

소프트웨어 응용

프로젝트 최종 보고서

8조

컴퓨터과학부 2017920019 김현창

컴퓨터과학부 2017920026 박영빈

컴퓨터과학부 2017920055 임철수

1. 개요

요리 레시피들을 그래프를 통해 데이터를 분석하여 레시피들간의 연관성을 파악하고, 기존 레시피 사이트에서는 제공하지 않는 새로운 기능을 제공하는 것을 목적으로 프로젝트를 진행하였다.

2. 데이터 셋

프로젝트 데이터 셋으로는 '메뉴판' 사이트의 대략 1,600개의 레시피를 선택하였다.

기존에 사용하려 했던 데이터 셋은 '만개의 레시피' 사이트였다. 하지만 만개의 레시피 사이트에는 몇 가지 문제점이 있었다. 첫 번째로는 레시피들의 분류가 거의 되어 있지 않아 데이터를 분석하기에는 적합하지 않았다는 점이다. 이런 점에서 비교하였을 때, '메뉴판' 사이트는 방법, 국가, 종류 등의 분류가 잘 되어있고, 재료 또한 주재료, 부재료, 양념으로 세분화되어 설명하고 있었다.

두 번째 문제는 '만개의 레시피'에서 제공하는 레시피들의 이름이었다. 만개의 레시피에서는 레시피의 이름에 잡다한 미사어구가 많이 들어가있어 데이터를 분석하는 데 방해가 주었다. 하지만 '메뉴판'에서 제공하는 레시피는 이름이 깔끔하게 되어있어 우리 프로젝트에 더 적합하였다.

3. Edge 연결 알고리즘

기존에 생각했던 Edge 연결 알고리즘은 레시피의 재료에서 1~2개만 추가하여 다른 레시피의 재료를 충족하면 연결하는 알고리즘이었다. 하지만 이 알고리즘에서는 필요한 재료가 많은 레시피만 연결될 확률이 높다는 문제가 있었다. 그래서 재료 간의 비율을 계산하는 방식으로 Edge를 연결하였다.

또한 주재료, 부재료, 양념 등 재료들이 세분화되어 적혀있는 메뉴판 사이트의 특성을 살리는 알고리즘을 사용했다. 주재료처럼 중요한 재료는 겹쳤을 경우, 부재료나 양념보다 Edge 연결에 더 영향력을 끼친다. 마찬가지로, 부재료는 양념보다 더 영향력을 끼치는 방식으로 Edge가 연결되는 알고리즘을 사용하였다.

사용한 알고리즘은 다음과 같다. 이렇게 계산한 weight값이 일정 값(α) 이상이면 Edge를 연결하였다. 이 프로젝트에서는 $\alpha=1.5$ 로 Edge를 연결하였다.

각 레시피들끼리:

weight \leftarrow 0

mainRatio $\leftarrow \frac{\text{len}(\text{레시피1의 주재료} \cap \text{레시피2의 주재료})}{\text{len}(\text{레시피1의 주재료} \cup \text{레시피2의 주재료})}$

weight \leftarrow weight + $e^{\text{mainRatio}}$

subRatio $\leftarrow \frac{\text{len}(\text{레시피1의 부재료} \cap \text{레시피2의 부재료})}{\text{len}(\text{레시피1의 부재료} \cup \text{레시피2의 부재료})}$

weight \leftarrow weight + subRatio

sourceRatio $\leftarrow \frac{\text{len}(\text{레시피1의 양념} \cap \text{레시피2의 양념})}{\text{len}(\text{레시피1의 양념} \cup \text{레시피2의 양념})}$

weight \leftarrow weight + (sourceRatio/2)

if weight $> \alpha$: ($\alpha = 1.5$)

edge 연결

4. 시각화

시각화, 분석을 위한 도구로는 Gephi를 사용하였다.

Gephi에서 그래프를 시각화하는 알고리즘으로는 ForceAtlas2 알고리즘을 사용하였다. ForceAtlas2는 강하게 연결된 노드를 함께 끌어오고, 약하게 연결된 노드를 분리하여 시각화하는 알고리즘이다.

