

텍스트 데이터 시각화의 표현 재료와 접근 방식에 관한 고찰

A Review on Expressive Materials and Approaches to Text Visualization

김효영, 박진완
중앙대학교 첨단영상대학원

Hyoyoung Kim(the.kimyo@gmail.com), Jin Wan Park(jinpark@cau.ac.kr)

요약

본 연구에서는 데이터 시각화 연구의 한 분야인 텍스트 시각화에 대하여 그 시각적 표현의 재료가 되는 텍스트 데이터의 종류 및 본질, 특성에 대하여 고찰하고, 다양한 텍스트 시각화의 사례 연구를 통해 텍스트 데이터의 시각화를 위한 표현적 접근 방식에 대하여 다각적으로 분석하였다.

텍스트 시각화 연구는 컴퓨터의 발달과 방대한 데이터의 공개, 그리고 시각화 툴의 범용화 등의 기류를 타고 급속도로 확산되고 있으며, 이에 따라 공학, 예술, 인문, 사회 등 학제 간 융합 연구를 통해 다양한 작품 또는 연구 성과물로서 창작되고 있다. 그럼에도 불구하고 텍스트 데이터를 비롯한 데이터 시각화에 관한 이론적 고찰과 접근 방식에 대한 체계적 분석은 거의 이루어지지 않고 있는 실정이다. 데이터는 이해와 해석의 대상이며, 가공 및 접근에 따라 무한한 정보와 가능성을 갖는다. 데이터의 중요성이 점차 증가하고 있는 현 시점에서 데이터의 이해와 해석에서 출발하는 텍스트 시각화라는 융합 학문적 연구 분야는 미래 인간 사회에서 데이터가 차지할 위상을 고려할 때, 보다 체계적인 연구와 이론적 축적을 필요로 한다.

■ 중심어 : | 텍스트 시각화 | 데이터 시각화 | 정보 예술 |

Abstract

In this study, we contemplated types, essence, characteristics of text data which is material for visual expression of text visualization part of data visualization research and also analysed the multidirectional means of expressive approach for it.

Studies of text visualization are spread dramatically under the influence of computer development, open data, wide use of visualization tools, etc. For these reasons, text visualization works have been creating as art works or output of research through various inter-discipline convergent research with engineering, art, humanities, sociology, etc. Nevertheless the theoretical studies on text data itself and its visualization, and also systematic analysis of its approach are rarely made. Data is target of understanding and interpretation, and it has infinite information and possibility with process and approach for it. Considering the attainable status of data in future human society, text visualization which is convergent academic field of study starting with understanding and interpretation of data needs further methodological research and theoretical accumulate.

■ keyword : | Text Visualization | Data Visualization | Information Art |

* 이 논문은 2010년도 중앙대학교 성적우수장학금과 2012년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2012S1A5A2A01020337)

접수번호 : #121109-003

접수일자 : 2012년 11월 09일

심사완료일 : 2013년 01월 09일

교신저자 : 박진완, e-mail : jinpark@cau.ac.kr

I. 서론

문자의 발명과 발전은 문명의 발전과 지식의 팽창을 가져왔으나 현대에 이르러 그 데이터의 양은 점차 방대해지고 거대해져 인간의 습득 수준을 넘어서게 되었다 [1]. 실제로 신문, 서적 등으로 대표되는 텍스트는 그 양이 방대해질 경우 독자가 전체적인 주제와 내용 및 그 데이터가 갖는 관계 등을 파악하기가 매우 어렵게 된다. 이러한 맥락에서 텍스트를 구성하는 개별 문자 자체의 의미 이외에 전체 텍스트가 갖는 여러 가지 정보를 파악하기 위해서는 텍스트 데이터의 분석이 요구된다. 이렇게 도출된 정보 및 데이터간의 관계를 사용자들에게 손쉽게 직관적으로 전달하기 위해서는 문자나 숫자 이외의 더욱 직관적인 형태로 표현되어야 할 필요성을 갖는데, 이를 위한 접근이 텍스트 시각화 연구이다.

이러한 텍스트 시각화 연구는 최근 들어 더욱 급속도로 이루어지고 있는 컴퓨터의 발달과 Open API를 이용한 방대한 데이터의 공개, 그리고 시각화 툴의 범용화 등의 기류를 타고 급속도로 확산되고 있으며, 이에 따라 공학, 예술, 인문, 사회 등 학제 간 융합 연구를 통해 다양한 작품 또는 연구 성과물로서 창작되고 있다. 본 논문에서는 이러한 텍스트 시각화의 다양한 사례들을 바탕으로 텍스트 시각화의 재료와 표현을 위한 접근 방식에 대하여 고찰해보고자 한다.

