MIAIS: A Multimedia Recipe Dataset with Ingredient Annotation at Each Instructional Step

Yixin Zhang[1], Yoko Yamakata[2] and Keishi Tajima[1]

[1]Kyoto University [2]The University of Tokyo

Background

- In recent years, more and more datasets on recipes are constructed.

- Food-101 dataset, Recipe 1M+, UEC Food256:
  - mainly focus on **completed dish images** after cooking rather than images in recipe instructions.

- **Cookpad dataset:**
  - 1, 642, 450 images of completed dishes
  - 3, 105, 594 instructional images

source: https://cookpad.com/recipe/6100648

Problem and Purpose

- Most of the datasets:

  - No sufficient **multimedia information (text and image) on the cooking instructions** of the recipe.

- However, the **<u>instructional information</u>** is one of the most **important** and **unique** characteristics of recipe data.

- **MIAIS Dataset**

- *Goal*: construct a recipe dataset with sufficient **multimedia data** and the **annotations** to them for every cooking step (based on Cookpad dataset)

- *Applications*: cooking flow graph generation, recipe text generation, and cooking action recognition

MIAIS Dataset

- Multimedia recipe data

- 12,000 recipes in Japanese

- The text and image data are originated from <u>Cookpad</u>.

  - Recipe selection rule:

