

벤 다이어그램 기반 퍼지 집합 시각화

Visualizing Fuzzy Set Based on Venn Diagram

박예슬, Yeseul Park*, 박진아, Jinah Park**

요약 퍼지 정보 시스템이 요구되는 다량의 데이터들은 퍼지 집합 시각화를 통한 다양한 분석을 필요로 한다. 따라서 본 연구는 퍼지 데이터 집합을 벤 다이어그램을 이용하여 시각화하는 방법을 제안하였다. 여러 주제에 관련되고 관련순위를 갖는 퍼지 데이터를 대상으로 여러 주제를 동시에 갖고 있는 데이터 집합과 여러 주제를 갖고 있는 모든 데이터들의 집합, 그리고 한 주제와 관련되지 않은 데이터 집합을 효과적으로 시각화하여 사용자가 원하는 결과를 제공할 수 있도록 하였다. 또한 서로 다른 주제 클러스터를 겹치거나 포개어서 결과 데이터를 표현하므로 사용자 중심의 정보 검색을 도모한다. 이러한 방법은 웹 검색 엔진 사용자와 웹 개발자에게 원하는 웹 문서를 보다 빠르게 전달 할 수 있는 장점을 갖으며 그 외, 정보 추출과 관련된 여러 목적으로 유용하게 쓰일 수 있는 발전 가능성이 있다.

Abstract Much amount of data which demand fuzzy information system requires various analysis through the fuzzy set visualization. Therefore, this study proposes how to visualize fuzzy data set using variation of Venn diagram. For the fuzzy data which are related to many topics and have ranking of relation, this way gives results that users want by visualizing intersection, union and complementary set. That is, it visualizes the set of fuzzy data which have many topics at once, or the set of all fuzzy data which has topics, or the set of fuzzy data not related to a topic. Users control these sets by overlapping or piling them; visualized with Venn diagram, which is user-oriented. One distinct advantage of this visualization is the fact that it delivers web documents which users of search engine and web developers want much quickly. Furthermore, its possibility can be expanded to several purposes by using for information retrieval.

핵심어: 퍼지 집합, 벤 다이어그램, 시각화, 데이터 마이닝, Fuzzy set, Venn diagram, visualization, data mining

본 논문은 2008년도 정부(교육과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임 (No. R01-2008-000-11678-0)

*박예슬 : 한국정보통신대학교 전산학과 컴퓨터 그래픽스 및 시각화 연구실 석사 과정 e-mail: ystopia@icu.ac.kr

**박진아 : 한국정보통신대학교 전산학과 컴퓨터 그래픽스 및 시각화 연구실 교수 e-mail: jinah@icu.ac.kr

1. 서론

비주얼 데이터 마이닝이라는 데이터 마이닝과 정보 시각화의 결합은 복잡하고 많은 양의 데이터의 분석과 이의 빠른 이해를 위해 점점 더 높은 관심을 받고 있다. 데이터 마이닝은 마케팅, 의학 분야, 회계 분야, 공학 분야, 바이오 정보학 분야에서 많이 쓰이고 있으며, 이러한 분야들의 데이터 마이닝을 이해하기 쉽게 사용자에게 전달하기 위해 정보 시각화를 필요로 하고 있다. 따라서 비주얼 데이터 마이닝은

사용자에게 큰 데이터 저장소로부터 발견될 수 있는 지식을 위한 새로운 해석을 얻을 수 있는 가능성을 제공한다. [1] 그 예로, 질병에 영향을 미치는 단백질을 대상으로 내용 기반 단백질 구조의 비주얼 데이터 마이닝 시스템이 있다. [2] 이 연구는 k번째로 가장 비슷한 구조를 찾기 위해 3D 단백질 구조를 이용한다.

실제 세계에서 많은 양의 데이터는 비선형적이고 불확실성을 갖고 있기 때문에 대부분 퍼지 이론을 따르고 있으며, 이 데이터들을 잘 수행하기 위한 퍼지 시스템이 요구된다.

