

의사결정트리를 활용한 정보시각화 디자인 교육방법 제안

A Education Method for Information Visualization Using Decision Tree

주 저 자: 김성곤 (Kim, Sungkon)

울산대학교 디지털콘텐츠디자인전공

sungkon@mail.ulsan.ac.kr

접수일자 2018. 11. 16. / 심사완료일자 2018. 12. 20. / 게재확정일자 2018. 12. 26.

이 논문은 2016년 울산대학교 연구비에 의하여 연구되었음

Abstract

For Information visualization education, we studied a step-by-step learning method from topic setting and data collection to visual content evaluation. In this process, we have developed a Visualization Contents Generation Tree(VCGT) educational model by applying decision tree theory including segmentation, classification, and prediction. This educational model is divided into 'trunk' phase for topic setting, 'node generation' phase for selection of contents production method, and 'conformity assessment' phase for selecting the best ideas. In the node creation step, there is three step such as, a 'visualization form' step for determining the main expression method for materializing the information visualization concept, a 'rhetorical effect' step for determining the effective expression method, and 'visual expression element' step for expressing the data in a specific visual form. The learner creates content combinations of various information visualization plans through content branching and evaluates each content combination value. The education model confirmed the effectiveness of the phase learning process if it was applied to 103 design undergraduate students.

Keyword

Decision Trees, Information Visualization Design Education, Visualization Contents Generation Tree (VCGT)

요약

정보시각화 교육을 위해, 주제 설정 및 데이터 수집에서 시각화 콘텐츠 평가에 이르는 단계별 학습 방법을 연구했다. 이 과정에서 본 연구자는 세분화, 분류, 및 예측 과정이 포함된 의사 결정 트리 이론을 응용하여, 시각화 콘텐츠 생성나무(VCGT) 교육 모델을 개발했다. 이 교육 모델은 주제 설정을 위한 트렁크 단계, 콘텐츠 제작 방법 선택을 위한 노드 생성 단계 그리고 최적의 아이디어 선택을 위한 적합성 평가로 구분된다. 노드 생성단계에는 정보시각화 콘셉트 구체화를 위한, 주된 표현의 방법을 결정하는 '시각화 형태 단계', 이를 효과적으로 표현할 방법을 결정하는 '수사학적 효과' 단계, 그리고 데이터를 구체적인 시각적 형식으로 표현할 방법을 결정하는 '시각 표현 요소' 단계가 있다. 학습자는 콘텐츠 분기를 통해 여러 정보시각화 기획의 콘텐츠 조합을 만들고, 각 콘텐츠 조합 값을 평가한다. 교육 모델은 디자인 학부생 103개의 과제에 적용하였고, 단계 학습 과정의 효율성을 확인했다.

목차

1. 서론

2. 교육내용 구성

- 2-1. 의사결정트리
- 2-2. 교육내용 구성도

3. 형식, 수사학, 요소의 설계

- 3-1. 시각화 형식의 종류

- 3-2. 수사학적 효과의 종류

- 3-3. 시각 표현 요소의 종류

4. 의사결정나무 이론의 활용

5. 결론

참고문헌

1. 서론

윌리엄(William Playfair, 1823)이 정보표현을 차트나 도표의 형식으로 사용한 이후, 정보 시각화 분야는 발전을 이루었다. 최근에는 네트워크와 멀티미디어기술의 발전으로 모션, 스토리텔링, 인터랙션 그리고 3차원 그래픽을 활용한 동적 정보시각화를 흔히 볼 수 있다.

연구자는 대학에서 인포그래픽과 정보시각화를 강의한다. 인포그래픽 강의에서는 스토리텔링이 포함된 2차원 그래픽 중심의 정보표현 방법을 강의하고, 정보시각화 강의에서는 멀티미디어 기능을 활용한 3차원 그래픽 중심의 정보표현 방법을 강의한다. 강의 내용에서 데이터베이스의 활용 유무도 차이가 있다. 수강생들은 막대그래프 혹은 차트 정도의 제작 경험이 있고, 정보시각화 제작의 전반적인 지식이 부족하다. 여러 제작 단계와 여러 미디어를 사용하는 정보시각화 디자인강의를, 경험이 적은 학생들에게 교육하기에 어려움이 있었다. 이를 위해, 오랜 시간동안 시각화 형태의 종류와 효과적인 시각적 표현의 방법 그리고 제작에서 범하기 쉬운 시각화의 오류와 해결안에 관한 교육 방법을 선행 연구하였다.

본 연구에서는 정보시각화 콘셉트 개발에 필요한 구체적인 제작 과정의 교육 모델을 제시한다. 수강생들에게 어떠한 시각적 형태의 표현이 적합하고, 어떠한 수사학적 기법을 응용하며, 그리고 어떠한 그래픽 요소를 사용하여 데이터를 표현하는 것이 적합한가를 제시한다. 교육모델을 개발이후, 디지털콘텐츠디자인 전공 3학년 학생을 대상으로 85개의 개별과제를 교육하고, 4학년 학생을 대상으로 2인 1조의 18팀을 교육하여, 전반적인 교육 모델의 효율성을 규명했다.

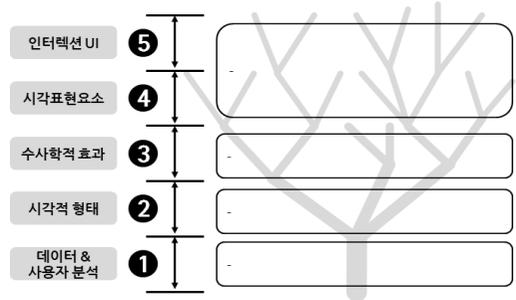
2. 교육내용 구성

2-1. 의사결정트리

본 연구에서는 교육 모델을 개발하기 위해 의사결정나무이론을 응용하였다. 응용 목적은 지식과 경험이 없는 학생들에게 정보시각화 콘텐츠 개발의 이해 폭을 넓히고, 그리고 정해진 강의시간 안에 단계적으로 교육시키기 위함이다. 의사결정나무는 의사결정 규칙을 도표화한 이후, 관심대상이 되는 그룹을 몇 개의 소단위 그룹으로 분류하여, 관심대상의 미래 행동을 예측할 때 사용한다.¹⁾ 또한, 미래 예측 분석의 정확도를 높이는

1) J. R. Quinlan, Induction of decision trees, Machine

방법 이외에, 분석과정을 설명하기 위해 사용되기도 한다.²⁾ 본 교육 모델에서는 미래의 결과 예측 보다는 분석 과정의 설명 방법에 중점을 두고 응용하였다.



