CookingSense: A Culinary Knowledgebase with Multidisciplinary Assertions

Donghee Choi[1]*, Mogan Gim[2], Donghyeon Park[3], Mujeen Sung[4], Hyunjae Kim[2], Jaewoo Kang[2†], and Jihun Choi[5†]

[1]Imperial College London, [2]Korea University, [3]Sejong University, [4]Kyung Hee University, [5]Sony AI
donghee.choi@imperial.ac.uk, {akim, hyunjae, kangj}@korea.ac.kr
parkdh@sejong.ac.kr, mujeensung@khu.ac.kr, jihun.a.choi@sony.com

2024 05 01
COLING-LREC 2024
Donghee Choi

DMIS Data Mining & Information Systems Laboratory   FNAI Lab   IMPERIAL   Sony AI

Why do we need CookingSense?

- Cooking is one of the most important human activities.
- Cooking is intertwined with many parts of our society, including restaurant business, food manufacturing, public health, and social media.
   - AI-driven culinary decision support systems ← Cooking Knowledge

Why do we need CookingSense?

Introduction: Cooking Knowledge and CookingSense

- But Cooking Knowledge is hard to be defined in a simple way.
  - ex) A chef who aims to make savory dishes for his guests
    - mastery of culinary arts
    - an in-depth understanding of ingredients and their food science
  - ex2) a diet researcher who focuses more on social well-being
    - understanding which ingredients or dishes are associated with health risks
- We need a multifaceted way to define Cooking Knowledge.
  - Many existing cooking-related knowledge resources tend to focus only on a specific aspect
    - e.g. recipe, nutrition, healthcare

Overview of CookingSense

CookingSense is built upon diverse culinary knowledge sources.
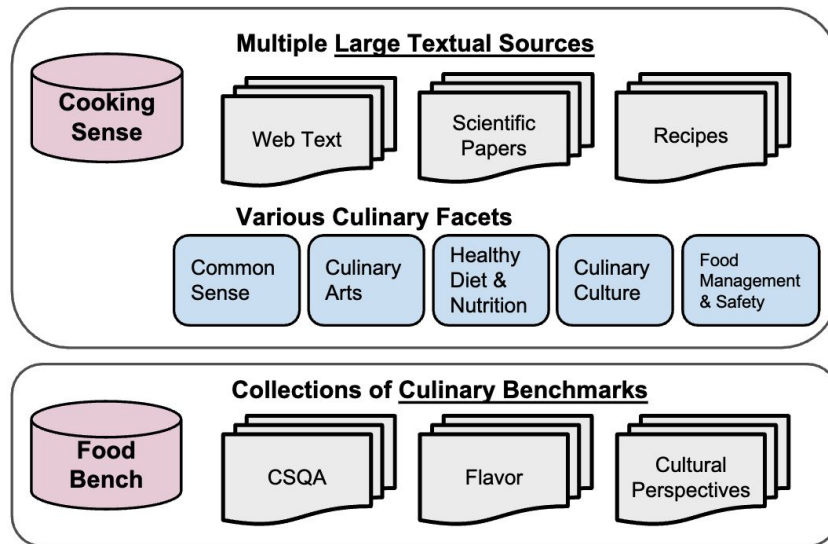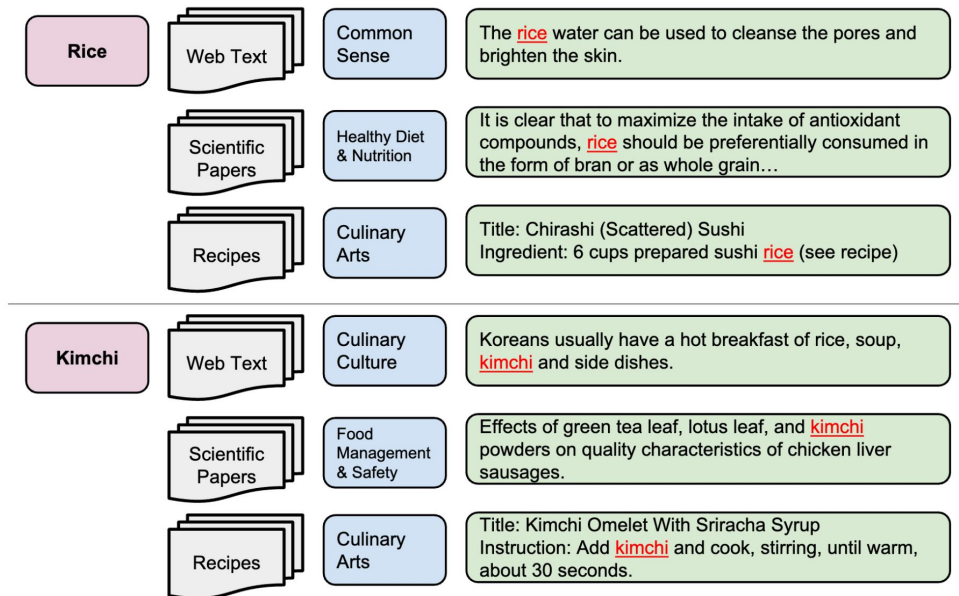
Examples of CookingSense

