimedia for cooking & eating activities, Barcelona, Spain: ACM, pp. 21-26.
- He, K., Zhang, X., Ren, S., and Sun, J. 2016. "Deep Residual Learning for Image Recognition," Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, USA, pp. 770-778.
- IJntema, W., Sangers, J., Hogenboom, F., and Frasincar, F. 2012. "A Lexico-Semantic Pattern Language for Learning Ontology Instances from Text," *Web Semantics: Science, Services and Agents on the World Wide Web* (15), pp. 37-50.
- Maedche, A., and Staab, S. 2001. "Ontology Learning for the Semantic Web," *IEEE Intelligent systems* (16:2), pp. 72-79.
- Miller, G. A. 1995. "Wordnet: A Lexical Database for English," *Communications of the ACM* (38:11), pp. 39-41.
- Nyarko, J. 2016. "Newspaper Review Show in the Broadcast Media Space in Ghana: An Exploratory Qualitative Study," *Sage Open* (6:2), p. 2158244016654952.
- Oh, K.-J., Hong, M.-D., Yoon, U.-N., and Jo, G. 2016. "Automatic Generation of Interactive Cooking Video with Semantic Annotation," *Journal of Universal Computer Science* (22:6), pp. 742-760.
- Zhu, G., and Iglesias, C. A. 2017. "Computing Semantic Similarity of Concepts in Knowledge Graphs," *IEEE Transactions on Knowledge and Data Engineering* (29:1), pp. 72-85.