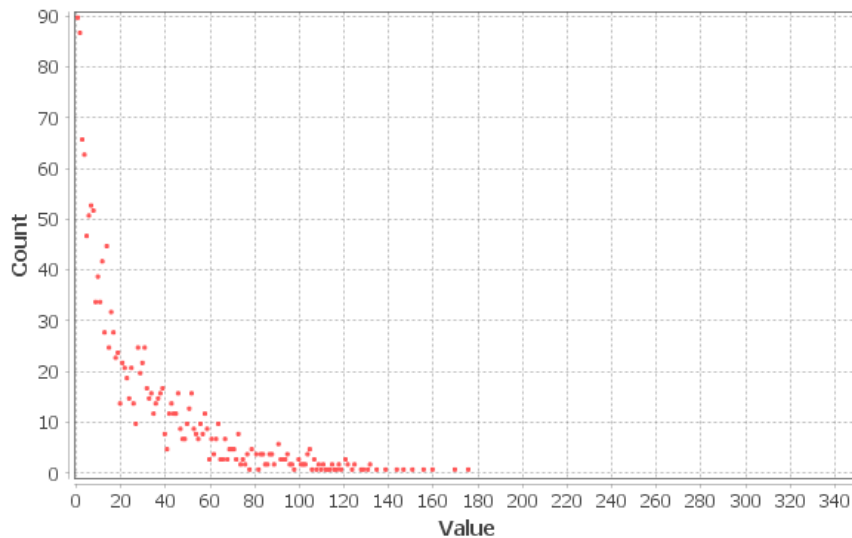


Figure 1. Degree

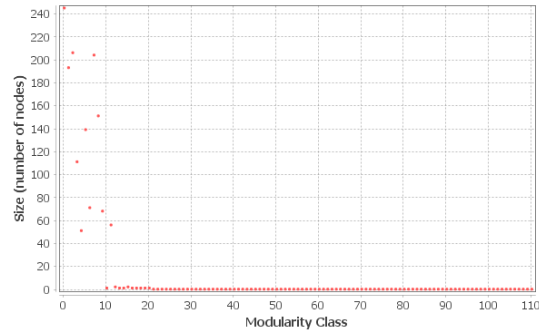
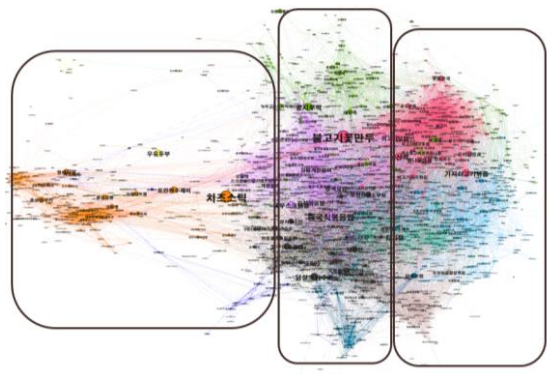


Figure 2. Modularity Class, Modularity : 0.559

시각화하였을 때 Degree와 Modularity Class는 위의 그래프처럼 나타났다. Communities는 111개로 나타났지만, Size가 1~2인 Community, 즉 다른 레시피와 연결된 것이 거의 없는 노드를 제외하면 실질적인 클러스터는 약 10개 정도 존재하는 것을 확인할 수 있다.