[그림 1] Decision Tree Analysis of Contents Hierarchy Classification

의사결정나무는 분류, 세분화, 예측, 그리고 상호작용효과의 관계 파악 등을 목적으로 사용하는데, 본 연구에서는 여러 변수에 활용하여, 목표 변수의 범위를 몇 개의 등급으로 나누는 '분류'의 방식을 특히 응용하였다. 본 연구에서 의사결정나무 이론을 활용한 교육모델을 시각화 콘텐츠 생성나무(Visualization Contents Generation Tree, VCGT)라 칭한다.

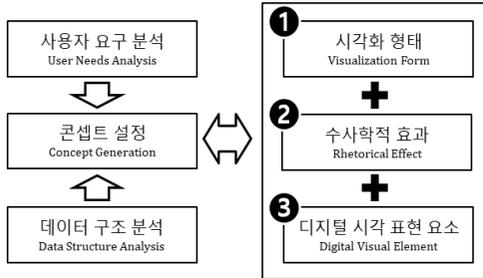
VGCT의 의사결정나무는 가지 형성단계, 콘텐츠 가지치기단계, 그리고 적합성 평가단계로 나뉜다. VGCT는 수강생들이 개별적인 정보시각화 주제를 설정한 이후에 적용이 가능하다. 주제 설정이후, 수강생은 데이터를 수집하고, 사용자의 환경과 특징을 조사한다. 그리고 사용할 미디어의 특성을 파악한다. 이 과정을 트렁크(Trunk) 단계라 칭한다. 두 번째부터 다섯 번째 단계까지를 마디(Node)라고 칭한다. 마디로부터 끝마디까지 연결된 마디를 가지(Branch)라 칭한다. 하나의 마디에서 분리된 마디를 자식마디(Child Node)라 칭하며, 그리고 자식마디의 상위 마디를 부모마디(Parent node)라 칭한다. 본 연구에서는 두 번째 가지마디를 시각적 형태의 결정 마디, 세 번째 가지마디를 수사학적 효과의 결정마디, 네 번째 가지마디를 시각적 표현 요소의 결정 마디, 그리고 마지막 다섯 번째 가지마디

Learning, 1986, Vol.1, Issue 1, pp.81-106.

2) Kim, Ki Sook, Kim, Kyuing Hee, A Prediction Model for Internet Game Addiction In Adolescents: Using a Decision Tree Analysis, Journal of Korean Academy of Nursing, 2010, Vol.40, No.3, pp.378-388.

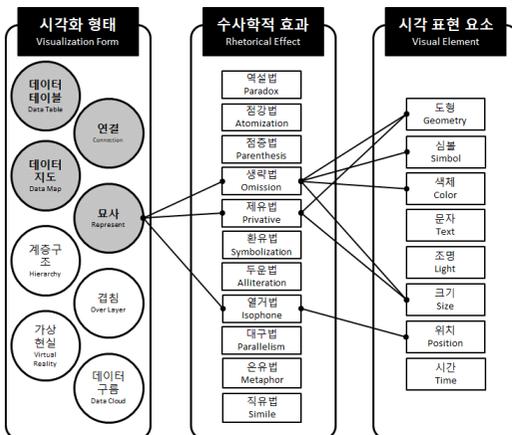
를 인터랙션 UI 결정 마디로 정했다.

2-2. 교육내용 구성도



[그림 2] Information Visualization Development Process Based on Concept Generation

정보시각화의 개념 결정은 데이터 분석으로 시작된다. 데이터 테이블을 구성하고 가장 핵심이 되는 주요(Key) 데이터를 설정한다. 이후, 주요 데이터를 중심으로 어떤 미디어를 사용하여 어떠한 형식으로 정보를 표현할 것인가를 결정한다. 이때 사용자의 요구사항 분석도 중요하다. 사용자마다 보고자하는 데이터의 종류와 형식이 다르기 때문이다. 이러한 데이터 구조 분석과 사용자 요구 분석을 통하여 정보시각화 개념 결정이 이루어진다.



[그림 3] Contents Factors (Visual Form, Rhetorical Effect, Visual Element) for Information Visualization

정보시각화 개념을 구체화하기 위해서는, 주된 표

현의 시각화 형태를 결정하고, 이를 효과적으로 표현할 수사학적 표현 방법을 정한다. 이후 데이터를 구체적인 시각적 형식으로 표현할 시각표현 요소를 결정한다. 과거 연구에서 시각화 형태, 수사학적 표현 방법 그리고 시각적 표현 요소에 종류와 표현 방법에 관하여 논하였다.

VGCT는 이러한 순차적인 선택의 과정을 수강생들이 실행하는 것이다. VGCT 과정 이전에, 수강생들에게 충분한 시각적 형태의 예와, 수사학적 표현 방법 그리고 데이터를 시각적 표현요소에 대응하는 방법을 강의한다. VGCT에 사용되는 형태, 효과, 요소들은 강의자의 선택에 따라 변형이 가능하다. 강의자가 생각에 따라서 각각의 내용을 첨부하거나 삭제할 수 있다. 왜냐하면, 멀티미디어 기술의 발전과 함께, 새로운 시각적 형태와 수사학적 효과가 등장하기 때문이다. 본 연구의 VGCT에서는 데이터 테이블을 포함한 8종류의 시각화 형태와 역설법을 포함한 11종류의 수사학적 효과 그리고 8종류의 시각 표현 요소를 제시한다.

3. 형식, 수사학, 요소의 설계

3-1. 시각화 형식의 종류

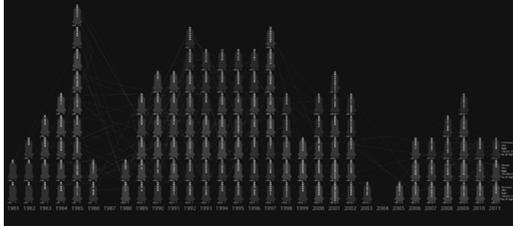
과거 선행 연구³⁾에서 정보시각화 종류를 분석해 보았다. 발표된 다이어그램 중에서, 정보를 효과적으로 표현하고 있는 다이어그램을 분석해 본 결과 8가지의 유형으로 분류되었다. 그 중에서 해부(Anatomy), 타임 시리즈(Time Series)를 삭제하고, 새롭게 많이 사용하는 겹침(Over Layer)과 데이터 구름(Cloud Data)을 포함시켰다. 이상 8개의 시각화 형태를 설명하면 다음과 같다. 본 논문에서 제시하는 예제의 모든 그림은, 본 연구자의 수강생이 제작한 인포그래픽 혹은 정보시각화 콘텐츠 결과물이다.