Figure 2: *CookingSense* examples. We present a selection of examples representing two culinary concepts, such as rice and kimchi, across various types in Table 3 from §3.2.

Comparison with Related works

CookingSense is a large-scale multi-faceted culinary textual knowledgebase

	Relevance	Source	Coverage	Relation	Volume
ConceptNet (Speer et al., 2017)		G	1 2 3 4 5	Structured	980K
FooDB (FooDB, 2020)	✔	C	3 4	Structured	6K
CANDLE (Nguyen et al., 2022b)		G	1 2 3 4 5	Textual	60K
Quasimodo (Romero et al., 2019)		G	1 2 3	Textual	627K
RecipeDB (Batra et al., 2020)	✔	C	2 3 4	Structured	118K*
CookingSense (Ours)	✔	G A C	1 2 3 4 5	Textual	54M

Table 1: Comparison of culinary KBs. **Relevance**: Direct relevance to culinary knowledge; **Source**: (G) General corpus, (A) Academic corpus, (C) Culinary-focused; **Coverage**: Each number implies (1) Food common sense, (2) Culinary arts, (3) Health & nutrition, (4) Culinary culture, (5) Food management & food safety; **Relation**: Structured indicates structured KB, while Textual is for textual KB; **Volume**: Number of sentences in the KB (*: Number of recipes).

Overview of CookingSense Construction Pipeline

Methodology
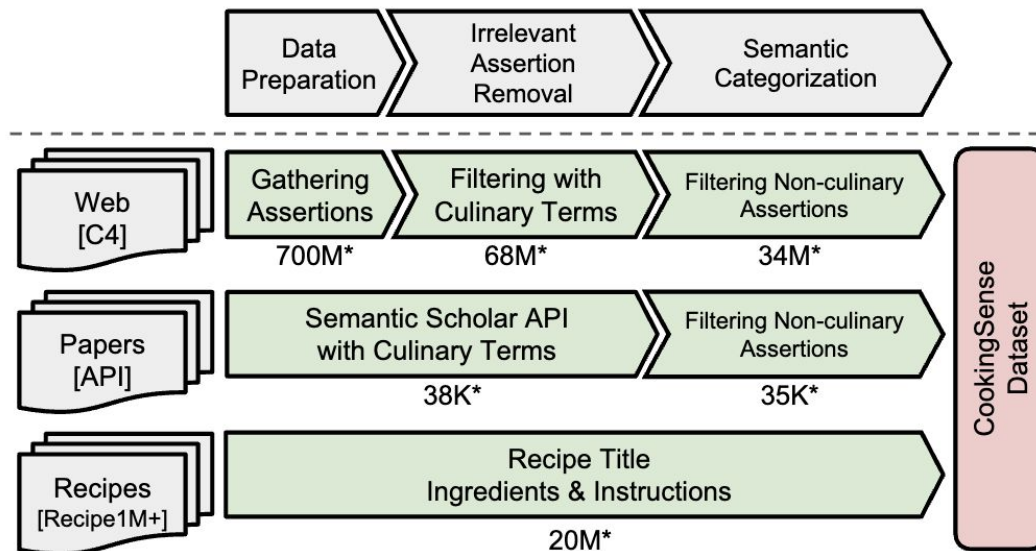
* Number of Assertions.