II. 텍스트 시각화의 재료와 표현

1. 텍스트 시각화의 재료

텍스트 시각화(Text Visualization)는 데이터 시각화(Data Visualization) 연구의 한 분야로, 방대한 데이터 종류 중에서도 텍스트 데이터를 시각화의 소재로 활용한 경우를 일컫는다. 기본적으로 텍스트 데이터는 서적, 뉴스, 신문, 간행물, 웹 사이트 등 텍스트로 이루어진 모든 데이터를 의미하는데, 이러한 텍스트 데이터의 경우 다른 데이터들과는 구별되는 독특한 특성을 갖는다. 이는 바로 텍스트 자체의 속성인 ‘형식’과 ‘내용’이다. 여기서 텍스트의 형식은 음운, 음절, 형태소, 단어, 어절,

구, 문장, 문단 등으로 대표되는 글의 형태론적 특성을 의미하는데, 텍스트 시각화의 경우 이러한 텍스트의 모든 형식적 요소가 모두 시각화의 재료가 될 수 있다.

뿐만 아니라 텍스트의 또 다른 속성인 ‘내용’ 또한 텍스트 시각화의 주된 소재로 활용된다. 텍스트의 내용 데이터는 예를 들면 소설의 줄거리, 뉴스의 주제, 블로그 게시글의 인기 키워드 등을 포함할 수 있다.

이렇듯 텍스트 시각화의 소재는 텍스트의 형식 그 자체, 그리고 내용을 이루는 텍스트의 구조뿐만 아니라, 개별 텍스트가 갖는 고유의 의미를 비롯하여 텍스트들이 이루어 창출해내는 콘텐츠 등 그 범위가 매우 방대하다. 이러한 텍스트 시각화의 재료들은 시각화의 목적, 즉 무엇을 표현하고자 하느냐에 따라서 선택되며, 이를 표현하기 위한 시각화 접근 방식 또한 매우 다양하게 나타난다.

2. 텍스트 시각화의 표현

텍스트 시각화의 사례를 분석해보면, 텍스트 시각화의 소재로서 가장 많이 활용되는 데이터 중의 하나가 바로 성경(Bible)의 텍스트이다. 성경은 전 세계 판매 부수 1위 서적임과 동시에 그 내용 또한 방대하여 시각화의 주된 소재로 활용되어왔다. 그런데 이러한 성경 시각화의 여러 사례를 분석해보면, 동일한 텍스트, 즉 동일한 데이터를 재료로서 활용함에도 불구하고 시각화의 목적이나 작가의 의도에 따라 매우 다양한 표현 방식을 갖는다.

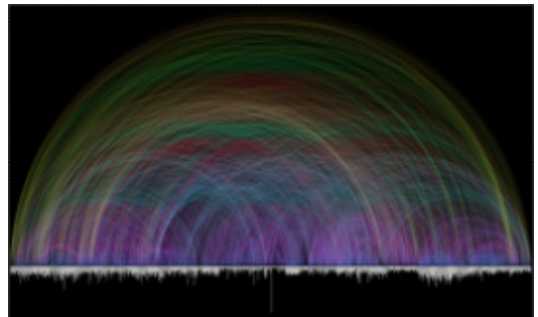


그림 1. Bible Cross-References(2008)
/ Chris Harrison, Christoph Romhild

[그림 1]은 성경 내에 교차적으로 언급되는 63,000여 개의 교차 언급을 막대(bar)와 호(arc)를 이용하여 시각적으로 표현한 ‘Bible Cross-References’[2]이다. 이 작업은 성경 텍스트가 갖는 내용 정보의 구체적 제시가 아닌, 성경이 갖는 텍스트 구조를 정해진 규칙에 의하여 나열한 시각화라고 볼 수 있다. 그러나 신약과 구약의 구분 및 교차 언급의 빈도에 의해 판단 가능한 각 권의 연관성을 효율적으로 표현하고자 하는 의도라고 보기에는 교차 언급에 대한 정보를 알아차리기 쉽지 않은 표현 방식을 지닌다. 따라서 본 시각화는 성경의 구조와 문장 자체를 시각화의 소재로 활용하되, 정보 전달의 기능적 측면보다는 심미적 목적에 초점을 둔 작업이라 볼 수 있다.