    - Recipes that have at least **five** cooking steps

    - Recipes that contain a text description and an image for **every step**

MIAIS Dataset

- 109,547 text-image paired instructional steps in total
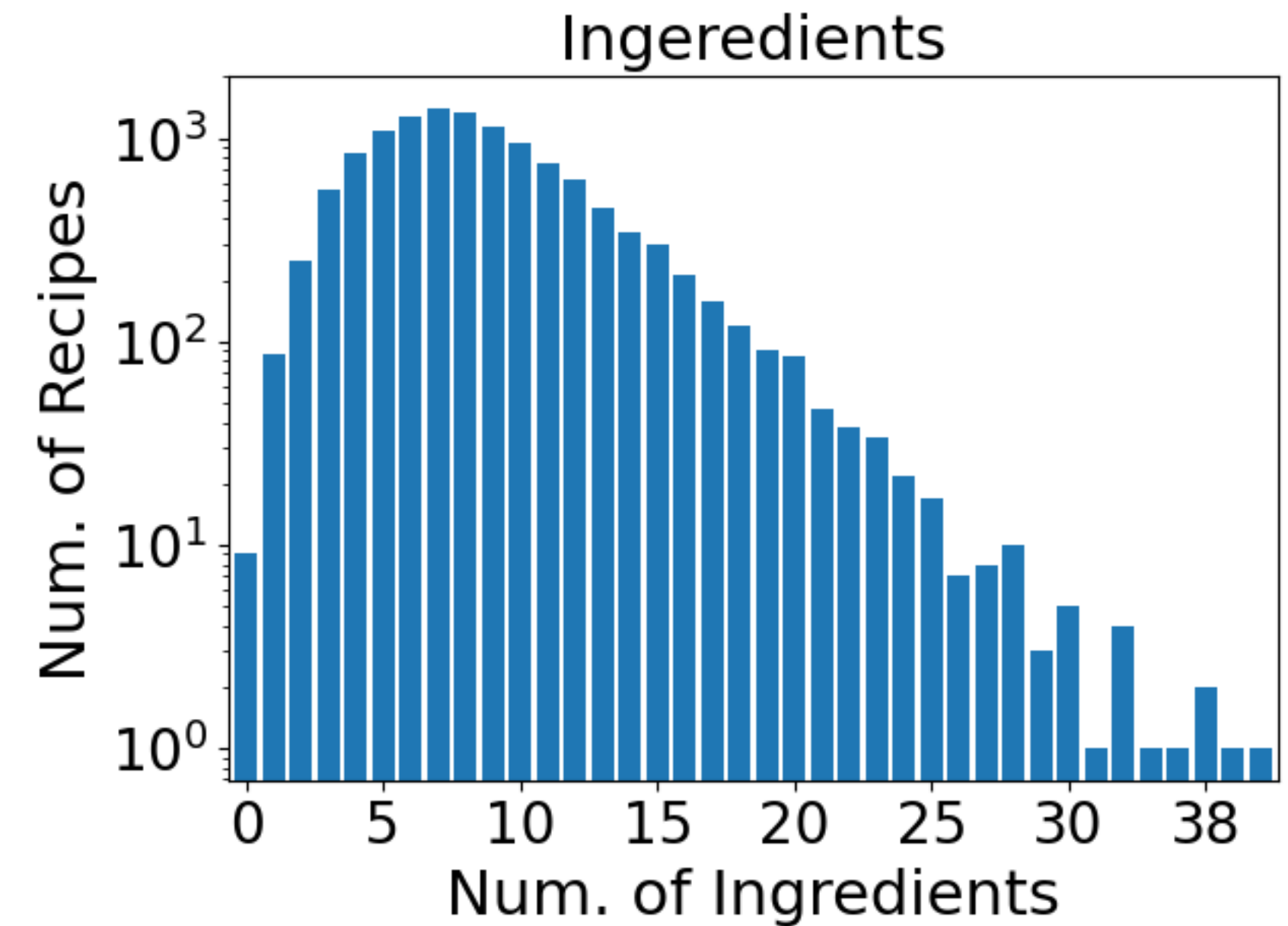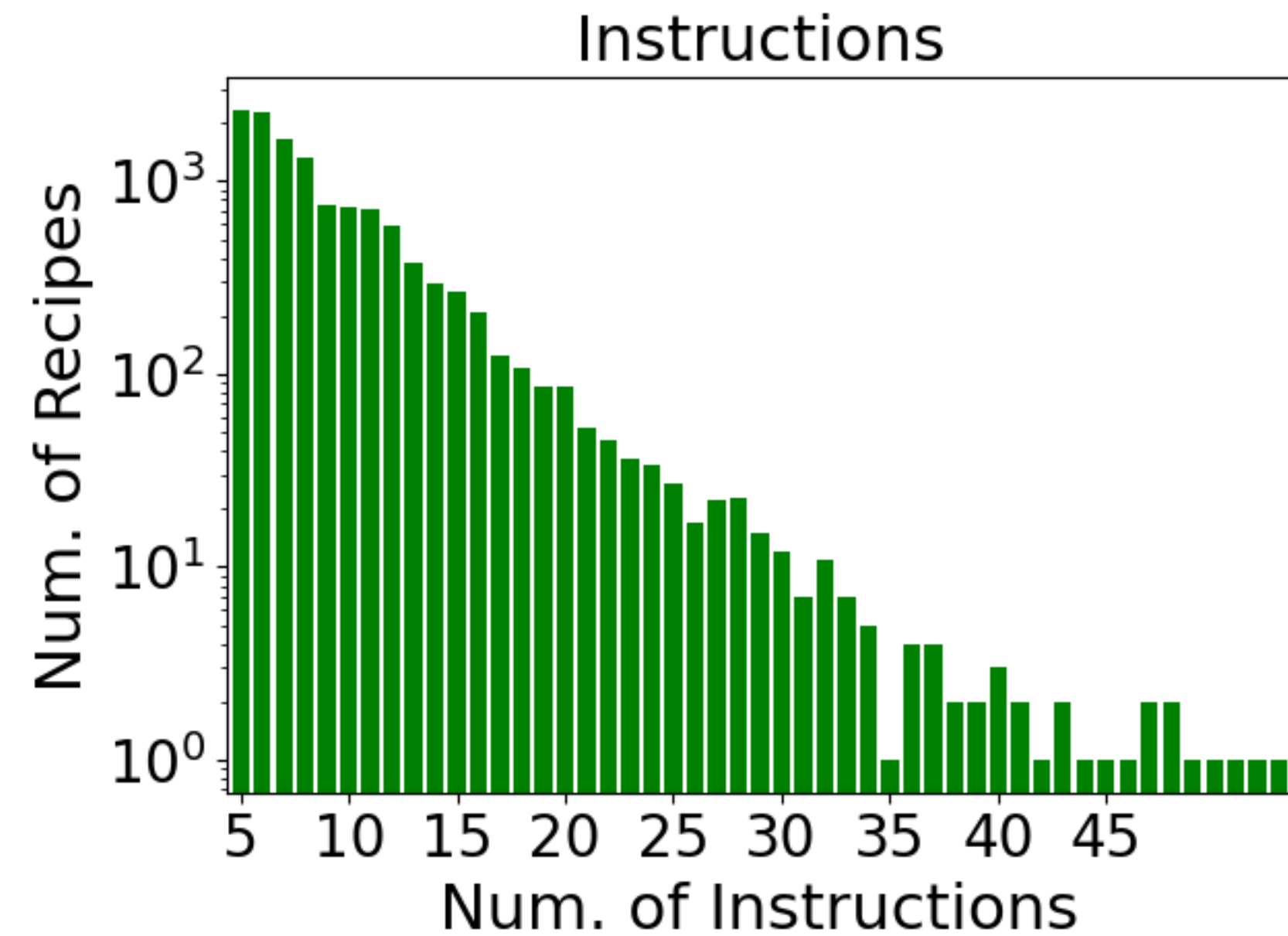
Cooking Flow Graph