그러나 이러한 정보의 불확실성으로 빚어지는 복잡성으로 퍼지 시스템이 수행하는 것 - data들 사이 상호관계의 불확실성을 해석하는 것, 그리고 그러한 불확실성의 파장이 모든 결과물의 확실성 수준에 어떠한 영향을 미칠 것인지에 대한 분석 - 을 사용자가 직접적으로 이해하는 것은 더 어려워진다. 이로 인해, 시스템 사용자 및 시스템을 최적화하는 방법을 찾는 디자이너들은 퍼지 시스템의 더 쉬운 이해를 위한 방법들을 요구하고 있다. 그 방법으로, 퍼지 시스템의 시각화 연구가 진행되면서 앞에서 언급한 문제점들이 해결되고 있다. [3] 그 예로, 그림 1에서 보듯이 병렬 좌표에서 퍼지 점들을 시각화하여 퍼지 데이터의 전체적인 요약과 더불어 소속도 까지 파악할 수 있는 연구가 된 바 있다. [4] 또 다른 연구로, 논문 [5]는 C-means 클러스터링 방법으로 그림 2와 같이 퍼지 소속도를 여러 개의 클러스터에 시각화하는 것을 제안하였다. 2003년에는 퍼지 시스템 시각화의 충족 요건에 대한 연구도 이루어졌다. [6]

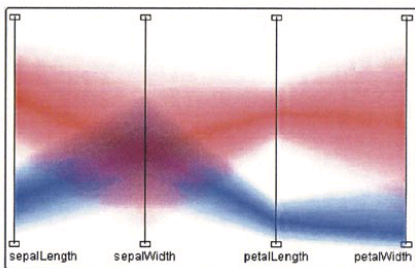


그림 1 퍼지 데이터 시각화 [4]

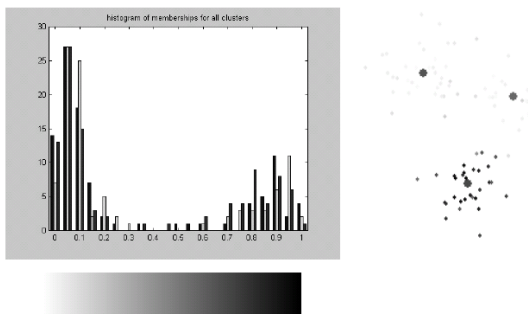


그림 2 퍼지 데이터의 클러스터링 및 시각화 [5]

퍼지 시스템에서 세부적으로 더 나아가 퍼지 집합의 시각화도 관심을 받고 있는데, 본 연구는 퍼지 집합의 시각화에 대한 방법을 제안하고자 한다. 즉, 퍼지 집합 이론의 데이터 마이닝을 통해 새로운 정보를 찾는 과정을 시각화를 통해 보다 쉽게 전달하고자 하는 것이다. 사용자는 퍼지 데이터들을 갖는 여러 클러스터들을 비교하여, 서로 다른 클러스터 사이에서 관련 있는 데이터들을 분석하고 그 결과를 알고자 할 경우가 있다. 즉 여러 클러스터의 퍼지 데이터들 중 공통으로 포함되는 데이터들로 이루어진 집합이나, 여러 클러스터의 퍼지 데이터들을 모두 합친 집합에 대한 분석이 필요할 때 퍼지 집합의 시각화가 필요하다. 이를 위한 방법

으로 본 연구는 벤 다이어그램을 이용하여 퍼지 데이터들의 전체적인 요약과 교집합과 동시, 퍼지 데이터들의 교집합, 합집합이 의미하는 바를 각각 데이터의 출신 클러스터와 또 그 클러스터와의 소속도를 통해 파악할 수 있도록 하였다. 이와 같은 결과물은 결과 데이터가 포함된 각각 클러스터에 얼마만큼의 소속도로 포함되어 있는지 그 정보를 사용자에게 전달한다.

2. 본론

2.1 벤 다이어그램의 변형

본 연구는 벤 다이어그램을 변형하여 퍼지 데이터들의 전체적인 요약과 교집합과 합집합을 시각화하는 방법을 제안하였다. 여기서 원의 성질을 이용한 벤 다이어그램은 데이터를 전체적으로 분포시키고 동시에 교집합, 합집합의 의미를 전달하고자 하는 본 연구의 시각적 도구로 쓰였다. 원 벤 다이어그램에서 원은 원의 중심으로부터 같은 거리에 있는 데이터들을 표현하기에 유용하고, 벤 다이어그램은 형식 논리에 있어 가장 기본적인 다이어그램으로 서로 다른 집합들 사이의 관계를 보여주기 위한 그림이다. 따라서 사용자에게 보다 직관적이고 정확한 정보를 전달하기 위한 시각적 도구로 벤 다이어그램이 적합하며, 이것을 변형시킨 새로운 벤 다이어그램으로 퍼지 데이터의 시각화를 가능하게 하였다.

