

그래프 플로팅을 통한 소스(Sauce)내 원재료 데이터 시각화

박건우, 최진서, 주승세, 신학종, 조성균, 최원규*
한국전자통신연구원

geonwoo0992@etri.re.kr, chojinseo@etri.re.kr, wntmdtp01@etri.re.kr, hakjong@etri.re.kr,
skjo@etri.re.kr, *wkchoi@etri.re.kr

Visualizing Raw Material Data in Sauce through Graph Plotting

Park Geon Woo, Choi Jin Seo, Joo Seung Sae,
Shin Hak Jong, Jo Seng Kyoun, Choi Won Kyu*

Electronics and Telecommunications Research Institute

요약

본 논문에서는 식품 소스(Sauce) 내 원재료의 상관관계 및 특징 추출을 위해 데이터의 시각화 방안에 대해 탐구하였다. 시중 제품 및 일반 레시피 데이터를 추출하고 그래프 플로팅을 통해 시각화 하였으며 노드의 크기와 위치, 엣지의 길이와 굵기로 원재료의 사용 빈도 및 함량을 표현하였다. 이를 통해 제품 및 일반 레시피들 내 원재료들의 사용 경향을 파악하고 맛의 구성을 파악하였다. 본 연구에서는 향후 사용자 선호도에 따른 소스 배합 및 개발 예측에 활용 가능성을 제시한다.

I. 서론

음식은 이제 단순히 포만감을 채우고 살아가기 위해서 섭취하는 것이 아닌, 미식과 건강을 중시하는 하나의 문화로 진화하고 있다. 특히 최근에는 대체당을 활용한 저당 및 저칼로리 제품들에 대한 관심이 높아지면서 기업들은 소비자들의 니즈를 충족시키려는 제품을 출시하기 위해 노력중이다. 그러나 모든 소비자들을 만족시키기에는 어려운 법이며, 사람마다 기호가 다르기 때문에 모든 수요를 충족시키기에는 어려움이 있다. 이 때문에 대중들이 요리를 할 때 각종 매체에 소개된 레시피에서 소스 제작 시 자신의 기호에 맞게 재료의 함량을 직접 조절하는 경우도 있다. 하지만 재료들을 조합하며 원하는 맛을 찾는 것은 전문지식이 부족한 일반 대중들에게는 어려운 일이다. 따라서 우선 어떠한 재료들이 사용되고 있는지, 그리고 얼마나 사용되는지 파악하는 것이 중요하다.

그리하여 대중들이 직접 업로드한 레시피에서 데이터들을 추출하였고, 이 결과를 바탕으로 시중에 판매되고 있는 제품들을 대상으로 원재료 데이터를 추출하였다. 효과적인 정보 제공을 위해 추출한 데이터들의 시각화를 진행하였으며, 이때 노드와 엣지의 시각화 및 공간화(layouting), label adjust 등의 기능이 포함되어 있는 Gephi를 사용하였다.[1]

II. 본론

1. 일반 레시피 대상 데이터 추출

농식품 빅데이터 거래소(KADX)에 업로드 된 일반 대중들이 작성한 ‘갈비’ 항목 레시피를 선택하여, 그 중 양념 레시피만을 추출하였다. 약 10 만 여개의 레시피 중 시판 양념소스를 사용하지 않고, 원재료의 함량 및