3-1-1. 데이터 테이블 (Data Table)

가장 많이 사용하는 시각화 형태이다. 일련의 데이터를 기준이 되는 한 축 혹은 두 축의 기준에 맞게 정렬해서 표현한다. 정렬 축을 기준으로 각 데이터를 비교 분석할 수 있는 표현 방법이다. X축, Y축 혹은 Z축의 정렬축의 항목 설정과 테이블에 그려지는 그래픽의 표현 방법이 중요하다. 'Space Shuttle Program'에서는 X축을 기준으로 1981년부터 2011년까지 미국의

3) 최홍석, 김성곤, 디지털 데이터의 동적 질의형 시각화에 관한 연구 : 개발 요소, 표현, 인터페이스 중심으로, 디자인학연구, 제17권, 제2호, 2004, p.440.

우주왕복선의 발사체를 연도별로 표현하였다. 테이블에 그려진 그래픽은 Columbia호, Discovery호와 같은 여러 발사체의 보여주며, 조종사의 남녀 성별비율, 나이 비율, 탑승 승무원의 수를 보여준다.



[그림 4] Space Shuttle Program (함중훈, 2015)

3-1-2. 연결 (Connection)

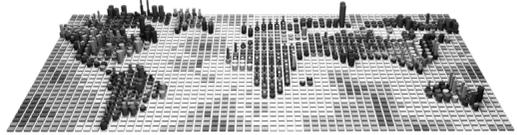
연관성이 있는 데이터 심벌들을 선 혹은 점선으로 연결하거나 테두리로 묶어 표현하는 방법이다. 연결선의 굵기, 색상, 길이, 형태, 그리고 방향성 등으로 여러 의미를 표현한다. 일반적으로 연결된 선의 길이가 짧거나, 두께가 두꺼우면 둘 데이터 사이의 연관성의 높음을 나타낸다. 'Cocktail Recipe'는 칵테일들의 개선되거나 진화된 연관성, 무드별 선택되는 칵테일의 종류, 칵테일의 도수 그리고 맛의 연관 관계를 보여준다.



[그림 5] Cocktail Recipe (윤아령, 2017)

3-1-3. 데이터지도 (Data Map)

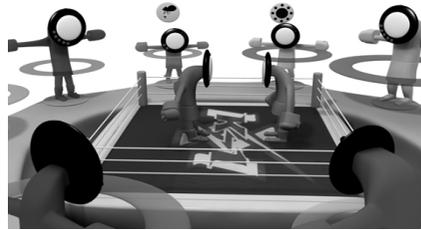
주로 지도를 사용하여 표현된다. 정보를 표현하는 도형 혹은 심벌들이 지도 위에 위치하여, 위치한 좌표를 이용하여 데이터를 비교한다. 지도의 영역 색상이나, 지도 위에 놓인 그래픽 심벌을 사용하여 부가적인 정보를 표현한다. '세계의 이상기온'은 1901년부터 2009년까지 기온의 변화에 따른 해수면 높이, 사막화, 홍수, 그리고 물 부족 현상을 보여준다.



[그림 6] 세계의 이상기온 (김아림, 2010)

3-1-4. 묘사 (Represent)

어떤 형태에 전달 내용의 의미를 담아 시각적 심벌로 표현하는 방법이다. 체르노프 페이스(Faces)⁴⁾가 가장 대표적인 예이다. 다변량 시각화 표현 방법들 중 하나로서 다차원 자료를 이차원 평면상에 표현하기 위해 처음으로 인간의 얼굴을 사용하였다. 묘사는 데이터를 새로운 의미 있는 형태(Meaning Form)로 표현하기 좋다. 'K-1 플레이어'는 승률, 펀치 주특기 종류와 강도, 킥 주특기 종류와 강도, 신장, 체중, 그리고 리치 길이 등의 선수별 능력치를 3차원 캐릭터 이미지에 심벌형태로 표현하였다. 선수들과의 장단점 분석으로 승패를 예상 할 수 있다.



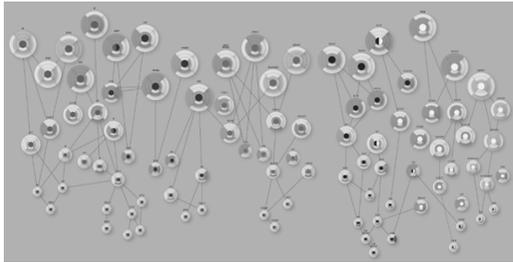
[그림 7] K-1 플레이어 (현광호, 2006)

3-1-5. 계층구조(Hierarchy)

계층구조는 성직자의 계급구조를 나타내는 말에서 유래되었다. 최상위의 한 점에서부터 시작해, 각 하위 레벨에 단계별로 연결되어 만들어진다. 상위 레벨을 Entity 레벨이라 칭하며, 하위 레벨을 Attribute 레벨이라 칭한다. 연결(Connection)은 수평적 관계이고, 계층구조는 상하의 수직적인 관계이다. 주로 그룹화 된 데이터의 조직적 구조를 표현할 때 효과적이다. 'Food Chain and Animal'은 환경 변화에 의해 파괴되는 생태계와 외래종에 의해 사라져가는 동물들의 먹이사슬

4) Chernoff, H., The use of faces to represent points in k-dimensional space graphically, Journal of the American Statistical Association, 1973, Vol. 68, No. 342, pp.361-368.

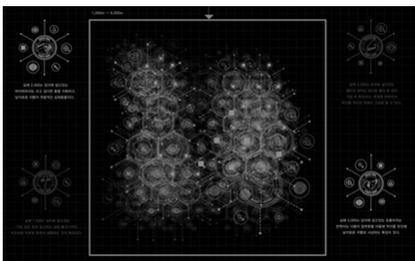
관계를 보여준다. 상위 레벨의 외래종의 등장이나 하위 레벨의 동물의 감소에 따른 생태계의 변화를 보여준다.



[그림 8] Food Chain and Animal (박연현, 2013)

3-1-6. 겹침(Over Layer)

겹침은 Z-thru 매핑기술을 활용한 표현 방법이다. 각각의 레이어 층의 투명도를 조절하여, 사용자는 원하는 내용을 선명하게 볼 수 있고, 더불어 다른 레이어 층의 정보와 비교 할 수 있다. 투명도 변화는 피사체의 초점 조절과 유사하다. 원하는 정보만을 선명하게 보고, 다른 정보는 희미하게 축약하여 본다. ‘해저 생물 분포와 특징’ 심해의 깊이 변화에 따른 생물 분포를 보여준다. 심해를 깊이 별로 레이어로 나누고, 다른 색상으로 표현하였다. 각각의 층에 각 생물체의 특징과 분포의 정도를 표현하였다.