먼저, 퍼지 데이터들을 전체적으로 분포시키기 위해 하나의 클러스터를 벤 다이어그램의 집합 A에 표시한다. 집합 A는 하나의 클러스터이므로, A의 데이터들이 모두 같은 주제를 갖고 있다는 의미에서 A의 전체적인 바탕을 한 가지 색으로 표시한다. 그리고 퍼지 데이터들의 소속도는 원의 가장자리로부터 중심까지의 거리로 정의 내렸다. 따라서 원의 중심에서 데이터들의 집합에 대한 소속도는 1이고, 원의 가장자리에서는 0이다. 데이터들을 표시하기 위해 데이터의 소속도가 가리키는 거리를 반지름으로 하는 동심원을 그리는 형태로 분포시켜서 데이터 소속도의 전체적인 요약과 가능하도록 하였다. 예를 들어 그림 1과 같이, 소속도 0.2인 퍼지 데이터는 원의 가장자리로부터 원의 전체 반지름의 0.2배에 해당하는 곳을 반지름으로 하는 원을 그리고, 소속도 0.4인 퍼지 데이터는 원의 가장자리로부터 원의 전체 반지름의 0.4배에 해당하는 곳을 반지름으로 하는 원을 그려서 데이터의 소속도가 어떻게 분포되었는지 확인할 수 있게 하였다. 이와 같이 퍼지 데이터들의 소속도를 의미하는 동심원들을 통해 퍼지 집합들은 각각 소속도별 데이터의 전체적인 분포를 전달할 수 있다.

2.2 퍼지 집합의 시각화

퍼지 데이터를 변형 벤 다이어그램에 분포시킬 때 그려지는 동심원들은 그림 3의 클러스터 A, B에서와 같이 소속도에 의해 선 투명도가 원의 중심으로부터 소속도가 0인 가장자리까지 점점 높아지도록 하였다. 동심원 선의 투명도가 높을수록 소속도는 작은 값을 의미하기 때문에 투명도 'alpha' 값은 1-소속도 값으로 할당하였다. 이것은 전체적인 소속도 수치의 분포를 파악하기 위한 방법이다.

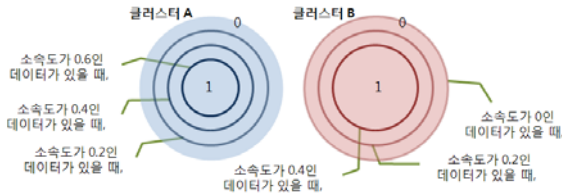


그림 3 소속도가 전체적으로 분포된 변형 벤 다이어그램

퍼지 데이터를 변형 벤 다이어그램에 위와 같이 분포시킨 다음, 클러스터 A와 B에 공통으로 포함되는 데이터들로 이루어진 집합, 즉 교집합과 클러스터 A와 클러스터 B의 데이터들을 모두 합친 집합, 즉 합집합을 추출하기 위해 사용자는 클러스터 A와 B를 마우스로 겹치거나 포개어 추출된 데이터들이 어떤 데이터들인지 분석하고자 한다. 또한 클러스터 A를 제외한 모든 데이터들의 집합, 즉 여집합의 시각화도 가능하다. 이들의 시각화 이전에 퍼지 집합 이론에는 합집합, 교집합, 여집합이 있고, 본 연구는 이 세 가지 논리 연산에 대해 다음과 같은 방법을 제안하였다.

2.2.1 교집합

여러 클러스터의 데이터들 중 공통으로 포함되는 데이터들로 이루어진 집합은 퍼지 집합 이론으로 찾을 수 있다. 퍼지 집합 이론의 교집합은 여러 클러스터에 동시에 포함된 퍼지 데이터 D 중 소속도가 가장 작은 값을 가리킨다. 예를 들어, 여러 클러스터에 동시에 포함된 퍼지 데이터 D의 소속도가 클러스터 A에서는 0.2, B에서는 0.3, C에서는 0.5일 경우, 소속도 0.2의 데이터가 교집합의 결과 값이 된다. 또 다른 예로, 동시에 포함된 퍼지 데이터 D가 없을 경우에는 여러 클러스터 중 적어도 한 개 이상의 클러스터는 D를 포함하지 않은 경우, 즉 소속도 0을 의미하므로 소속도 0의 데이터가 교집합에 표시된다. 이를 전체집합 U의 퍼지 부분집합인 A와 B, 그리고 U의 원소 데이터 u의 집합 기호로 표현하면 다음과 같다.