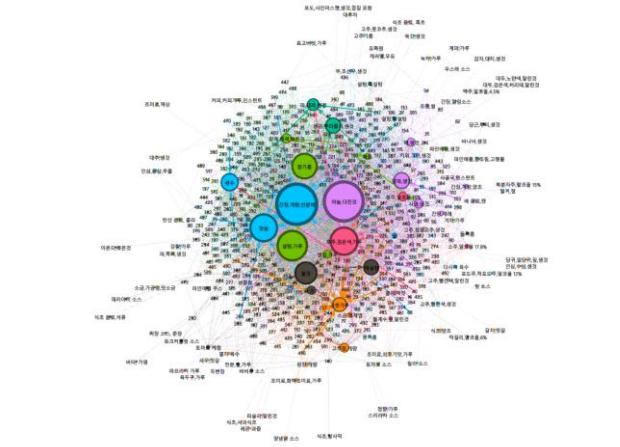
단위가 기록되어 있는, 육류가 사용된 ‘찜’과 ‘구이류’의 총 529 가지의 레시피 데이터를 추출하였다. 추출한 원재료 데이터들을 국가표준식품성분표에 따라 다시 분류하였으며, 각각의 재료들의 함량을 백분율로 나타내어 각각 가중치를 부여하였다. 이후 Gephi 내 layout 인 ForceAtlas2 를 적용하여 그래프 플로팅을 진행하였다.[2]

2. 시중 판매 제품 대상 데이터 추출

현재 대중들이 소비하는 제품들의 데이터를 파악하기 위해, 식품생산 실적에 따른 상위 기업들의 제품들을 대상으로 하여 데이터를 추출할 제품들을 선정하였다. 특정 맛을 확인하기 위해 타겟은 ‘갈비류’ 소스로 선정하였으며 매운맛이 첨가된 ‘매운 갈비’, ‘불갈비’ 소스도 포함하였다. 또한 2022년 기준 1인 가구 비중은 34.5%를 차지하였는데, 이들을 겨냥한 HMR(Home Meal Replacement) 제품들 내 소스 또한 포함하였다. 추출한 원재료 데이터들을 국가표준식품성분표에 따라 다시 분류하였으며, 이후 마찬가지로 동일한 layout 인 ForceAtlas2 를 적용하여 그래프 플로팅을 진행하였다.

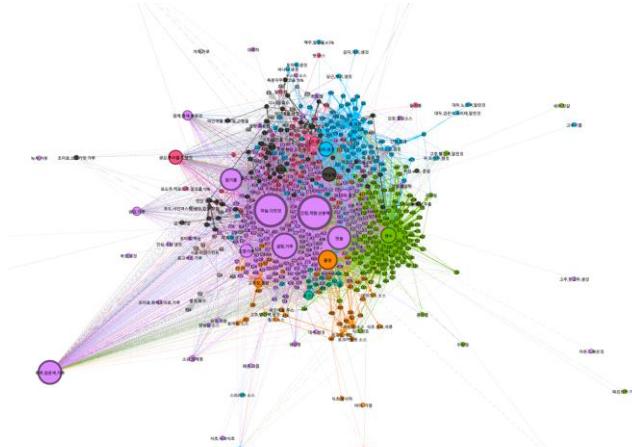
3. 그래프 플로팅

그림 1 의 (a)와 (b)는 대중들이 작성한 레시피 속 원재료 데이터들을 추출하여 플로팅한 결과다. (a)는 먼저 함량 정보 없이 플로팅을 진행한 결과로, 노드의 크기에 따라 원재료들의 사용빈도를 파악할 수 있다. ‘간장, 개량, 산분해’와 ‘마늘, 다진것’의 사용빈도가 높음을 확인할 수 있었다. 이 결과는 어떠한 원재료들이 사용되는지만 알 수 있을 뿐 함량 정보가 포함되어 있지 않아 레시피 별로 맛을 추측하는데는 한계가 있었다. 그리하여 함량 정보를 추가하여 (b)에서 결과를 나타냈다.



(a) 일반 레시피 대상 그래프 플로팅 (함량 정보 미포함)

<그림 1> 시중 판매 제품 및 일반



(b) 일반 레시피 대상 그래프 플로팅 (함량 정보 포함)