Removal of Irrelevant Assertions

- Culinary Dictionary
  - Ingredient and Food names
    - We use recipe titles and ingredient names in FooDB[1].
  - Bigrams that occur more than 3 times (food names) or 2 times (ingredient names)
  - 1,914 food names and 5,482 ingredient names
- Terms Collected from AI Assistant
  - We also need general terms like "Food" and "Nutrition"
  - 1,600 culinary terms using ChatGPT[2].
  - *ex) Please provide an exhaustive list of verbs related to cooking actions and techniques, such as chopping, slicing, seasoning, and garnishing.*

Web	Paper	Recipe
ice cream	fatty acids	olive oil
olive oil	gut microbiota	teaspoon salt
hot water	systematic review	black pepper
weight loss	oxidative stress	brown sugar
blood sugar	antioxidant activity	finely chopped

Table 4: Top 5 bigrams by frequencies in *CookingSense* by its sources.

1) https://foodb.ca/
2) https://chat.openai.com

Semantic Categorization

- Randomly sampled 10,000 assertions from the KB

- Labeling using GPT-4 model (version as of September 30, 2023)
    - Classify them into six distinct types: (a) Food CommonSense, (b) Culinary Arts, (c) Healthy Diet & Nutrition, (d) Culinary Culture, (e) Food Management & Food Safety, and (f) Irrelevant or None
    - Randomly chose 218 sentences from each category, resulting in a balanced dataset.
    - We split the dataset into 8:2 / Train:Test
        - We trained a classifier using bert-large-uncased. → 0.76 Acc for Test split
    - We applied this classifier to 68M assertions after removing irrelevant assertions, resulting in 34M categorized assertions, excluding those labeled as "Irrelevant or None".
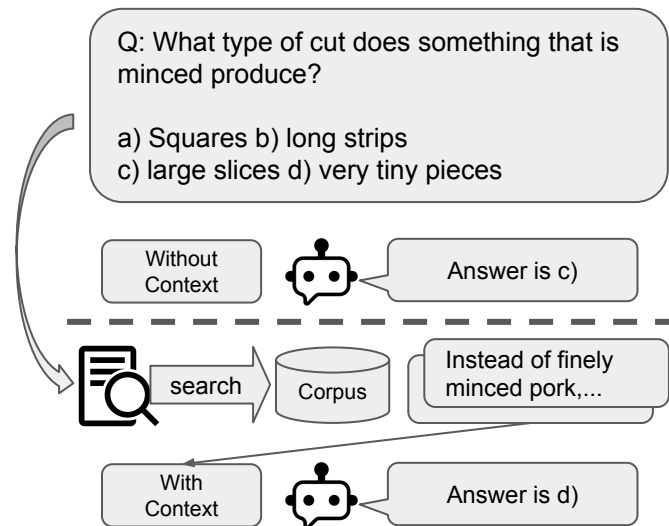
DMIS Data Mining & Information Systems Laboratory   FNAI Lab   IMPERIAL   Sony AI

Number of assertions in CookingSense

Distribution by types and sources

Type Description	Web	Paper	Recipe	Total
Food Common Sense	7,210,883	3,262	-	7,214,145
Culinary Arts	5,630,583	703	20,372,992	26,004,278
Healthy Diet & Nutrition	6,211,601	21,317	-	6,232,918
Culinary Culture	4,414,988	212	-	4,415,200
Food Management & Food Safety	10,846,348	9,596	-	10,855,944
All	34,314,403	35,090	20,372,992	54,722,485

FoodBench

- To assess the effectiveness of our KB, we adopt the context-augmented language model setup inspired by the work of Retrieval Augmented Generation (RAG[1]), where a context retrieved from a retriever system is augmented with the input to generate text.
  - Retriever system: Okapi-BM25
  - Language model: Flan-t5 model to generate answer.