$$\mu_{A \cap B}(u) = \min\{\mu_A(u), \mu_B(u)\} \quad (1)$$

이를 적용하여, 퍼지 데이터가 분포된 변형 벤 다이어그램들을 서로 얼마나 겹쳤는지에 따라 결과 데이터를 의미적으로 다르게 제공하여 보다 정밀한 정보 분석을 가능하게 하였다. 각 클러스터의 가장자리 근처만 겹쳤을 경우, 즉 겹

친 면적이 작을 경우에는 가장자리의 소속도 0과 그 주변의 비교적 작은 소속도를 대조한 결과가 교집합이다. 따라서 가장자리 근처 겹친 부분은 각 클러스터의 소속도를 동시에 낮게 갖고 있는 데이터끼리의 교집합을 의미한다. 각 클러스터의 중심에 가깝게 겹쳤을 경우, 즉 겹친 면적이 클 경우에는 중심 주변의 비교적 큰 값들도 포함하여, 소속도 0부터 소속도 1과 그 주변의 비교적 큰 값까지 모두 대조한 결과가 교집합이다. 따라서 가장자리로부터 중심 근처까지 겹친 부분은 각 클러스터의 소속도를 동시에 낮게 갖고 있는 데이터들과 높게 갖고 있는 모든 데이터끼리의 교집합을 의미한다. 예를 들어, 그림 4와 같이 동심원 0에서 0.6사이가 겹쳤을 때, 공통으로 두 클러스터에 포함된 0에서 0.6까지의 소속도를 갖는 데이터들 중, 작은 값을 갖는 데이터가 교집합이다. 0에서 0.6까지 데이터들 중 공통으로 포함된 데이터 D는 클러스터 A에서 0.2의 소속도를, 클러스터 B에서 0.5의 소속도를 갖고 있으면 클러스터 A의 데이터가 교집합의 결과가 된다.

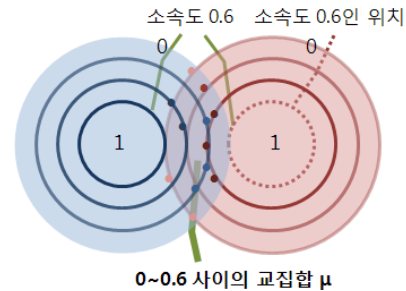


그림 4 퍼지 데이터의 교집합

사용자가 마우스를 두 클러스터가 겹쳐진 교집합 부분에 올리면, 그림 4와 같이 교집합 데이터들은 각 동심원 선의 색과 같고, 해당 동심원 위의 교집합 부분에 위치한다. 클러스터가 겹쳐진 부분의 바탕색은 두 클러스터의 교집합 관계를 직관적으로 나타내기 위해 두 클러스터의 전체적인 바탕색을 합친 색으로 표시하였다. 또한 사용자의 마우스 오버에 의한 결과 데이터 확인을 가능하게 하여 겹친 부분이 교집합을 의미함을 보여주었다. 이로써 교집합 데이터들이 어떤 색의 동심원 위에 있는지에 따라 어떤 클러스터에 포함되어 있는지 각각 자신의 출신 클러스터를 확인할 수 있으며, 동심원 선의 투명도를 통해 교집합 데이터들의 소속도를 파악할 수 있다.