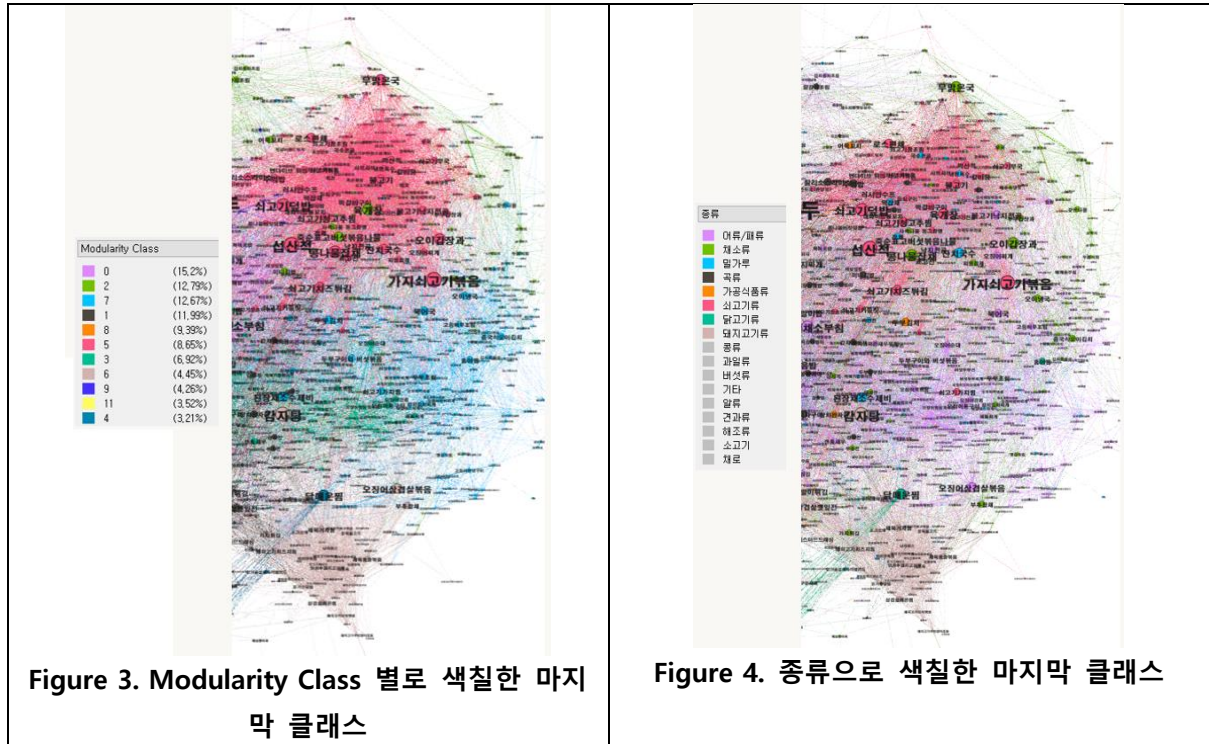
5. 데이터 분석



그래프는 크게 세 가지 클래스로 나누어 살펴볼 수 있다. 왼쪽 클래스는 대부분 서양식과 이탈리아식이 분포하고 있었다. 이 클래스에는 '설탕, 버터, 계란'이 재료로 많이 사용되었고, 재료에서도 유추할 수 있듯이 대부분 빵, 과자 레시피들이 분포하고 있었다. 레시피들의 종류도 역시 밀가루 종류가 대부분을 차지하고 있었다.

가운데 클래스는 다양한 클러스터들이 섞여 있는 클래스이다. 이 부분은 재료나, 방법, 국가 등으로 나누어 살펴봐도 뚜렷하게 영역이 나누어져 있지 않고 레시피들이 얹혀 있는 것을 확인할 수 있다. 퓨전, 한식, 서양식 등 다양한 레시피가 섞여있는 영역인 만큼 여러 종류의 음식들이 종류와 방법을 가리지 않고 섞여 있다고 판단할 수 있다.

마지막 클래스는 대부분 한식으로 구성된 클래스였다. 마지막 클래스의 색깔을 방법별로 구분하여 분석한 결과, 어떠한 뚜렷한 특징은 나타나지 않았다. 우리나라 사이트인 만큼 볶음, 찌개, 찜, 부침, 국 등 한식에 최적화된 다양한 분류들이 존재하였다. 이 클래스를 종류별로 색깔을 칠하여 분석한 결과 채소류 클러스터, 쇠고기류 클러스터, 돼지고기류 클러스터, 어류/패류 클러스터가 존재하는 것을 확인할 수 있었다. 이를 통해 한식에 다양한 재료가 사용된다는 것을 알 수 있다.