반면 [그림 2]의 ‘Dispensational Chart’[3]는 예수가 그의 일생동안 했던 말(speech)들을 살펴볼 수 있도록 웹 기반으로 제작한 인터랙티브 시각화 작업으로, 이용자들에게 구체적 성경 구절을 직접 제공하고자 하는 목적으로 제작된 실용적 시각화로 볼 수 있다.



그림 2. Dispensational Charts/ And Dang

또 다른 성경 시각화의 사례로 ‘Mapping the New Testament’[4][그림 3]을 들 수 있다. 이는 IBM사의 Collaborative User Experience 연구 그룹에서 운영하는 ‘Many Eyes’라는 일종의 데이터 시각화를 위한 툴과 유저 인터페이스를 제공하는 웹 사이트에 소개된 작업의 하나로, 성경의 신약전서에 등장하는 인물들 간의 사회적 관계를 시각화하였다. 실제로 인물 간의 사회적 관계는 성경 문장 내에서 이웃하여 등장한다고 해서 가까운 것이 아니다. 한 예로 구약서에 등장하는 요셉(Joseph)이 신약의 예수(Jesus)와 동 시대에 존재하지 않음에도 불구하고 사회적 관계망으로 연결될 수 있는 것이다. 이러한 시각화 작업은 성경 텍스트 데이터에

직접적으로 드러나지 않은 숨은 데이터 간의 관계를 나타내고자 하는, 앞선 성경 시각화 사례들과는 또 다른 시각화 목적을 갖는다.

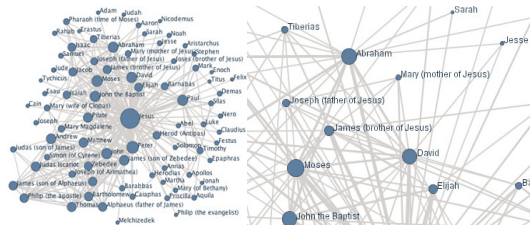


그림 3. Mapping the New Testament(2006)/crossway

지금까지 살펴본 동일한 성경 텍스트에 대한 다양한 시각화 사례를 통해서도 알 수 있듯이, 시각화의 목적에 따라 텍스트 데이터의 선택과 시각적 표현을 위한 접근 방식이 달라진다. 이러한 점에 착안하여 다음 장에서는 다양한 텍스트 시각화의 사례 분석을 통해 텍스트 시각화를 위한 구체적인 접근 방식에 대해 고찰해보고자 한다.

III. 텍스트 데이터 시각화의 접근 방식

1. 정보의 직관적 전달

데이터 시각화는 복잡하고 거대한 데이터 집합을 시각적으로 표현함으로써 그 데이터 내부에 숨은 정보를 보다 쉽게 파악하고자 하는 목적에서 출발한다. 텍스트 시각화의 일차적 목표 또한 인간의 시각 지능을 이용한 시각적 자극의 형태로 텍스트 데이터를 표현함으로써 방대한 텍스트 데이터가 갖는 다양한 정보 - 텍스트의 형식 및 내용 정보, 데이터 패턴, 관계, 트렌드 등 - 를 직관적으로 전달하는 것이다.

1.1 텍스트 형식 및 구조 정보

텍스트는 그 데이터의 특성상 텍스트의 형식과 구조에 대한 정보를 갖는다. 따라서 텍스트 시각화 연구에서는 이러한 텍스트의 형식 또는 구조 자체를 집약하여

가시화하기 위한 시도들이 지속되어왔다.

‘Textual Visualization based on Readability’[5][그림 4]는 2011년 ACM Siggraph Asia에 발표된 연구로, 서적의 문장 가독성을 시각화하는 방법을 제안하였다. 이를 위해 문장의 가독성을 판별하는 여러 요인들에 대한 이론적 조사를 통하여 문장의 ‘길이’ 속성이 핵심 요인임을 도출한 후, 한 문장이 시작되면서부터 마침표로 끝이 나는 지점까지 문장의 길이에 비례하여 명도가 낮아지도록 설계하여 표현하였다. 이러한 시각화 결과물을 통하여 서적 텍스트를 직접 읽지 않고도 책의 가독성 정보를 직관적으로 파악할 수 있다.