2.2.2 합집합

퍼지 집합 이론의 합집합은 여러 클러스터의 데이터들을 모두 합친 것으로 여러 클러스터에 동시에 포함된 퍼지 데이터 D 중 소속도가 가장 큰 값을 가리킨다. 예를 들어, 여러 클러스터에 동시에 포함된 퍼지 데이터 D의 소속도가 클러스터 A에서는 0.2, B에서는 0.3, C에서는 0.5일 경우, 소속도 0.5의 데이터가 합집합의 결과 값이 된다. 또 다른 예로, 동시에 포함된 퍼지 데이터 D가 없을 경우, 즉 여러 클

러스터 중 클러스터 A는 D를 포함하지 않고, 클러스터 B는 D를 0.2의 소속도로 포함하고 있을 때, 0.2의 데이터가 합집합에 표시된다. 이를 전체집합 U의 퍼지 부분 집합인 A와 B, 그리고 U의 원소 데이터 u의 집합 기호로 표현하면 다음과 같다. ↓

$$\mu_{A \cup B}(u) = \max\{\mu_A(u), \mu_B(u)\} \quad (2) \downarrow$$

합집합의 시각화는 사용자들이 클러스터들을 포개어 겹쳤을 경우로 정의 내렸다. 클러스터들이 포개어 겹치면 교집합에서와 같이 클러스터들 사이의 합집합 관계를 직관적으로 나타내기 위해 클러스터들의 전체적인 바탕색을 합친 색으로 표시하였다. 또한 각 클러스터의 동심원 선들도 포개어 지므로 동심원 선들의 색도 합쳐지도록 한다. 그러나 이 경우에 사용자가 마우스를 포개어진 클러스터 위에 올리면, 그림 5와 같이 합집합 결과 데이터는 포개기 전 클러스터의 동심원 선 색을 유지한다. 그리고 합집합 결과 데이터의 위치는 해당 동심원 위에 서로 겹치지 않게 무작위로 분포시킨다. 교집합과 마찬가지로 합집합의 데이터들도 동심원 선 색, 투명도에 따라 출신 클러스터와 소속도를 바로 알 수 있다.

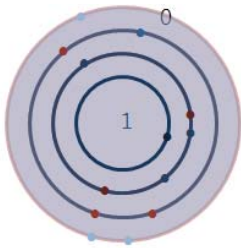


그림 5 퍼지 데이터의 합집합

2.2.3 여집합

퍼지 집합의 여집합은 한 클러스터의 퍼지 데이터들을 제외한 다른 모든 퍼지 데이터들의 집합을 의미하므로, 변형 벤 다이어그램을 통해 한 클러스터에 소속되어 있지 않은 나머지 다른 데이터들을 시각화하면 된다. 먼저, 전체집합 U의 퍼지 부분 집합인 A, 그리고 U의 원소 데이터 u의 집합 기호로 표현하면 다음과 같다.

$$\mu_{\bar{A}}(u) = 1 - \mu_A(u) \quad (3) \downarrow$$

이는 전체 집합에서 하나의 클러스터 A를 제외시킨 나머지 부분이므로 그림 6과 같이 A의 모든 데이터들의 소속도를 1에서 뺀 값들로 다시 할당하여, 새로운 소속도로 다시 동심원을 위치시킨다. 예를 들어, 클러스터 A에 소속도 0.2로 포함된 데이터 D가 아닌 다른 클러스터에 소속된 나머지

데이터 D는 소속도 0.8로 존재할 것이다. 따라서 클러스터 A 속 데이터의 소속도를 1에서 뺀 값인 나머지를 갖는 데이터들을 동심원에 표시하여 여집합을 시각화한다. 이와 같은 시각화 방법을 통해 다른 클러스터와의 집합 관계를 다룰 때, 직접적인 시각적 연산이 가능하다.

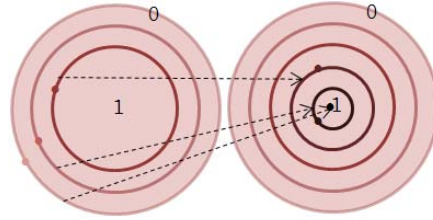


그림 6 좌: 집합 A, 우: 집합 A의 여집합

2.3 비주얼 데이터 마이닝

본 연구가 비주얼 데이터 마이닝 관점에서 어떤 의미를 갖는지 다음과 같은 질문을 통해 파악하고자 한다.

(1) 퍼지 집합 이론을 시각화 하여 사용자가 얻을 수 있는 정보는 무엇일까?

일반적으로 퍼지 집합 이론을 통한 결과 값들은 내재적으로 데이터들의 중요도와 출신 클러스터 정보를 포함하고 있다. 이러한 정보들을 눈으로 확인하고자 할 때, 이에 적합한 정보 시각화가 이를 가능하게 한다.

(2) 그렇다면 퍼지 집합 이론을 통해 새로운 정보를 찾자 할 때, 벤 다이어그램 변형이라는 시각적 도구는 어떤 의미와 중요성을 갖는가?