레시피 대상 그래프 플로팅 결과

(b)에서는 노드(node)의 크기와 위치, 엣지(edge)의 길이와 굵기에 따라 레시피 속 원재료의 함량과 사용빈도를 알 수 있었다. 노드의 크기가 클수록 사용빈도가 높으며, 엣지에 함량별 가중치를 둬서 노드끼리 연결된 길이가 길수록 함량이 낮고, 가까워서 군집을 이루도록 함량이 높았다. 예를 들어 ‘간장, 개양, 산분해’는 노드의 크기가 크고 중앙에 위치하고 있어 사용빈도와 함량 모두 높은 것을 확인할 수 있었다. 반면 ‘후추, 검은색, 가루’는 노드의 크기는 크나 다른 노드들과 달리 중앙이 아닌 바깥으로 위치하고 있어서 사용빈도는 높으나 함량 자체는 낮은 것을 파악할 수 있다. ‘생수’ 같은 경우는 앞선 노드들에 비해 크기가 작아 사용빈도는 낮으나 주변 레시피 노드들과 연결되어 있는 엣지의 길이가 짧고 굵으며, 군집을 이루고 있었다. 이로 보아 사용빈도 자체는 ‘간장’이나 ‘후추’ 보단 낮으나 한번 사용될 경우 많은 양이 사용되며 그렇게 사용되는 레시피들 또한 확실한 특색을 가지고 있음을 알 수 있었다.

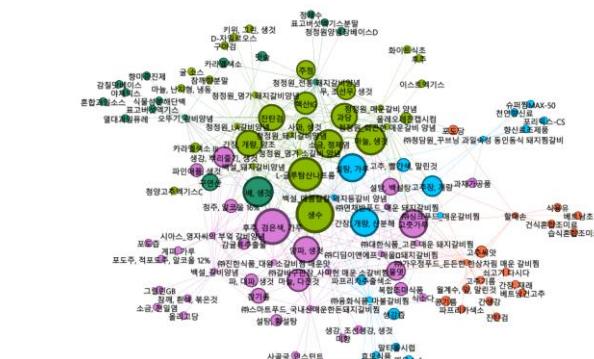
그림 2에서는 대중들의 레시피에서 데이터를 추출한 결과를 바탕으로, 시중에 판매중인 제품에 적용을 한 결과다. 그림 1의 (a)와 마찬가지로, 제품들에 사용되는 원재료의 종류와 그 빈도를 알 수 있었다. 또한 같은 색의 노드들끼리 군집을 이뤄 비슷한 맛을 내는 제품들끼리 군집을 이루는 것을 파악할 수 있었다. 하지만 기업에서 판매중인 제품들 같은 경우에는 배합비율이 알려져 있지 않아 자세한 정보를 파악하는데는 한계가 있었다. 따라서 함량 데이터가 추가된다면 그림 1의 (b)처럼 제품별 맛의 특징을 파악할 수 있을 것이다.

III. 결론

본 논문에서는 식품 소스(Sauce) 데이터의 상관관계 및 특징을 그래프 플로팅을 통해 알아보고자 하였다. 온라인 플랫폼에서 일반 대중들이 작성한 레시피들과 시중에 판매중인 소스 제품들을 대상으로 하여 레시피, 제품 및 원재료들의 데이터 값들을 시각적 속성으로 변환한 후 최종적으로 그래프로 나타냈다. 원재료들의 사용빈도, 각 소스들과 원재료들 간의 상관관계 등 여러 데이터들을 시각화하여 정보를 제공하였다. 따라서 추후 원재료 별로 특정되는 맛 성분들의 데이터를 추가하여

군집을 설정하고 사용정도와 범위를 확인하여 맛을 예측할 수 있을 것이다.

ACKNOWLEDGMENT



<그림 2> 시중 판매 제품 대상 그래프 플로팅

본 연구는 과학기술정보통신부 및 정보통신산업진흥원, 전북특별자치도 및 익산시의 재원으로 수행중인 '농식품 메타버스 기반 기술실증 지원사업(R1201-25-1001)'의 지원을 받아 작성되었습니다.

참 고 문 헌

- [1] Bastian, M., Heymann, S., Jacomy, M. "Gephi: An Open Source Software for Exploring and Manipulating Networks," Proceedings of the Third International ICWSM Conference, p[. 361–362. 2009.
- [2] Jacomy, M. "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software," PLOS One Vol 9, Issue 6, Jun. 2014.