Related datasets proposal and integration example

Speaking about cuisine and recipes, it is not difficult to generalize and add or integrate data about it. Starting from adding data, a similar process of web scraping can be enough to increase the number of recipes in the system. As long as main characteristics are present, like a title and a list of ingredients, the main features of the system are preserved. Obviously, some secondary queries like those about nutritional information could be limited whenever some recipes cannot be filtered among these values, but this is already happening with the provided datasets. A concrete example is “ambitiouskitchen.com”, a similar website to the ones that has been used so far. The structure is compatible for the searching of urls and within the recipes, the most important fields and values are there. Adapting the provided scripts and appending (or merging) the results to the existing data after some cleaning should be enough.

The integration with smart fridges is neither required in this project nor feasible since it is a fresh technology and still a niche. Nevertheless, adapting the data structure from an eventual smart fridge to the one of the ingredients table, could already be enough to store the quantity, the remaining quantity and the expire date of ingredients. The technology to actually store items in a smart fridge is left to the relative producer. As for the recipes part, if new ingredients need to get inserted to the system, it should be enough to adapt the data structure of those to the already existing data provided in this project.

A very important information that is not present at all in the source data is the cuisine type (French, Japanese, thai, …). A useful way to get this information is provided now. It is possible to predict the cuisine type given the ingredients. To do that, a training set was downloaded from Kaggle to train a model and apply it on the ingredients table. After auto-modelling with naïve bayes in rapidminer, a prediction model was created with around 77% of accuracy on the training set itself. The ingredient table got adapted to the training dataset and then the prediction algorithm got started. It is not possible to have an accuracy of the results, but checking manually a couple of predictions, it seems to work quite well. This procedure is explained more precisely now.

Bla blab la blab la, model, bla bla bla, adapting, bla bla bla, predicting, bla bla bla.