- A **challenge** in understanding the cooking process from recipes:

  - present the complicated operations of cooking activities in **a succinct workflow.**

The original form sometimes makes it difficult to understand recipes directly and efficiently.

Recipe form: Sequential

Actual actions: sometimes parallel
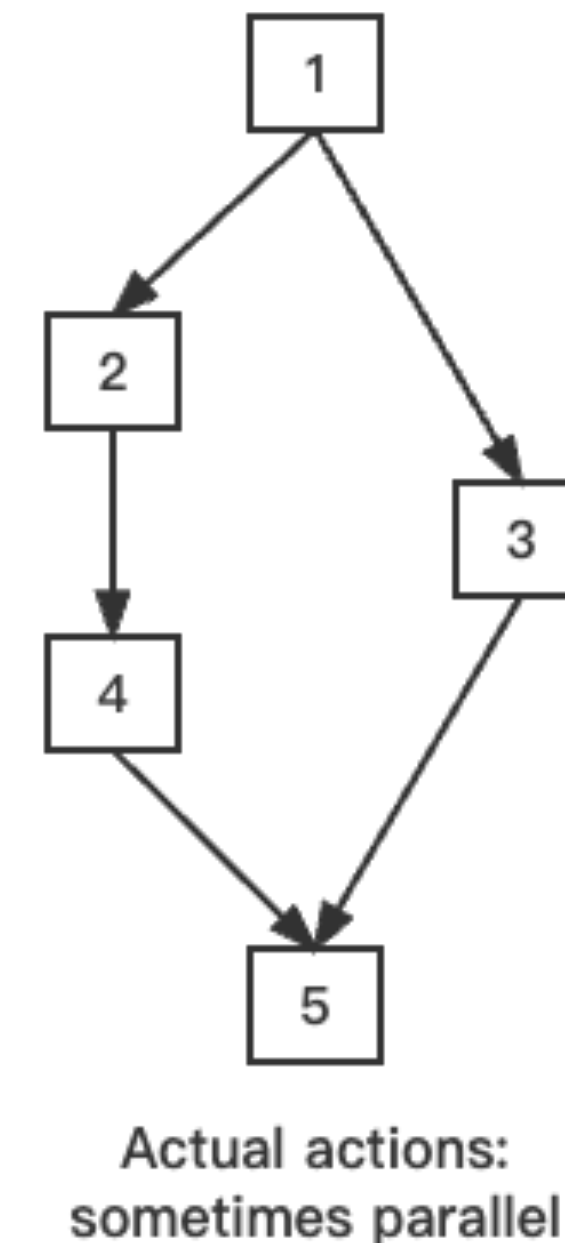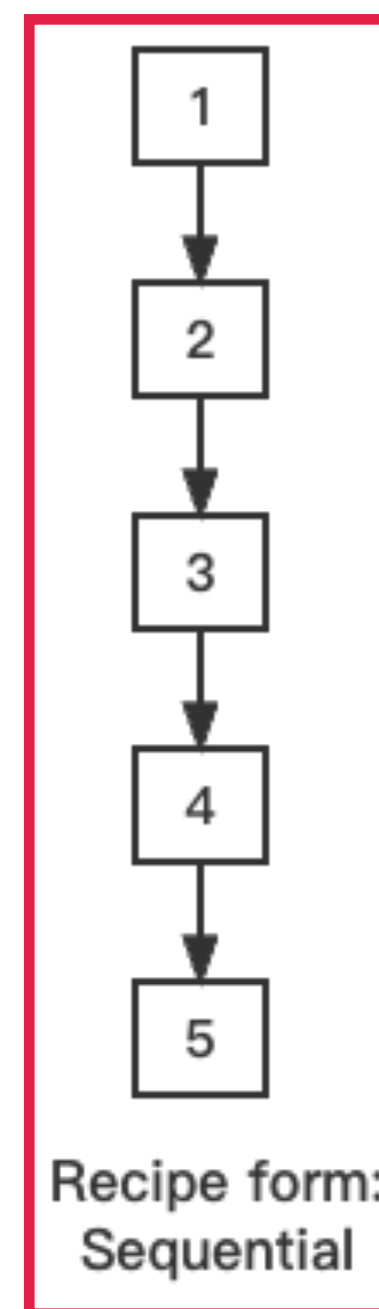
- A common solution to this issue is to generate a flow graph to help them understand the cooking process.