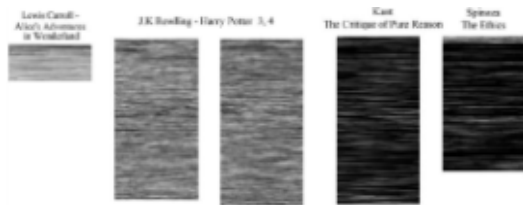


그림 4. Textual Visualization based on Readability (2011)/ Hyoyoung Kim, Jin Wan Park

또 다른 텍스트 구조 시각화 사례로서 ‘The Evolution of The Origin of Species’[6][그림 5]를 들 수 있다.

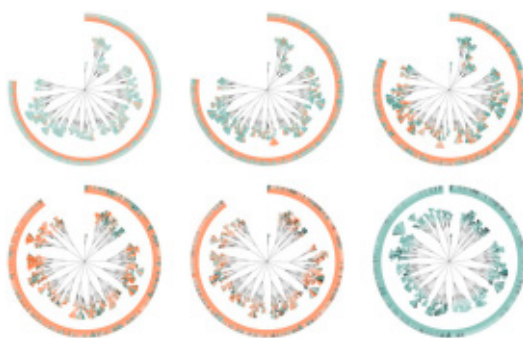


그림 5. The Evolution of The Origin of Species (2009)/ Stefanie Posavec, Greg McInerney

영국의 생물학자 찰스 다윈(Charles Robert Darwin)의 대표 저서인 ‘종의 기원(The Origin of Species)’의 1판(1st edition)부터 6판(6th edition)까지를 시각화 한

이미지가 순차적으로 나열되어 있다. 한 텍스트가 갖는 각 단락(paragraph)들을 꽃잎으로 표현하였고, 바깥쪽 원형 테두리에 둘러있는 블록들은 책이 갖는 각각의 한 문장을 의미한다. 각 블록은 다음 판(edition)과 비교했을 때 그대로 존재하는 문장일 경우 푸른색으로, 삭제된 문장일 경우 주황색으로 구분하여 표현하였다. 이러한 시각화를 통해 매번 판의 수가 증가할 때마다 문장의 길이가 짧아졌으며, 각 판별로 바로 이전 판에 비하여 문장의 개정이 얼마나 이루어졌는지에 대한 정보를 알 수 있다.

1.2 텍스트 내용 정보

텍스트는 형식 및 구조 정보 외에도 내용에 관한 정보를 갖는다. 실제로 데이터이기 이전의, 커뮤니케이션 수단으로서의 텍스트를 바라볼 때 가장 중요한 것은 그가 갖는 의미, 즉 내용 정보라고 할 수 있다. 그러나 서적이거나 신문 기사 등 그 데이터의 양이 거대해질 경우, 그 의미 또는 내용의 흐름을 파악하는 것이 매우 어려울 뿐만 아니라 그러한 정보를 획득하기 위해서는 상당한 시간이 소요 된다. 이러한 이유로 방대한 텍스트가 갖는 내용 정보의 손쉬운 파악을 위한 수단으로서의 시각화 접근이 다양하게 이루어지고 있다.

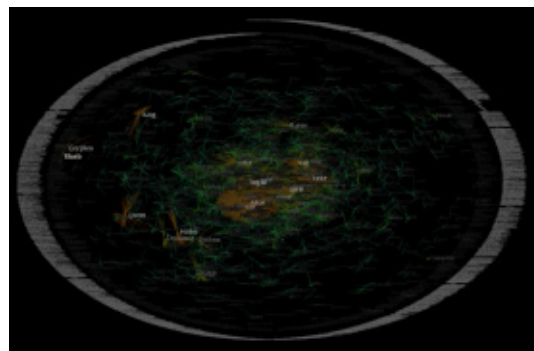


그림 6. TextArc: Alice's Adventures in Wonderland (2009)/ W. Bradford Paley

텍스트의 내용을 시각화하는데 가장 많이 활용되는 방법은 텍스트 데이터를 단어로 분리하여 그 출현 빈도를 가시화하는 것이다. [그림 6]은 텍스트 시각화의 가