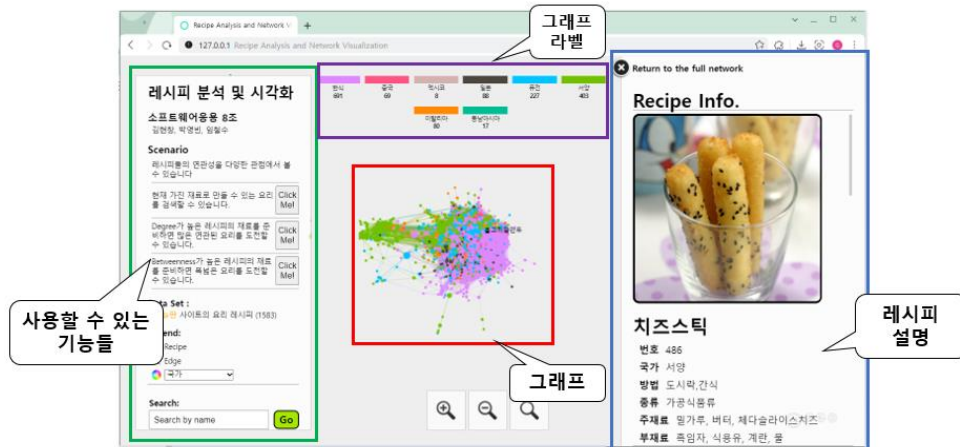


6. 어플리케이션

Gephi에서 제공하는 Sigma Exporter 플러그 인을 사용하여 제작하였다.

Sigma Exporter는 분석한 그래프를 웹 브라우저에서 볼 수 있도록 Export해주는 플러그 인이다. Export한 후에, 우리가 구현하고자 하는 기능들을 그 위에 새롭게 구현하였다.

그래프 영역은 분석한 그래프가 표시되는 영역이고, 화면에 나타나는 노드들의 라벨은 위에 표기가 된다. 사용자가 기능을 사용하기 위한 리모콘은 왼쪽에 나타나고, 현재 선택한 노드(레시피)에 대한 설명은 오른쪽 패널에 나타난다.



7. 시나리오

해당 프로젝트에서 존재할 수 있는 시나리오는 4가지를 준비하였다.

1. 레시피 이름이나 재료로 검색하여 레시피를 찾을 수 있다.
2. 레시피들의 연관성을 다양한 관점에서 볼 수 있다.
3. Degree가 높은 레시피의 재료를 준비하면 다른 많은 레시피를 도전할 수 있다.
4. Betweenness가 높은 레시피의 재료를 준비하면 폭 넓은 다른 레시피를 도전할 수 있다.

1번 기능은 대부분의 레시피 사이트에서도 제공하는 기능이다. 해당 프로젝트에서도 입력한 키워드로 레시피의 이름이나 재료들을 검색하여 찾는 기능을 제공하고 있다.

2번~4번 기능이 다른 레시피 사이트에서는 제공하지 않는 해당 프로젝트만의 새로운 기능이라고 할 수 있다. 2번 기능은 레시피들의 관계를 분석한 그래프를 간단하게 웹 페이지에서 사용자가 직접 보며 레시피들의 연관 관계를 파악해볼 수 있는 기능이다. 레시피를 클릭하면, 레시피에 대한 설명이 패널에 나타나면서, 선택한 레시피와 연결된 관계가 그려진 그래프가 그래프 영역에 업데이트되고, 연결된 노드(레시피)들이 무엇인지 클릭하여 계속 이어갈 수 있다.

3번 기능, 4번 기능은 Degree, Betweenness를 이용한 기능이다. Degree가 높은 노드는 연결된 다른 노드가 많다 즉, 재료가 비슷하게 연관된 다른 레시피가 많다는 의미이다. 사용자가 Degree가 높은 레시피의 재료를 준비하여 요리를 한 이후에 해당 재료에서 약간의 다른 재료만 추가한다면 많은 다른 레시피를 도전할 수 있다.

Betweenness가 높은 노드는 서로 다른 클러스터링을 연결하는 다리 역할을 하는 레시피이다. 사용자가 Betweenness가 높은 레시피를 선택하고 재료를 준비하여 요리를 한다면, 이후에는 약간의 다른 재료를 추가하면 폭 넓은 레시피들을 도전할 수 있다. 예를 들어 Betweenness가 높은 '불고기꽃만두'는 '서양' 레시피가 밀집된 클러스터와 '한식' 레시피가 밀집된 클러스터를 연결하는 역할을 하고 있다. 사용자가 퓨전 요리인 '불고기꽃만두'의 재료를 준비하여 요리를 하면, 이후에는 약간의 재료 추가만으로 '서양' 레시피, '한식' 레시피 등 도전할 수 있는 폭이 넓어진다는 것을 의미한다.

8. 사용 예시

사용 예시는 보고서에 첨부하지 않음.