퍼지 집합 이론은 소속도를 갖는 데이터들의 집합을 대상으로 하기 때문에 결과 데이터와 그것의 소속도를 효과적으로 전달해야 한다. 결과 데이터가 어느 클러스터 소속이고 그 클러스터에서 얼마나 중요한지에 대한 정보를 시각화하고자 할 때, 나열식으로 전달한다면 각각 데이터에 대한 정보는 제공할 수 있지만 여러 데이터들의 전체적인 분포는 보여주기 어렵다. 또한, 나열식은 각 클러스터마다 데이터들이 어떤 소속도를 갖고 있는지에 대한 전체적인 요약 전달하기 어렵다. 벤 다이어그램 변형이라는 시각적 도구는 바로 이러한 어려움을 해결한다. 즉, 퍼지 집합 이론 시각화에 적합한 벤 다이어그램은 각 클러스터의 전체적인 요약과 결과 데이터의 전체적인 분포 및 각 데이터들의 개별 정보에 효과적인 설명 방식이다.

(3) 본 연구의 시각화를 통해 추출된 정보는 사용자로 하여금 어떠한 도움을 주는가?

결과 데이터의 소속도와 출신 클러스터의 분포를 통해 사용자는 결과 데이터가 어떤 군집 안에서 어떤 내용을 주로 다루고, 얼마만큼의 중요성을 갖는지 한 눈에 파악이 가능하

다. 예를 들어 사용자가 A와 B에 대해 동시에 얼마만큼의 중요성을 갖는 데이터를 찾고자 할 때, A에 포함되지 않지만 C에 대해 얼마만큼의 중요도를 갖는 데이터들을 찾고자 할 때, 시행착오 없이 정확한 데이터를 빠르게 얻을 수 있다.

3. 적용 및 해석

위에서 제안한 퍼지 집합 이론의 시각화 방법은 웹 검색 엔진에서 검색어를 통해 결과 문서를 찾는 과정과 밀접한 관련이 있다. 웹 검색 엔진은 기존 볼리언 모델에 순위를 부여하는 기능을 넣어 확장시킨 퍼지 집합 모델을 사용하는 사례가 많기 때문에 검색 결과 문서들의 심층적인 분석을 위해 퍼지 집합 시각화가 유용하다.

예를 들어 그림 7과 같이 검색어에 '영화AND음악'을 넣으면 영화와 음악에 둘 다 동시에 관련된 웹 문서들을 얻게 된다. 이들 영화, 음악과 관련된 웹 문서들의 전체적인 순위 분포와 검색어 '영화AND음악'의 결과 문서들을 분석하고자 본 논문이 제안한 방법을 적용해보았다. 사용자가 검색어에 필요한 '영화', '음악'만 입력하면 검색 결과로 그림 3과 같은 변형 벤 다이어그램을 제공한다. '영화AND음악'의 검색을 위해 사용자는 영화와 음악 클러스터를 서로 원하는 만큼 겹쳐서 영화와 음악이 동시에 얼마나 웹 문서에 포함된 건지를 파악한다. 작게 겹치면 영화와 음악이 조금씩 웹 문서에 포함된 것이고, 크게 겹치면 많이 포함된 웹 문서들까지 포함된 것이다.

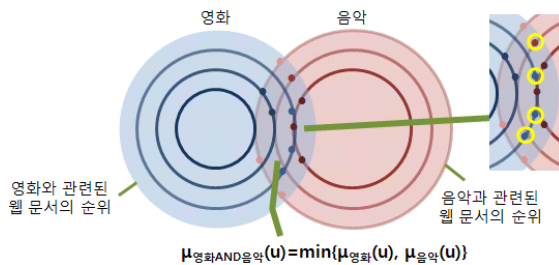


그림 7 검색어 '영화AND음악'을 통한 웹 문서들의 퍼지 집합 시각화 예

사용자가 영화와 음악에 동시에 속하는 결과 웹 문서들 중에 영화에 대한 소속도 0.2를 갖는 문서나 음악에 대한 소속도 0.2를 갖는 문서를 찾고자 한다면, 표시된 포인트들의 색깔과 포인트가 속한 동심원의 투명도를 통해 찾을 수 있다. 그림 7의 노란색 원으로 표시한 부분이 찾은 데이터이며 이로써 얼마만큼의 소속도를 갖는 문서들인지 보다 빠르게 얻을 수 있다.

