RECIPE SUGGESTION TOOL

CS298 PROGRESS REPORT
FALL 2010

BY

SAKUNTALA GANGARAJU

DECEMBER 2010

Contents

1. Introduction

Search engines have made access to information easy. One only needs to get connected to the Internet to get the information one needs. When searching for cooking recipes, sometimes user may prefer to search based on ingredients. It will be more helpful if the search also suggests similar recipes.

In my project, Recipe Suggestion Tool, I am planning to implement the above-suggested features. My aim is to suggest clusters of recipes based on the user query. The query can be a search for recipes or for recipes based on an ingredient. For instance, if a user searches for 'cake', my aim is display different clusters of cake recipes, like strawberry cakes, chocolate cakes etc. Similarly, It displays results for searches based on ingredients.

In my project, I am using Yioop Search Engine developed by my Project Advisor, Dr.Chris Pollett, for crawling and displaying the results. I am using PHP for coding and MySQL for managing the database. I have implemented Clustering through Minimum Spanning Tree using Kruskal's algorithm, which I will be using for clustering the recipes.

2. Progress Made In Fall 2010

In Fall 2010, I started by modifying Yioop Crawler so that when it crawls a recipe page, it extracts the ingredients from the page. Using XPath, the crawler checks whether the page has 'Ingredients' tag in it. If it has, the crawler extracts the ingredients from that page.

The ingredients list extracted has adjectives and adverbs along with the main ingredient. My next step was to extract the main ingredient from the ingredients list. I used regular expressions to extract the main ingredient.

To Cluster the recipes, ingredients are represented as vectors and dot product of their vectors is calculated. This product is used to represent the distance between the recipes when represented as a spanning tree. Clustering is done after calculating the Minimum Spanning Tree using Kruskal's algorithm.

Yioop Crawler includes a feature to add user-defined meta-words. Here meta-word means a word which was not in a downloaded document, but which is added to the inverted-index as if it had been in the document. I am using this feature to add meta-word 'recipe' to the recipe pages.

3. To do In Spring 2011

In Spring 2011, I will create a new Doc Id for recipe pages and attach the meta-word 'recipe' to the pages. Using this meta-word, I will extract the recipes from the web archives and perform the clustering of the recipes. The clusters of the recipes are saved in web archives. I will also improve the extraction of main ingredient from the ingredient-list by implementing recursive descent parsing.