장 유명한 사례 중의 하나로 꼽히는 W. Bradford Paley의 'TextArc'[7]이다. 이는 판타지의 고전인 영국 작가 Lewis Carroll의 소설 '이상한 나라의 앨리스(Alice's Adventures in Wonderland)'의 텍스트 데이터를 단어 단위로 가공하여 그 출현 빈도수에 따라 해당 글자의 크기가 커지도록 함과 동시에 전체 텍스트 구조 내에서 단어가 출현하는 위치 정보까지 반영함으로써 소설 내에서 함께 등장하는 등장인물의 이름이 서로 가깝게 나타나도록 설계하였다. 이를 통해 도출된 결과 이미지를 살펴보면, 중앙에 'Alice'라는 단어가 크게 자리 잡고 있음을 볼 수 있는데, 이는 Alice가 주인공의 이름으로 책 전반에 걸쳐 지속적으로 등장하기 때문이다. 또한 책 속에 등장하는 다양한 캐릭터들은 위치와 색깔을 달리하며 각각의 특성과 관계들을 시각적으로 표출해낸다. 예를 들면 해터(Hatter)와 도어마우스(Dormouse), 3월의 토끼(March Hare)는 비슷한 글자체로 가까이 위치하고 있어 주로 함께 등장함을 짐작할 수 있는 것이다.

이와 유사한 시각화 사례로서, [그림 7]은 2010년 NordiCHI 학회에 발표된 논문 'Visualizing the Text of Philip Pullman's Trilogy "His Dark Materials"[8]에서 제안된 서적 시각화의 결과 이미지를 보여주고 있다. 본 시각화 연구에서는 Phillip Pullman의 3부작 소설 'His Dark Materials'의 텍스트를 그대로 나열한 후, 두 남녀 주인공의 이름인 Lyra, Will이 나타나는 지점에 각각 파란색과 빨간색의 점을 찍어 이를 곡선으로 연결하였다.

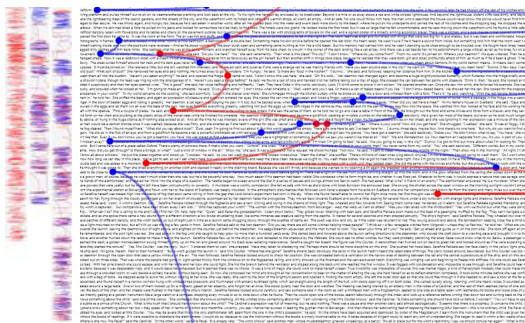


그림 7. Visualizing the Text of Phillip Pullman's Trilogy "His Dark Materials"(2010)/ Tim Legan, Linda Baker

[그림 7]의 왼쪽에 위치한 이미지는 3부작의 1, 2, 3권 각각에 대한 시각화 결과물로, 이 일련의 이미지를 통해 소설의 1권에서는 남자 주인공이 등장하지 않은 채, 여자 주인공에 대한 묘사 및 서술로 진행되고 있다는 점과, 두 주인공이 모두 등장하지 않는 공백 부분에서는 다른 등장인물(Mary Malone, Lord Asriel, Mrs. Coulter 등)에 관한 서사가 이루어지고 있음을 파악할 수 있다.

이렇듯 텍스트 시각화에 있어서 텍스트 내용 및 흐름 파악을 위한 접근은 거대한 텍스트 데이터의 주요 내용을 직관적으로 파악하는데 용이하게 활용될 뿐 아니라, 텍스트를 직접 접하지 않고도 일정 수준의 정보를 획득하는데 도움을 줄 수 있다.

1.3 텍스트 데이터의 패턴, 관계, 트렌드 정보

앞서 제시한 텍스트의 형식 또는 내용과 같이 텍스트 데이터가 직접적으로 갖는 정보 이외에도 거대한 텍스트 데이터의 집합은 다양한 패턴, 트렌드 또는 데이터와 데이터간의 관계에 관한 정보를 갖는다. 이러한 정보들은 시각적 요소를 통해 가시화되어야만 그 존재나 의미를 드러내는 경우가 많다. 따라서 이러한 복잡한 데이터를 정제하여 그 안에 숨은 정보를 파악하기 위해 시각화는 필수적으로 거쳐야 하는 관문이라고 볼 수 있다.