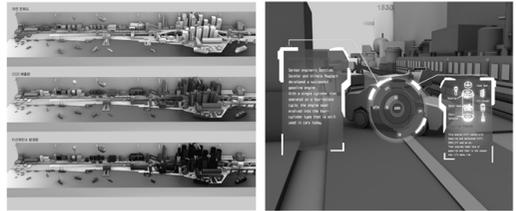


[그림 9] 해저 생물 분포와 특징 (이소민, 2017)

3-1-7. 가상현실(Virtual Reality)

3차원 그래픽을 사용하여 현실감 있게 표현한 것을 가상현실이라 한다. 때로는 사용자의 정보 이해를 돕기 위해, 적절한 형태 변화 혹은 생략 등을 사용한다. 사용자가 망원경 혹은 현미경을 사용해야만 볼 수 있는 자연 현상이나 공간적 혹은 시간적 제약으로 볼 수 없는 과거 혹은 미래 역사 속의 장면 등을 표현 할 때 사용한다. 제작 기간이 긴 단점이 있다. ‘이산화탄소 배

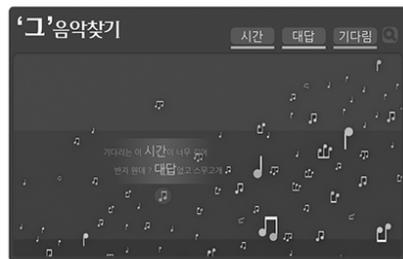
출과 관련 기술은 1900년부터 2010년까지 이산화탄소 배출과 연관 있는 어떤 기술이 등장했고, 얼마만큼의 영향을 주었는가를 그래픽 묘사와 함께 보여준다.



[그림 10] 이산화탄소 배출과 관련 기술 (김병재, 2010)

3-1-7. 데이터 구름(Data Cloud)

데이터 구름에는 문서구름, 태그구름, 숫자구름, 도형구름 그리고 심벌구름 등이 존재한다. 단어구름의 경우, 문서에 사용된 단어의 빈도를 계산하여 시각적으로 표현한다. 많이 나오는 단어는 크게 표시되기 때문에 문서의 핵심 내용을 알 수 있다. 주식시장의 데이터의 경우, 주가의 등락과 거래량을 반영해, 회사명의 크기와 색상을 변화 시킨다. ‘그’ 음악 찾기 SW는 노래 가사 속의 포함된 ‘그’ 단어의 빈도 정도를 음표의 크기로 표현하여 보여준다. 음표의 모양과 색상 또한 상징적 여러 의미를 포함한다.



[그림 11] ‘그’ 음악 찾기 (구진모, 2010)

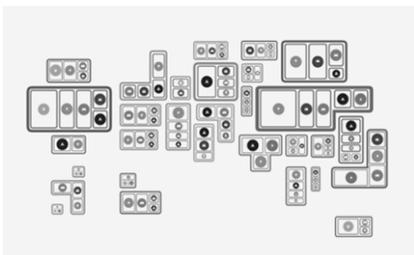
3-2. 수사학적 효과의 종류

시각화 형태는 데이터베이스의 내용에 따라서 대부분 정해져 있다. 그러나 수사학적 효과는 내용의 자세한 설명보다는 내용의 효과적인 전달에 중점을 두어서, 데이터베이스의 내용과는 독립적으로 표현이 가능하다. 수사학을 말의 표현에 관련된 기법을 다루는 학문으로 보는 견해와 삼단 논법과 같은 사고의 논리성과 추론과 관련된 학문으로 보는 견해⁵⁾가 있다. 본 연구에서

는 수사적 표현효과를 연사(言辭)의 수식의 의미로 말과 글을 아름답게 꾸미는 기법에 관한 견해에 국한하여 사용한다. 수사학적 표현에는 대구법, 생략법, 설의법, 대유법, 은유법, 직유법, 열거법, 점강법, 중의법, 의인법, 제유법, 역설법, 문답법, 영탄법, 도치법, 점층법, 풍유법, 반어법, 비약법, 그리고 환유법 등 무수히 많다. 수많은 수사학적 표현 중에서 정보시각화에 응용 가능한 수사법만을 수강자들에게 제시한다.

3-2-1. 역설법(Paradox)

‘이것은 소리 없는 아우성’은 역설법(逆說)이다. 얼핏 보기에는 이치에 어긋난 것처럼 보이면서도 그 속에 진리가 담겨 있게 표현하는 방법이다. 누구나 아는 사물의 크기나 위치를 왜곡되게 표현하여, 데이터 내용을 강조 할 수 있다. ‘The World in Carbon Dioxide Emissions’에서 데이터 맵 형식을 사용하여 각국의 이산화탄소 배출량을 표현하였다. 2011년의 경우 중국의 이산화탄소 배출량이 일본에 비하여 2배가량 많은 것으로 보인다. 대부분의 사용자는 중국의 국토와 일본의 국토 면적의 실질적인 크기와 이산화탄소배출량으로 표현되는 크기를 비교하여 정보를 습득할 것이다.



[그림 12] The World in Carbon Dioxide Emissions (김준목, 2011)

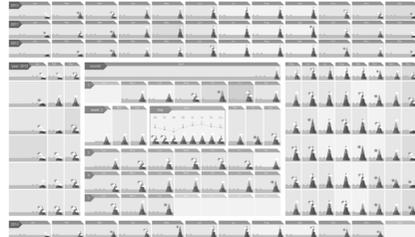
3-2-2. 점강법(Atomization)

점차로 여러개, 작게, 얇게, 낮게 벌려 나가는 법이 점강(漸降)법이다. 부수적인 정보들을 요약하여 보다 작은 공간에 작게 표현하거나, 혹은 축약하여 표현할 수 있다. 투시벽(Perspective Wall)⁶⁾은 대표적인 점강법 효과를 표현할 수 있는 그래픽 표현 방법이다. 사용자가 보고자하는 주요정보를 정면에 배치하고, 나머지 부수적인 정보는 비스듬히 눕혀서 축약하여 표현한다.

5) 박창성, 수사학, 문학과 지성, 2000, p.13.

6) Robert Spence, Information Visualization, ACM press Books, 2001, pp.117-118.

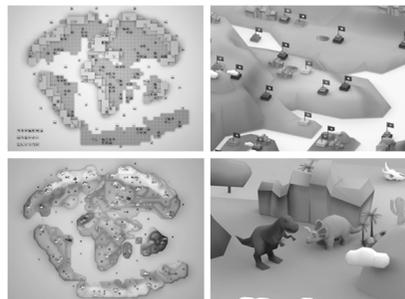
정보의 탐색 UI에 사용하면 유용하다. ‘연간 날씨 다이아리’에는 사용자가 원하는 날씨의 시간별 온도변화와 기상 상태를 알 수 있다. 또한 주간 날씨 변화, 월별 날씨 변화, 분기별 날씨 변화, 그리고 연간 날씨 변화의 평균값을 알 수 있다.



