# Task 3: Dish Recognition

Xiaoming Ji (xj9)

# Overview

The goal of this task is to mine the data set to discover the common/popular dishes of a particular cuisine. Typically, when you go to try a new cuisine, you don’t know beforehand the types of dishes that are available for that cuisine. For this task, we would like to identify the dishes that are available for a cuisine by building a dish recognizer.

For comparison, two approaches: [SegPhrase](https://github.com/shangjingbo1226/SegPhrase) and [AutoPhrase](https://github.com/shangjingbo1226/AutoPhrase) are used to extract dishes from **Chinese** cuisine. SegPhrase and AutoPhrase are algorithms developed by UIUC and achieved SOTA performance on mining quality phrases. SegPhrase needs a small amount of manually labelled samples. While AutoPhrase reduced human effort by leveraging existing general knowledge bases.

We can see both approaches give us a very good results but AutoPhrase performs better in term of finding new phrases with less effort.

# Data Preparation

Cuisine corpus

Labeling

Knowledge Base

# Phrase Mining

SegPhrase

AutoPhrase

# Discussion

<https://en.wikipedia.org/wiki/List_of_Chinese_dishes>