

Proposers: Alison O'Shea & Brian Murphy

Contact: Alison.OShea@mtu.ie & BrianM.Murphy@mtu.ie

Project Title:

Automatically identifying allergens in ingredient lists

Research Question/ Problem Addressed:

Can text processing be used to automatically identify allergens from ingredient lists?

Project Abstract:

In the Biological Sciences Department within MTU, there is a large amount of nutritional data collected on consumer foods. This includes images of food brands for labelling and data logging purposes. These labels contain valuable information, which includes potential allergen ingredients. Allergens are often identified by marking text using bold or underline, alternatively foods are marked as being prepared or packed in facilities where allergies are present.

The aim of this project is to develop an allergen dataset which includes the different types of allergens and their variations. Following this, a text matching system can be developed to identify and list the different allergens that are contained within store bought products.

Typical tasks a student will carry out include:

- Research of the different text recognition/matching processes
- Investigate national and international guidelines around allergen identification and labelling
- Creation of allergen dataset
- Creation of simulated product ingredient lists
- Result logging
- Testing and evaluation
- Document work completed
- Weekly team meetings
- Work as part of a multi-disciplinary team of computer scientists, engineers, and nutritional scientists

Technologies Utilised:

Students choice but example Python (Numpy, Scikit-Learn, Scipy, Pandas)