**REPORT ABOUT DATASET RECIPENLG**

**1. Dataset format**

Cooking recipes have a specific format which consists of: a title, a list of ingredients with given amounts, and the instructions in a step by step format. There is some more data but not important.

- The title should accurately name and summarize the recipe content.

- The ingredients list has to contain entities consisting of the quantity, unit name, and ingredient name format.

- The instructions section needs to accurately present the order of steps (one step is one sentence).

Example format:

Recipe i-th {

title: “Cheeseburger Potato Soup”,

ingredients: [

“6 baked potato”,

“2/3 cup of butter or margarine”,

...

],

instructions: [

"Wash potatoes; prick several times with a fork",

"Microwave them with a wet paper towel covering ...”,

...

]

}

**2. Creating**

From Recipe1M+ do:

- Data cleansing:

+ Remove recipes without any ingredients or instructions.

+ Removed the excessivewhitespace characters.

+ Replaced unicode symbols, (e.g., fractions) with their ASCII equivalents.

+ Removed recipes identified by the same URL, the same sequence of characters in instructions and ingredients.

- Collecting more recipes.

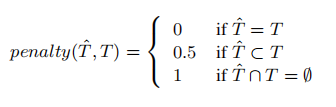
Result: The RecipeNLG dataset contains 2, 231, 142 distinct cooking recipes which is greatly expands the number of recipes available from Recipe1M+.

**3. Experiment**

a) Identifying food entities using NER

Training by a subset of 500 recipes was manually annotated food entities. This training data allowed authors to extract food entities from the rest of the dataset. In total, the chosen recipes contained about 2,400 individual ingredients.

The authors created the penalty metric to evaluate how precisely the model extracts a food entity (set of tokens Tˆ) from an ingredient, based on a test set (set of tokens T):



b) Generating recipes from food entities using pretrained GPT-2

Before traning, the authors decided to remove recipes with very short titles or instructions sections. They also removed recipes which contain phrases: “step” in instructions, to remove the possibility of cross-step references based on ordinal numbers, and ’mix all’, which lead the model to a preference of mixing everything over preparing detailed instructions. The model was given a set of food entities and ordered to generate full recipes.

The model was given a set of food entities and ordered to generate full recipes.

c) Evaluation

Select a set of 100 recipes that were not used in training, base on food entities of each record from the 100 recipes set mentioned, the author generate 10 recipe for each record from each models (1 trained by RecipeNLG, 1 trained by Recipe1M+). The result is 1000 recipes from each model.

They calculate cosine similarity to meansure the similarity between the generated recipe and the 100 recipes set. The results have shown that a RecipeNLG model generates recipes more similar with the original recipes (0.666 for RecipeNLG model and 0.589 for Recipe1M+ model average cosine similarity).

The authors used the LanguageCheck spell and grammar checker to calculate the amount of linguistic mistakes to estimate the overall performance of the model. They calculated the average number of errors per recipe. There were fewer errors in the RecipeNLG model (2.78) than in the Recipe1M+ (7.35).

The last approach to the evaluation was the utilization of translation metrics (BLEU, GLEU and WER).

