Computational Gastronomy

Assignment 1

You may use Python and Jupiter Notebook to complete the assignments and documentation.

Notes: You are responsible for backing up the data and results, which will be used for evaluation.

Follow the rubric diligently while submitting. Name the files with the question numbers.

Answers to questions numbers 4—8 are to be presented as a single scanned PDF file from handwritten answers.

---

1. Complete the following analysis using the recipes' data.
   (a) Scrape (using libraries such as BeautifulSoup) any **10,000 recipes**. **Submit the raw data.** **[5]**
   Include recipe titles, ingredient phrases, cooking instructions, and other relevant details.
   (b) Write a script to extract information about the 'name of the ingredients' from the ingredients section using Named Entity Recognition. **[8]**
   (c) **Store recipes in the form of a (Recipe ID)—(Ingredient Name) form**. **[2]**

   NOTE: You must not scrape/use data from RecipeDB

2. Analyze the data obtained for the following.
   (a) Find the number of unique ingredients. List them with their frequencies. **Submit the file.** **[2]**
   (b) Plot the recipe size distribution for these recipes and the average size of the recipes (s).
   Properly label the axes. **Submit the file.** **[3]**
   (c) Plot cumulative distribution of recipe size (label axes properly). **Submit the file.** **[5]**

3. For the data of recipes obtained in the above question (1):
   (a) Plot the frequency-rank distribution. Scale and label axes properly. **Submit the file.** **[5]**

4. **Compare** the nutritional profile of a 'boiled egg' versus that of a 'boiled rice and *daal*'. **[2]**

5. List **five** most uncommon food ingredients that you (or your parents or grandparents) know of. List their (a) Common/Vernacular/Local Name, (b) English Name (if available), (c) Seasonal Use (if known), and (d) Nutritional Values (if available). **[5]**

6. Consider the following statements and **provide scientific arguments** about them being a scientific truth or a myth. **[3]**

   (a) Cooking food in microwave destroys its 'nutritional value'.
   (b) Refrigerating food destroys its 'nutritional value'.
   (c) Genetic modifications in plans or animals are 'bad'.

7. **Write a brief (5-8 lines) summary** on the technique used for finding out the calorific content of food products. **[2]**

8. **List eight specific technologies/products, with 1-3 lines of description for each** that you envision could emerge from the application of the Computational Gastronomy paradigm. Be as specific/narrow as possible. **[8]**