Tagset for COMER (Corpus of Meal Entity Recognition)

There are three main types of tags: ingredient, equipment and recipe.

- Ingredient (IN):

The ingredients used for the recipes are tagged as such. They are usually named at the beginning of the recipe, in the "ingredients" section, and then repeated multiple times throughout the body.

The second part of the ingredient is the type of ingredient. There are 22 tags and it is mandatory; no ingredients can be tagged without its type. This list has been adapted as the corpora has been tagged. For example, sugar was not considered a category at the beginning of the project, but we thought it could be a useful ingredient to tag, as there are health conditions like diabetes related to the consumption of sugar.

- Fats and oils (FO): Vegetable oils like olive or sunflower oil.
- Nut (N) : Dried fruits like hazelnuts or pistachios.
- Milk Product (M): Any ingredient that is either animal milk or derived from it, like yogurt or butter.
- Pulse (P): Edible seed that are usually encased in pods like lentils, chickpeas, beans, ...
- Mushroom (MR): All types of edible fungi such as enoki, oyster, shiitake, ...
- Fruits and vegetables (FV): A broad category that includes any ingredient that is considered either a fruit, like apple or pear, or a vegetable, like lettuce or artichoke.
- Dough (D): a mixture of a liquid (water, milk, etc.) with flour, meal or powdery ingredients.
- Grain (G): Any cultivated cereal used as food such as wheat, oat, quinoa, …
- Pasta (PT): Pasta that is used as an ingredient, like macaroni or spaghetti.
- Rice (R): All types of rice (bomba, jasmine, long, …).
- Baking product (BP): All sorts of flours and yeast.
- Fish (F): Examples of this category include mackerel, codfish, salmon, …
- Shellfish (S): Examples of this category include clam, lobster, shrimp, …
- Meat (ME): Examples of this category include chicken, lamb, beef, …
- Alcoholic beverage (AB): Drink with alcoholic content such as wine, beer, champagne, …
- Non-alcoholic beverage (NAB): Drinks without alcoholic content such as juice, vegetable milks, sodas, …
- Egg (E): This category includes different types of eggs such as chicken egg, ostrich egg and quail egg.
- Herb and Spice (HS): Different types of aromatic herbs used to add flavour to food like oregano, thyme, cumin, …
- Sugar (SU): natural sweetener that usually comes from plants and is used for cooking in a white crystalline form.
- Sweetener (SW): Any sweetener that isn't sugar, like saccharin, agave syrup, stevia, …
- Other (O): Any ingredient that can't be categorized as any of the other types.

Optionally, ingredients can have a third component if they contain any of the following lists' allergens. Since this is dependent on the type of the ingredient (if an ingredient has been categorized as "egg" it will have "egg" as an allergen, but never "milk product"), a list of the possible combinations of tags was created, which is shown in Table 1. This list of 14 allergens is based on those that are required to be declared as allergens according to the EU 1169/2011 Regulation on the provision of food information to consumers (or FIC Regulation).

- Milk product (M)
- Gluten (G)
- Mollusc (MO)
- Egg (E)
- Crustaceans (CR)
- Fish (F)
- Peanut (P)
- Soya (S)
- Nut (N)
- Celery (C)
- Mustard (MU)
- Sesame seed (SS)
- Sulfur dioxide (SD)
- Lupin (L)

- Equipment (E):

The nouns that refer to the tools used for cooking receive the tag "equipment". There are two subtypes of equipments, thatthaet must always be categorized:

- Utensil (U): Equipment that does not need electricity to function. This category includes cutlery, dishes, trays,...
- Appliance (A): Cooking devices that are powered by electricity, such as an oven or a microwave.

- Recipe (R):

This tag is used for the name of recipes that are normally mentioned multiple times throughout the body of the recipe. Recipe names will usually contain names of ingredients, but they will be tagged as "recipe" since they are being named as such. Entities tagged as "recipe" have no further specifications, since it is a very general tag.

Tag	Meaning
IN-FO	Ingredient-Fats and Oils
IN-N	Ingredient-Nut
IN-N-AN	Ingredient-Nut-Allergen Nut
IN-N-AP	Ingredient-Nut-Allergent Peanut
IN-M-AM	Ingredient-Milk Product-Allergen Milk Product
IN-P	Ingredient-Pulse
IN-P-AS	Ingredient-Pulse-Allergen Soy
IN-P-AL	Ingredient-Pulse-Allergen Lupin
IN-MR	Ingredient-Mushroom
IN-FV	Ingredient-Fruits and Vegetables
IN-FV-AC	Ingredient-Fruits and Vegetables-Allergen Celery
IN-D	Ingredient-Dough
IN-D-AG	Ingredient-Dough-Allergen Gluten
IN-D-AE	Ingredient-Dough-Allergen Egg
IN-D-AM	Ingredient-Dough-Allergen Milk Product
IN-G	Ingredient-Grain
IN-G-AG	Ingredient-Grain-Allergen Gluten
IN-G-ASS	Ingredient-Grain-Allergen Sesame Seed
IN-PT-AG	Ingredient-Pasta-Allergen Gluten
IN-R	Ingredient-Rice
IN-BP-AG	Ingredient-Baking Product-Allergen Gluten
IN-F-AF	Ingredient-Fish-Allergen Fish
IN-S-AMO	Ingredient-Shellfish-Allergen Mollusc
IN-S-ACR	Ingredient-Shellfish-Allergen Crustaceans
IN-ME	Ingredient-Meat
IN-AB	Ingredient-Alcoholic Beverage
IN-AB-ASD	Ingredient-Alcoholic Beverage-Allergen Sulphur Dioxide
IN-NAB	Ingredient-Non Alcoholic Beverage

IN-E-AE	Ingredient-Egg-Allergen Egg
IN-HS	Ingredient-Herb and Spice
IN-SU	Ingredient-Sugar
IN-SW	Ingredient-Sweetener
IN-O	Ingredient-Others
IN-O-ASD	Ingredient-Others-Allergen Sulphur
E-U	Equipment-Utensil
E-A	Equipment-Appliance
R	Recipe

Table 1: Final list of the possible combinations of tags.

* All these tags can begin with "B-" or" I-", there is also an "O" tag.