[그림 13] (조예슬, 2014)

3-2-3. 점층법(Parenthesis)

말하고자 하는 내용의 강도를 점차 높이거나 넓혀 그 뜻을 표현하는 기법이 점층(漸層)법이다. 작고 좁은 곳에서 크고 넓은 곳으로 표현을 확대해가는 표현의 방법이다. 시멘틱줌(Semantic Zoom)⁷⁾은 대표적인 점층법 효과를 표현할 수 있는 그래픽 표현 방법이다. ‘시대별 대륙 변화에 따른 공룡 분포에 관한 지도’에서는 시기별로 이동한 대륙의 변화와 등장한 대표적인 생물의 종류를 보여준다. 사용자는 줌 기능을 사용하여 4단계로 확대해 볼 수 있다. 점차적으로 묘사가 구체화된 그래픽으로 탐색 할 수 있다.

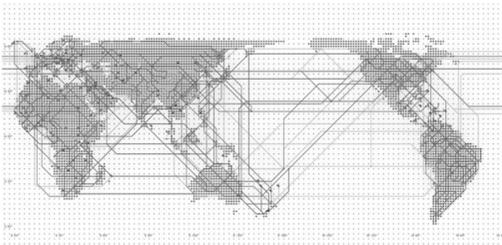


[그림 14] 시대별 대륙 변화에 따른 공룡 분포에 관한 지도 (김지명, 2012)

7) Robert Spence, Information Visualization - Design for Interaction, Second Edition, ACM press Books, 2007, p.119.

3-2-4. 생략법(Omission)

덜 중요한 부분을 생략하고, 나머지는 상상에 맡기는 수사법이 생략(省略)법이다. 어떤 말이 없어도 뜻의 내용이 간결해져서 함축과 여운을 지니게 하는 방법이다. 정보시각화 표현에서 사람의 동맥 혈관과 정맥 혈관의 흐름을 설명할 때, 생략법을 사용하면, 인체의 구조를 그릴 필요 없이 심장을 중심으로 혈관의 형태만을 그려도 사용자는 이해한다. ‘세계 주요 항공기 노선도’는 단순히 비행경로만을 표시하였고, 또한 경로의 두께나 색상으로 비행시간과 고객 수를 표현하였다.



[그림 15] 세계 주요 항공기 노선도 (김현우, 2015)

3-2-5. 제유법(Privative)

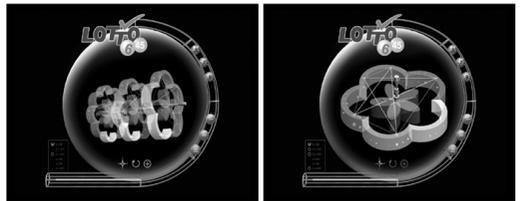
사물의 명칭을 직접 쓰지 않고, 사물의 일부분이나 특징만으로 전체를 나타내는 방법이 제유(提喻)법이다. 부분을 가지고 그 사물 전체를 나타내는 방법이다. 예를 들어 ‘빵만으론 살 수 없다’라는 문장이 대표적인 제유법이다. 빵은 식량을 나타내는 단어이다. ‘유럽 와인 생산지와 생산량은 유럽에서 국가별 와인 원료인 포도의 재배환경과 생산량, 수입량 그리고 소비량을 표현하였다. 주요 포도생산 도시를 하나의 포도 농장으로 표현하였다. 울타리의 모양, 꽃의 개화 정도, 농장 바닥 색상, 저장 탱크 모양, 농가의 집 모양 등으로 도시의 포도 생산과 관련된 데이터를 제유하여 표현하였다. 심벌의 형태로서 사용자에게 관련 정보를 전달하였다.



[그림 16] 유럽 와인 생산지와 생산량 (최시내, 2006)

3-2-6. 환유법(Symbolization)

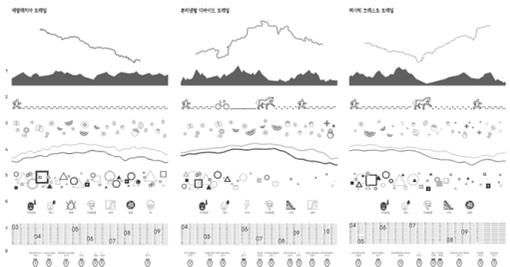
하나의 사물을 다른 명칭을 들어 비유하는 수사법이 환유(換喩)법이다. ‘상아탑’이라는 단어를 대학교에 환유하여 표현한다. ‘로또와 당첨 확률’은 번호별 당첨 확률과 기대치를 보여준다. 꽃모양에서, 암술 크기는 끝자리 패턴을, 꽃잎의 색은 당첨된 번호의 높은 수와 낮은 수 패턴을, 꽃무늬 길이는 로또 번호의 합계 패턴을 그리고 수술의 각도는 짝&홀수 패턴을 표현한다. 만개한 꽃의 모양을 로또 당첨 환희에 환유하여 표현하였다.



[그림 17] 로또와 당첨 확률 (아혜선, 2003)

3-2-7. 두운법(Alliteration)

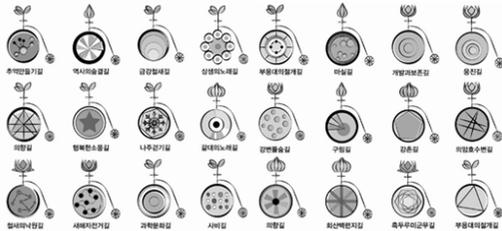
머리 두(頭)를 써서 운율 구절의 첫 부분을 비슷하게 표현하는 수사법이 두운(頭韻)법이다. 중간부분이 반복되는 요운(腰韻)법과 끝부분이 반복되는 각운(脚韻)법을 합쳐 운율(韻律, Meter)법 이라고도 한다. 운율이 일정한 거리를 두고 반복되면서 율격(律格)을 만드는데, 이러한 율격을 통하여 글의 내용을 강조한다. ‘미국 3대 장거리 트레일 분석’은 미국의 유명한 애플래치아 트레일, 콘티넨탈 디바이더 트레일, 그리고 퍼시픽 크러스트 트레일의 온도, 기간, 위험요소, 이동수단, 계절, 그리고 소요시간을 동일한 형식으로 반복하여 표현하였다. 일정한 율격을 통하여 트레일의 내용을 비교하였다.