Cooking Flow Graph

- The flow graph is usually generated by **extracting the relationships between cooking actions and ingredients**.

- Most existing methods only focus on the **ingredients that appear in the instructional text**.

Quick and moist Nikujaga (meat and potatoes)

Ingredients
- potato
- water
- olive oil
- beef
- sugar
- salt
- Shirataki
- Mirin
- soy sauce
- honey

Cooking Instructions

1. Wash the potatoes with a bowl.

2. Roughly cut all the potatoes into quarters. After washing with water, add 4 tablespoons of water and cover over with a wrap.

3. Microwave for 6 minutes. Meanwhile, heat a frying pan and sauté the beef in olive oil.

4. After boiling, the potatoes are placed in a bowl of water and peeled them by hand.

5. Sprinkle 4 tablespoons of sugar on the beef. When blended, add a teaspoon of salt. Season the beef well.

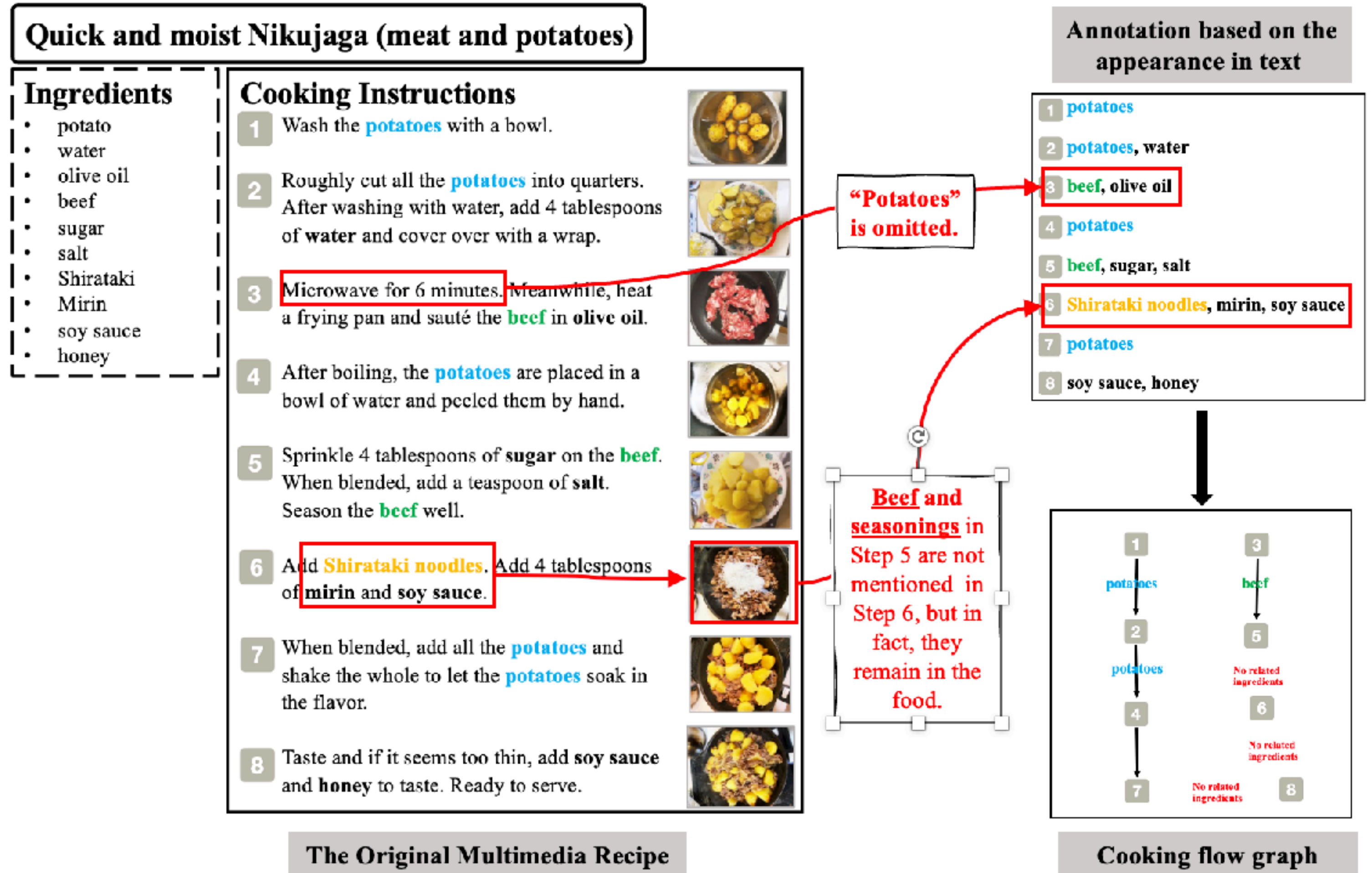6. Add Shirataki noodles. Add 4 tablespoons of mirin and soy sauce.