Q: What type of cut does something that is minced produce?

a) Squares b) long strips
c) large slices d) very tiny pieces

Without Context — Answer is c)

search → Corpus → Instead of finely minced pork,...

With Context — Answer is d)

[1] Lewis, Patrick, et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks." NeurIPS (2020)

FoodBench

- To evaluate the utility of CookingSense and other baseline KBs, we have developed a benchmark, FoodBench.
  - a collection of culinary-related benchmark tasks covering question answering, flavor perspective prediction, and cultural perspective prediction.
- Multiple choice question answering format.
  - For some tasks, we convert the task into the format.

FoodBench

- Question answering
  - We collected 429 question-answer pairs from user-generated content on the web that reflects the real-world perspective of culinary knowledge; namely CookingSenseQA (CSQA).
- Flavor perspectives
  - We gathered and constructed flavor-related binary classification problems from the following resources (ASCENT++[1], The Good Scents Company; TGSC[2])
- Cultural perspectives
  - Cultural knowledge quizzes(CANDLE[3], FORK[4])

1) Nguyen, Tuan-Phong, et al. "Refined commonsense knowledge from large-scale web contents." TKDE (2022)
2) https://www.thegoodscentscompany.com/
3) Nguyen, Tuan-Phong, et al. "Extracting cultural commonsense knowledge at scale." Proceedings of the ACM Web Conference 2023
4) Palta, Shramay, and Rachel Rudinger. "FORK: A bite-sized test set for probing culinary cultural biases in commonsense reasoning models." Findings of the ACL 2023

DMIS Data Mining & Information Systems Laboratory   FNAI Lab   IMPERIAL   Sony AI

Results of FoodBench

Wow! CookingSense wins!

	CSQA	ASCENT++	TGSC	CKQ	FORK	Avg.
Without Context	16.08	24.52	13.60	14.38	28.80	19.48
ConceptNet (Speer et al., 2017)	47.79	22.90	47.60	54.25	46.20	43.75
FooDB (FooDB, 2020)	48.25	20.97	45.80	52.29	**58.70**	45.20
CANDLE (Nguyen et al., 2022b)	48.48	41.29	51.40	54.58	39.67	47.08
Quasimodo (Romero et al., 2019)	50.35	40.65	63.40	53.59	53.80	52.36
CookingSense (Ours)	**68.30**	**56.77**	**65.40**	**64.38**	50.00	**60.97**

Table 6: Experimental results for *FoodBench*. The **bold** values indicate the highest scores within each benchmark dataset. All scores represent result accuracy.

DMIS Data Mining & Information Systems Laboratory   FNAI Lab   IMPERIAL   Sony AI

Results of FoodBench

Examples of CSQA with retrieved contexts from CookingSense

Source	Question	Retrieved Context
Web	If you double your recipe, what ingredient should you not double?	Recipe can be doubled but don't double the salt in the cooking water.
Recipe	"Soft Ball" Stage of Cooked Sugar occurs in which temperature range?	Barley Sugar Cook to 240F or soft-ball stage.
Paper	What can you use as a substitute for real sugar?	...alternatives to sugar with special consideration of xylitol.
Web	The forest in France whose oak trees are used to make barrels for aging wine is known as the:	The wine is then distilled and given to age in French Limousin oak barrels.
Web	Which of the following ingredients is not considered a major eight allergen?	Milk is considered one of the eight major food allergens by the FDA. Caution: nuts and peanuts are two of the top eight major food allergens.

Conclusion

- We have constructed the CookingSense, a large-scale KB that encompasses a comprehensive collection of culinary-domain assertions obtained from various data sources.