[그림 18] 미국 3대 장거리 트레일 분석 (김혜인, 2017)

3-2-8. 열거법(Enumeration)

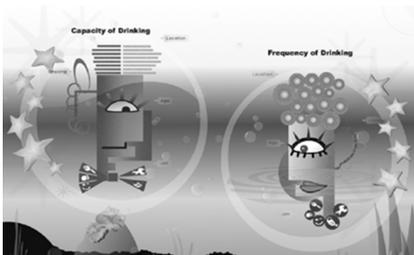
어떤 분류나 계통상 같거나 비슷한 맥락에 있는 것들을 늘어놓는 수사법이 열거(列擧)법이다. 이때, 열거된 각각의 말은 동격이고, 의미상 연관성을 가져야 하며, 그리고 동일한 문장 성분으로 이어져야 한다. 하나 하나가 모여 전체로서의 뜻을 강조한다. ‘한국 자전거길 중주 인증 스탬프’는 국내 주요 24개의 자전거길의 정보를 열거하여 보여준다. 꽃잎이 피는 과정을 5 단계로 나누어 코스 길의 난이도를, 꽃잎의 색상으로 거리를, 바퀴모양으로 코스의 특징을, 그리고 나비의 유무로 무인센터의 존재를 표현한다. 유사한 형태를 연관성을 가지고 열거하여 정보를 표현하였다.



[그림 19] 한국 자전거길 중주 인증 스탬프 (이재은, 2012)

3-2-9. 대구법(Parallelism)

어조가 비슷한 문구를 나란히 두어 문장의 변화와 안정감을 주는 수사법이 대구(對句)법이다. 예를 들어, ‘호랑이는 죽어서 가죽을 남기고, 사람은 죽어서 이름을 남긴다.’라는 문장이 대표적인 대구법이다. ‘남녀 술&담배 소모량’은 국내 음주량의 연도별 변화를 남성과 여성의 모습 변화를 통해 표현하였다. 머리모양에서는 지역별 데이터가, 눈썹 모양에서는 연령별 데이터가, 그리고 옷깃 모양에서는 직업별 데이터가, 사용자의 UI를 통해 선택한 연도에 따라서 표현된다. 대구법은 비슷한 두 내용을 비교 관찰하기 용이하다.



[그림 20] 남녀 술&담배 소모량 (이정균, 2003)

3-2-10. 은유법(Metaphor)

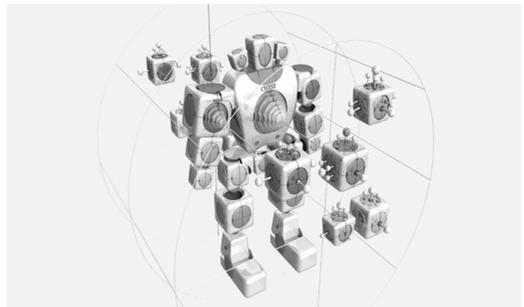
‘A(원관념)는 B(보조관념)이다’와 같이 비유하는 말과 비유되는 말을 동일한 것으로 단언하듯 표현하는 수사법이 은유(隱喩)법이다. 예를 들어 ‘내 마음은 호수.’라는 문장이 대표적인 은유법이다. 원관념은 숨기고 보조관념만 드러내어서 대상을 설명 혹은 특징을 표현한다. ‘세계 각국의 맥주 맛 비교’는 주요 맥주의 생산 지역, 맥주 분류, 알코올 농도, 칼로리, 그리고 향에 관한 정보를 보여준다. 특히 맥주의 꽃, 과일, 커피, 초콜릿, 그리고 캐러멜 등의 다양한 맛을 색채 데이터에 은유하여 표현하였다.



[그림 21] 세계 각국의 맥주 맛 비교 (조재형, 2013)

3-2-11. 직유법(Simile)

두 개의 사물을 직접적으로 비교하여 표현하는 수사법이 직유(直喩)법이다. 내포된 비유가 아닌, 겉으로 정확히 구체적으로 드러내는 비유 방법이다. ‘A는 B와 같다’ 또는 ‘A와 같은 B’ 하는 식의 비유적 표현 방법이다. ‘헬스 운동기구의 근육 생성량’은 헬스 운동기구 종류별, 운동량 세트 횟수별, 그리고 운동의 강도에 따른 근육 생산량을 보여준다.

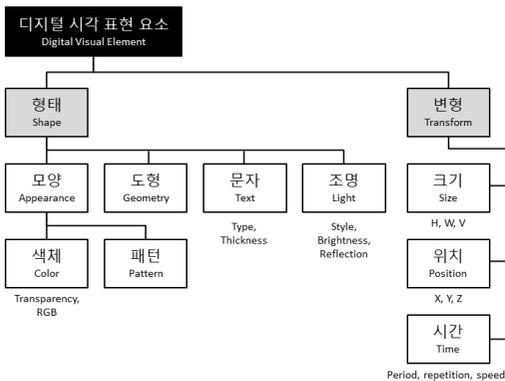


[그림 22] 헬스 운동기구의 근육 생성량 (이지연, 2005)

조이스틱 모양의 손잡이를 조절하여, 각각의 헬스 운동기구의 운동 세트 횟수와 강도를 설정 할 수 있다. 운동을 통하여 생성된 근육의 양에 비례하여 로봇처럼 생긴 강철 근육이 앞으로 돌출된다. 로봇처럼 생긴 강철 근육을 근육 생산량 데이터에 직유 표현하였다.

3-3. 시각 표현 요소의 종류

시각화 형태를 선택하고, 효과적인 전달을 위한 수학적 표현을 선택한 후, 그래픽 묘사를 위해서 구체적인 데이터를 표현할 시각적 요소를 설정해야 한다. 데이터는 실수와 정수로 표현되는 숫자(Numerical) 데이터와 요일이나 서수를 표현되는 분류(Ordinal) 데이터 그리고 카테고리(Categorical) 데이터의 형식으로 구성된다.⁸⁾ 컴퓨터에서 구현되는 그래픽은 이진법으로 표현되는데, 시각 표현 요소 또한 이진법으로 표현되어야 한다. 씬 그래프(Scene Graph)⁹⁾ 시스템은 이러한 멀티미디어 그래픽을 표현하기 위한 대표적인 시스템이며 다음과 같은 구조를 가진다.