7. When blended, add all the potatoes and shake the whole to let the potatoes soak in the flavor.

8. Taste and if it seems too thin, add soy sauce and honey to taste. Ready to serve.

The Original Multimedia Recipe

"Potatoes" is omitted.

Beef and **seasonings** in Step 5 are not mentioned in Step 6, but in fact, they remain in the food.

Annotation based on the appearance in text

1. potatoes
2. potatoes, water
3. beef, olive oil
4. potatoes
5. beef, sugar, salt
6. Shirataki noodles, mirin, soy sauce
7. potatoes
8. soy sauce, honey

Cooking flow graph

7

Cooking Flow Graph

- To generate flow graph from multimedia recipes , one way is to focus on the **flow of ingredients** in recipes.

- Sufficient information is needed.

- However, the ingredient information in text **tends to be insufficient or omitted** when written by users.

Our Annotation Method

Quick and moist Nikujaga (meat and potatoes)

Ingredients
- potato
- water
- olive oil
- beef
- sugar
- salt
- Shirataki
- Mirin
- soy sauce
- honey

Cooking Instructions

1. Wash the potatoes with a bowl.

2. Roughly cut all the potatoes into quarters. After washing with water, add 4 tablespoons of water and cover over with a wrap.

3. Microwave for 6 minutes. Meanwhile, heat a frying pan and sauté the beef in olive oil.

4. After boiling, the potatoes are placed in a bowl of water and peeled them by hand.

5. Sprinkle 4 tablespoons of sugar on the beef. When blended, add a teaspoon of salt. Season the beef well.

6. Add Shirataki noodles. Add 4 tablespoons of mirin and soy sauce.

7. When blended, add all the potatoes and shake the whole to let the potatoes soak in the flavor.

8. Taste and if it seems too thin, add soy sauce and honey to taste. Ready to serve.

The Original Multimedia Recipe

Our proposed annotation method

1. potatoes
2. potatoes, water
3. potatoes, water, beef, olive oil
4. potatoes
5. beef, olive oil, sugar, salt
6. beef, olive oil, sugar, salt, Shirataki noodles, mirin, soy sauce
7. beef, olive oil, sugar, salt, Shirataki noodles, mirin, soy sauce, potatoes
8. beef, olive oil, sugar, salt, Shirataki noodles, mirin, soy sauce, potatoes, honey

**Our method complement those omitted information**

- Our new annotation rule:

  Each instructional step is annotated with

  all ingredients that physically exist in the intermediate products processed in the step

  no matter whether they are mentioned in the instructional text or not.

Flow Graph Generation

- Cooking flow graphs

  - One essential component of MIAIS, which is **derived from our annotation data**

- We <u>track the paths</u> of ingredients in order to generate the cooking flow graph.

- If a specific ingredient

  - is contained in a previous step;

  - can also be detected in one of the subsequent steps

- we can consider that **there is continuity between the two steps**.

- The path is created for each ingredient and the final flow graph is a union set of all the paths.