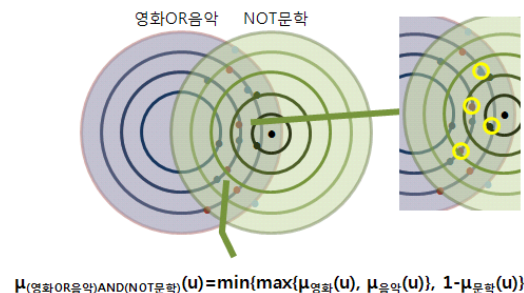


그림 8 검색어 '(영화OR음악)AND(NOT문학)'을 통한 웹 문서들의 퍼지 집합 시각화 예

또 다른 예로, '(영화OR음악)AND(NOT문학)'의 검색을 위해 사용자는 영화, 음악, 문학을 입력한다. 먼저 영화와 음악 관련 웹 문서들의 모든 문서 집합으로서 두 클러스터를 포개어 겹친 결과와 문학이 관련되지 않은 —문학 클러스터를 서로 원하는 만큼 겹치게 되면 거기에 해당하는 웹 문서들이 표시된다. 그림 8은 위 예에 대한 설명으로 소속도 0에서 1까지의 웹 문서들을 대상으로 교집합을 구하였다.

사용자가 찾고자 하는 결과 웹 문서들 중, 출신 클러스터가 영화이고 소속도가 0.2인 문서들, 출신 클러스터가 음악이고 소속도가 0.4인 문서들, 문학을 제외한 문서들 중 0.8의 소속도를 갖는 문서들은 그림 8의 노란색 원이 가리키는 포인트들에 해당한다.

4. 시스템 구조

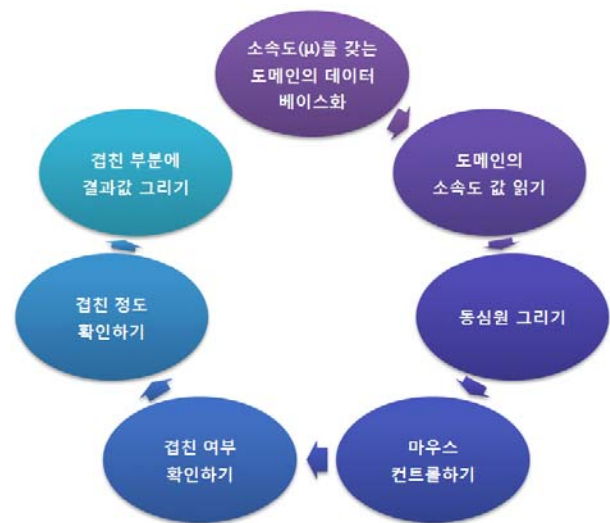


그림 9 시스템 구조

본 연구의 도메인은 소속도를 갖고 있고, 내용에 따라 분류한 데이터들이다. 따라서 그림 9와 같이 먼저 소속도를 갖는 도메인을 대상으로 데이터베이스를 만들어야 한다. 데이터베이스는 크게 분류 주제(클러스터)를 담고 있는 Vector와 겹친 클러스터들을 담는 Array List를 두었다. 주어진 분류 주제로 클러스터들을 원으로 시각화한 후에, 도메인 데이터의 소속도 값을 읽고 동심원을 그린다. 그리고, 마우스로 원들을 움직여 서로 겹쳤을 때와 마우스를 겹친 부분 위에 올릴 때의 마우스 컨트롤을 고려하여야 한다. 원들이 서로 겹쳤을 때에는 겹친 클러스터들을 담는 Array List에 추가하고, 겹친 원들이 다시 분리되었을 때에는 분리된 정보를 Array List에 넣어 업데이트 시킨다. 겹친 정도는 각각 원들의 중심 거리를 비교하여 판단하고, 합집합인지 교집합인지를 결정한다. 집합 연산이 이루어지면 클러스터의 겹친 부분에 마우스를 올렸을 때 결과 값을 그린다. 이와 같은 시스템 구조를 바탕으로 벤 다이어그램 변형을 이용한 퍼지 시각화를 구현하고 있다.