[그림 23] Digital Visual Element for Information Data Representation

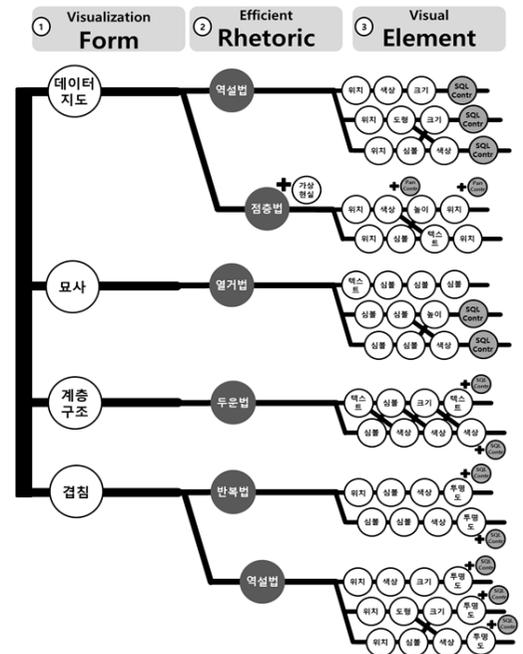
씬그래프 시스템은 형태요소와 변형요소로 이루어져 있다. 형태요소(Shape)는 RGB와 투명도로 표현되는 '색채'와 사용자가 만든 여러 무늬의 '패턴'을 포함한 모양과, 정사각형, 삼각기둥, 그리고 원과 같은 기하학적 '도형', 나뉠 고딕, 맑은 명조와 같은 서체와 크기로 표

8) Spence, Robert, Op. cit., pp.4-5.

9) OpenSG, Open Inventor, 그리고 SGL 등과 같은 C++ 기반의 그래픽 오픈 소스 프로그래밍 시스템. 씬 그래프가 관리하는 데이터는 노드 개념을 통하여 그래프로 구조화 된다.

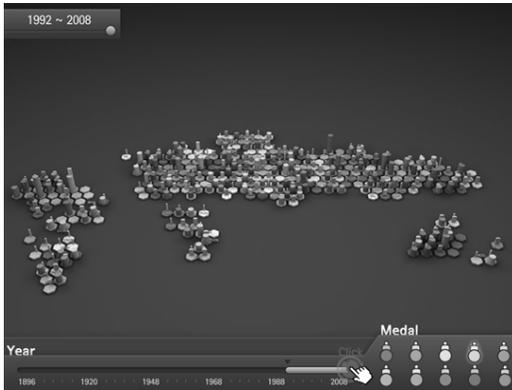
현되는 '문자' 그리고 '조명'으로 구성된다. 변형요소(Transform)는 높이, 길이, 폭, 볼륨 등의 스케일 개념이 포함된 '크기', 그래픽 3차원 공간속의 좌표 값인 '위치' 그리고 애니메이션에 필요한 '시간'으로 구성된다. 데이터의 형식에 따라서, 분류 데이터는 숫자의 조합과 범위가 제한적인 '색채'와 '조명'으로 표현이 가능하고, 카테고리 데이터는 종류가 다양한 '문자', '패턴'으로 표현이 가능하며, 그리고 숫자 데이터는 '도형'의 형태, '크기', '위치', 그리고 애니메이션 '시간' 등으로 표현이 가능하다. 또한 시각 표현 요소에는 UI를 활용할 컨트롤러가 포함된다. 데이터베이스에서 SQL 조건식을 실행 명령할 'SQL Controller', 화면을 좌우상하로 움직일 'Pan Controller', 화면을 'Z'축을 이용하여 위아래로 움직일 'Zoom Controller', 걸어가면서 보는 시점으로 변환할 'Navigation Controller', 애니메이션을 제어할 'Time Controller', 그리고 사용자가 마음대로 시점의 축을 변환시킬 수 있는 'View Controller'가 포함된다.

4. 의사결정나무 이론의 활용



[그림 24] An Example of Contents Generation Tree Using Olympic Games Contents

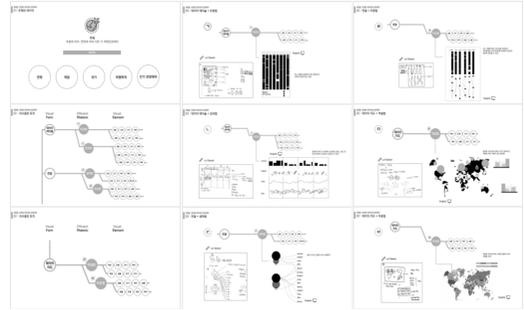
위의 그림은 의사결정나무를 응용한 VCGT의 활용 예이다. '국제 올림픽'을 주제로 국가별, 연도별, 그리고 종목별 금, 은, 동메달 수를 보여주는 정보시각화이다. 수강생들에게 예제로서 제공하기 위해 연구자가 만들었다. 여러 변수 중에서, '시각화 형식', '수사학적 효과' 그리고 '시각 표현 요소' 등의 변수를 선택 사용하였다. 선택된 변수를 '분류'와 '세분화' 방식에 응용한 VCGT의 활용 예이다. '국제 올림픽'에 사용한 정보시각화 콘셉트 콘텐츠로 총 15가지의 가지치기(Branch)가 생성됐다. 샘플 VCGT에는 2번째 단계 '가지'로 '데이터 지도', '묘사', '계층구조' 그리고 '겹침'의 네 종류를 선택하였다. 그리고 3번째 단계 '가지'로 '데이터 지도 - 역설법', '데이터 지도 - 점층법', '묘사 - 열거법', '계층구조 - 두운법', '겹침 - 반복법', 그리고 '겹침 - 역설법' 여섯 종류를 선택하였다. '시각화 형식'으로 '데이터 지도'를 선택하고, '수사학적 효과'로 '역설법'을 선택하며, '국가는' '위치 좌표', '종목은' '색상', '메달 수는' '크기' 그리고 연도는 'SQL Controller'를 사용한 '가지'가 그림의 첫 번째 콘셉트 콘텐츠이다. 이를 구체화한 정보시각화가 아래의 그림이며, 수강생들에게 구체화된 예로 보여주고 있다.



[그림 25] 1896-2008년 올림픽 메달 수 (이은경, 2010)

실제로 3번째 단계 '가지'까지가 가지마다 차별성이 있고, 4번째 단계 '가지'부터는 세부적인 내용이어서 차별성이 적다. 강의에서는, 3번째 '가지'까지의 내용으로 수강생들에게 구체적인 아이디어 스케치 과제를 제시한다. 아래 그림은 수강생이 제작한 아이디어스케치의 예이다. 수강생은 주제를 '계절에 따라, 연령에 따라, 선호하는 세계 관광지'로 선택하였다. 주요 데이터로 '연령', '계절', '국가', '여행목적', 그리고 '인기 관광지'를 설정했다. 그리고 2단계와 3단계 가지치기를 통하여

여섯 가지의 콘셉트 콘텐츠를 정했다. 구체적인 콘셉트를 유사한 형태의 인포그래픽과 아이디어 스케치를 사용하여 설명하고 있다. 6가지의 아이디어 중에서 콘셉트에 적합한 방식은 무엇인가를 판단해야 한다. 이를 위해 적합성 평가가 필요하다. VCGT 적합성 평가에서는 다음의 사항을 염두 해야 한다.