Flow Graph Generation

Our proposed annotation method
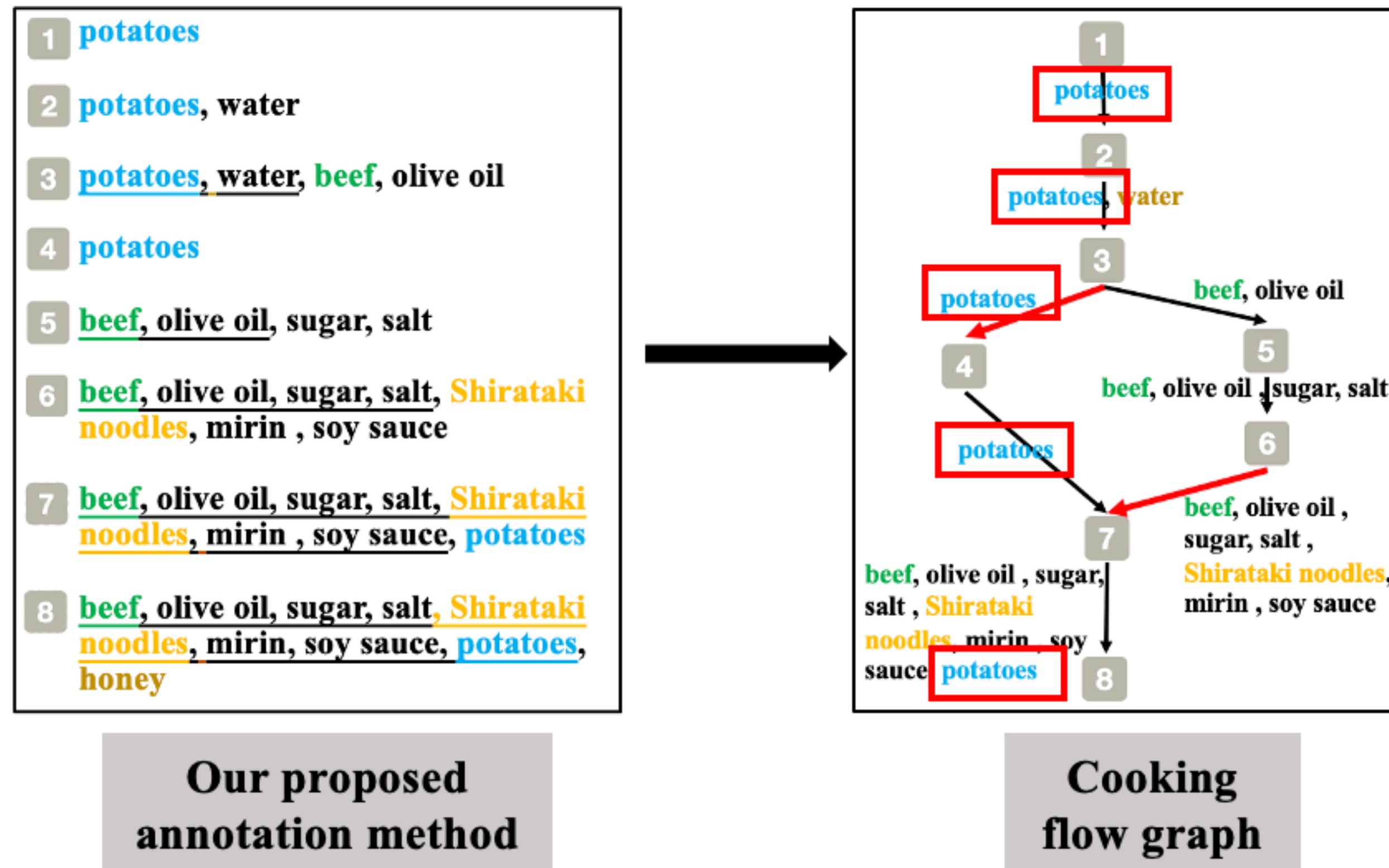
Cooking flow graph

- "Potato": 1-2-3-4-7-8

- "Beef": 3-5-6-7-8

- The flow graph of this recipe can be generated based on the above tracking information.

- The order in which the actual cooking is performed is not sequential, but with several branches.

Conclusion

- A multimedia recipe dataset with ingredient annotation for every instructional step

    - Containing **both text and image data for every cooking step**

    - Supplementary information about text-image pair of every step

- Cooking flow graphs

    - Directly derived based on the annotation information

    - These flow graphs can become the ground truth for the method of recipe flow graph generation

Thanks for listening!