5. 구현

본 연구를 구현하기 위해 미리 구해진 데이터들의 소속도가 필요하다. 위에서 언급한 웹 엔진을 이용한다면 웹 문서들의 소속도를 토대로 벤 다이어그램의 시각화를 구현할 수 있다. 그러기 위해선 단어에 따른 웹 결과 문서들의 소속도를 문서 각각의 단어 빈도수와 단어 당 갖는 문서 빈도수를 이용하여 계산해야하는 큰 작업이 필요하다. 따라서 구현에 쓰일 도메인 데이터들의 소속도를 임의로 정해주어 시각화 시스템 구현에 임하였다. 구현은 정보 시각화 디자이너에게 많이 쓰이는 MIT에서 개발된 자바 기반 Processing [7] 환경에서 진행하였다.

5. 결론

본 연구는 실제로 웹 검색 엔진 사용자나 웹 문서들을 통해 정보를 추출하고자 하는 개발자에게 원하는 결과를 빠르게 제공하고자 벤 다이어그램에 기반한 퍼지 집합의 시각화 방법을 제안하였다. 하지만 웹 문서 뿐만 아니라 더 나아가 기본적으로 주제별 분류가 가능하고 한 주제에 대한 여러 데이터가 순위를 갖으며, 각 데이터들이 동시에 다양한 주제로 분류가 되는 도메인을 대상으로, 서로 다른 주제 클러스터를 겹치거나 포개어서 다양한 분석 결과를 가져올 수 있다. 본 연구의 시각화를 통해 사용자가 얻는 정보는 교집합, 합집합, 여집합의 복합 연산에 의해 결정된 결과 데이터들의 직관적인 시각 요소들에 의해 파악이 가능하다. 그리고 얻은 정보를 이용한 다양한 자료 분석이 이루어진다. 따라서 정보 특성상 웹 문서와 비슷한 여러 데이터들을 대상으로 사용자가 원하는 세부 분석을 할 수 있다는 점이 본 연구의 장점이다.

현재 벤 다이어그램 기반 퍼지 집합 시각화 연구는 소속도 값을 임의로 갖는 임의 데이터를 대상으로 구현되었기 때문에 실제 웹 검색 엔진이나 그 외 객관적인 소속도 값이 계산되어 있는 도메인은 아직 시각화해보지 않은 상태이다. 따라서 본 연구를 실제 도메인에 적용시킨 후, 그 결과에 대

한 사용자 평가를 추진할 계획이다. 또한 본 연구가 사용자에게 편리한지에 대한 분석이 필요하므로, Norman의 사용자 인터페이스의 설계 원칙 [8]과 Jacob Nielson의 휴리스틱 평가 방법 [9]를 토대로 효과적인 시각화와 시각요소에 따른 사용자 평가를 통해 개선시키고자 한다.



참고문헌

- [1] Viktor H.L. and Paquet E., "Visual Data Mining: From Visualization to Visual Information Mining", Encyclopedia of Data Warehousing and Mining (2nd Edition), NRC 49882, 2008.
- [2] Paquet E. and Viktor H. L., "Exploring Protein Architecture using 3D Shape-based Signatures, ECMB, pp.1204~1208, 2007.
- [3] Pham, Binh L. and Brown, Ross A., "Visualization of Fuzzy Systems: Requirements, Techniques and Framework", Future Generation Computer Systems 21(7), pp.1199~1212.
- [4] Michael R. Berthold and Lawrence O. Hall, "Visualizing Fuzzy Points in Parallel Coordinates", Vol. 11, No. 3, IEEE Transactions on Fuzzy Systems, 2003.
- [5] Cox, Z., Dickerson, J.A. and Cook, D., "Visualizing Membership on Multiple Clusters after Fuzzy C-means Clustering", Proceedings of Visual Data Exploration and Analysis VIII, pp.60~68, 2001.
- [6] Pham, Binh L. and Brown, Ross A., "Analysis of Visualization Requirements for Fuzzy Systems", Proceedings of GRAPHITE Conference, pp. 181~187, 2003.
- [7] R. Casey, F. Ben, Processing: a programming handbook for visual designers and artists, The MIT Press, Cambridge, USA, 2007.
- [8] Norman, D.A., The Design of Everyday Things, Basic Books, New York, USA, 1988.
- [9] Nielson, J., Usability Engineering, Morgan Kaufmann, San Francisco, USA, 1993.