[그림 26] An Student's Idea Sketch Example

우선 정보시각화의 사용자와 사용 환경을 명확히 설정해야 한다.¹⁰⁾ 사용 미디어, 사용자의 연령대, 사용자의 교육 정도, 학습시간, 정보 탐색 시간, 그리고 사용 장소를 설정해야 한다. 다음으로 단계별 설정 기준을 마련한다. '시각화 형식', '수사학적 효과' 그리고 '시각 표현 요소'의 각각의 가지마다 단계에 얼마의 가중치를 두고 평가할 것인가를 설정한다. 예로서 제시한 '국제 올림픽'의 경우, '시각화 형식'의 경우 40점, '수사학적 효과'의 경우 35점 그리고 '시각 표현 요소'의 경우 25점을 할당했다. 일차적으로 '시각화 형식'이 전반적으로 주제에 가장 부합되고, 이차적으로 '수사학적 효과'가 잘 선택 포함되어야 가치 있는 정보시각화 콘셉트로 평가된다. '국제 올림픽'의 15개 가지치기의 적합성평가 결과는 최고 90점에서 최저 55점으로 평가 되었다. 적합성 평가가 상위 20% 이상의 결과가 나온 가지 중에서 하나를 선택하여 구체적인 시각화 작업을 한다.

5. 결론

정보시각화 강의에서 두 가지의 어려움이 있다. 첫째는, 학생들이 정보시각화에 대한 제작 경험의 부족이

10) 김성곤, 데이터베이스 자료 시각화 방법과 그 표현 언어들에 관한 연구, 디자인학연구, 2000, 제13권, 제3호, p.196.

다. 단순한 막대그래프, 파이그래프 혹은 좌표축을 가진 포지션 맵(Position Map) 정도의 지식을 가지고 있다. 둘째는, 정보시각화 콘셉트 개발 과정에서 고려해야 할 점이 많아서, 수강생들은 무엇을, 어떠한 과정으로, 어떻게 해야 하는가에 대한 어려움이 있다. 수강생들에게 정보시각화 분야를 소개하고, 정보시각화 콘텐츠 기획 과정을 짧은 시간에 이해시키기 위해 의사결정나무를 응용한 VCGT 교육모델을 제작했다.

디지털콘텐츠디자인 학부생을 대상으로 VCGT 교육 모델을 다음과 같은 3가지 경우로 각각 교육해 보았다. 학생은 약 100여명이며, 개별 과제로 진행했다. 첫 번째 경우는, 본 연구자가 시각화 주제를 제시하지 않고, 학생들이 자료조사를 하고, 이를 바탕으로 주제를 선택하는 것이고, 두 번째 경우는, 본 연구자가 과거 수강생 선배들이 과제 진행에서 성공적으로 정보시각화 콘텐츠 기획이 가능했던 35개의 주제를 선별하여 제시하는 것이며, 세 번째 경우는 ‘국제 올림픽’, ‘스포츠 야구’ 그리고 ‘항수의 종류’ 등 3가지 주제만 활용하여 콘텐츠를 기획하는 것이다. 첫 번째 경우는 수강생들이 주제 선별에 너무 많은 시간을 보냈으며, 또한 사용할 데이터가 그래픽 표현에 적합하지 않는 경우가 많았다. 데이터베이스 개념이 부족해서, 학생들은 주제 속에 데이터 얼마나, 어떻게 존재하는지에 관한 지식이 부족했다. 세 번째의 경우는 제작된 콘텐츠 기획 내용이 수강생들 사이에 차이가 적었다. 다른 수강생이 개발한 내용을 바탕으로 비교 학습하는 다양성 학습이 부족했다. 결론적으로 강의자는 학생들에게 정보시각화 콘텐츠 기획 학습에 적절한 주제를 미리 선별하여 학생들에게 제시하는 것이 좋다.

VCGT 교육과정에서 유의해야 할 두 가지가 있다. 첫째는, 수강생들에게 ‘시각화 형식’과 ‘수사학적 효과’에 대한 충분한 사전 교육이 필요하다. 이를 위해, 강의 전반부 5주차 이전에 수강생들에게 수백 개의 예를 보여주며, 형식과 효과에 대하여 반복적으로 설명해야 한다. 둘째는 적합성 평가의 가중치 설정과 평가 합에 수강생들은 많은 관심을 가진다. 수강생들에게 평가의 합은 콘텐츠 기획의 전반적인 평가 혹은 선택을 돕기 위한 방법이지, 절대적인 해결안이 아님을 강조한다. 수강생들에게 최대한 많은 데이터를 요약(Summarize)

하고, 그 형식을 단순(Simple)하게 표현하며, 또한, 사용자가 내용에 흥미를 가질 수 있게 묘사(Describe)하고 그리고 콘텐츠 속의 데이터 비교를 위한 탐색(Explore)이 가능한 정보시각화 콘텐츠 기획을 선택하게 한다.

의사결정나무를 응용한 VCGT 교육모델을 정보시각화 콘텐츠 기획 교육에서 수월성과 다양성을 학습하기 위한 방법으로 제시하였다. 향후, VCGT 교육모델을 발전시키기 위해, 보다 다양한 연령대와 직업군을 대상으로 교육 실험할 것이다.

참고문헌

1. 박창성, 수사학, 문학과 지성, 2000.
2. Robert Spence, Information Visualization - Design for Interaction, Second Edition, ACM press Books, 2007.
3. 김성곤, 데이터베이스 자료 시각화 방법과 그 표현 언어들에 관한 연구, 디자인학연구, 2000, 제13권, 제3호.
4. 최홍석, 김성곤, 디지털 데이터의 동적 질의형 시각화에 관한 연구 : 개발 요소, 표현, 인터페이스 중심으로, 디자인학연구, 2004, 제17권, 제2호.
5. Chernoff, H., The use of faces to represent points in k-dimensional space graphically, Journal of the American Statistical Association, 1973, Vol. 68, No. 342.
6. Kim, Ki Sook, Kim, Kyung Hee, A Prediction Model for Internet Game Addiction In Adolescents: Using a Decision Tree Analysis, Journal of Korean Academy of Nursing, 2010, Vol.40, No.3.
7. Quinlan, J. R. Induction of decision trees, Machine Learning, 1986, Vol.1, Issue 1.