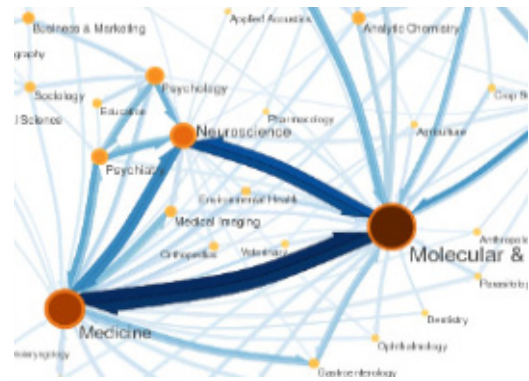


그림 8. Map of Science(2007) / Martin Rosvall, Carl Bergstrom

[그림 8]은 2004년 Thomson Reuters's Journal Citation Reports가 집계한 6,128개의 과학 관련 저널에

인용된 6,434,916개의 인용 데이터 간의 상호 관계를 지도화하여 나타낸 것이다[9][10]. 문헌의 인용은 학문 관련 출판에 있어서 보편적 관행임과 동시에 학계에서 저서의 인기도와 신뢰도를 측정하는 중요한 수단이기도 하다. 뿐만 아니라 각 주제간 유사성의 관계를 확인할 수 있는 정확한 단서가 될 수 있다. 만약 서로 다른 두 작업이 제 3의 작업에서 인용되었다면, 이는 비록 서로 간의 인용이 존재하지 않더라도 두 작업의 연관성은 존재한다고 볼 수 있는 것이다. 따라서 이러한 인용 데이터의 적절한 시각화를 통해 학문간, 주제간 연관성의 유무 및 정도를 파악할 수 있으며, 이는 학문적으로 매우 중요한 정보가 된다. 이러한 정제된 정보는 각 데이터 노드(node)간에 존재하는 연관성 데이터라는 또 다른 데이터를 시각적 요소로 변환하여 나타내는 과정을 통하여 비로소 정보로서의 역할이 가능하다.

유사한 시각화의 예로, 개인의 이메일 아카이브 내의 데이터를 보낸 사람(from), 받는 사람(to), 참조(cc)로 구분하여 주소록 내의 각 개인 간 관계 정보를 시각화한 'Email Map'[10]을 들 수 있다(그림 9). 이 작업도 앞선 사례와 마찬가지로, 데이터 자체만으로는 파악하기 불가능한 관계를 관계 생성 규칙에 따라 가시화하여 나타냄으로써 데이터 내부에 가려진 관계 정보를 직관적으로 보여준다.

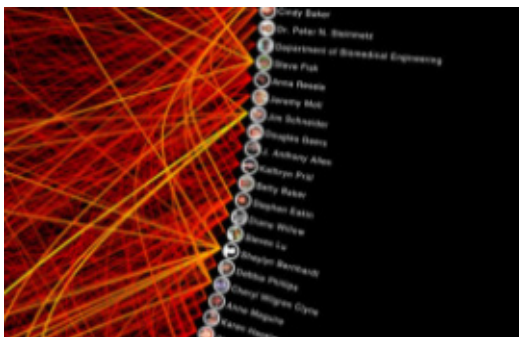


그림 9. Email Map(2007)/ Christopher Paul Baker

이러한 접근 방식은 논문의 인용(citation), 이메일, 온라인 채팅 데이터 등과 같이 송신자-수신자/ 인용자-피인용자/ 출발-도착 관계 등 데이터 단위 간 명확한

연결 관계가 존재하는 데이터 집합과, 신문이나 뉴스, 웹 사이트의 태그(혹은 검색 키워드)와 같이 그 콘텐츠가 시대상을 반영하고 시간의 흐름에 따라 누적되는 데이터 등을 주 재료로 하여 활용된다.

2. 스토리텔링의 창출

텍스트는 데이터의 특성상 텍스트 그 자체로서 어떠한 '의미(또는 내용)'를 갖는다. 그러나 이렇듯 의미를 지니는 텍스트 데이터의 시각적 가공을 통해 도출된 시각화 결과물은 다양한 해석 가능성을 지닐 수 있으며, 이러한 다양한 해석 가능성의 존재는 데이터에 대한 새로운 스토리텔링을 가능하게 한다.

NYTimes: 365/360[11]이 그 좋은 예이다(그림 10). Jer Thorp은 뉴욕 타임즈(New York Times)의 검색 API 데이터를 이용하여 각 연도별 기사에 상위로 언급된 단체들과 유명인사들 간의 유대관계를 선으로 연결하여 나타내었다. 이러한 시각화 결과물은 독특한 관점 또는 해석 방식이라는 도구를 통해 특정 단체와 개인에 대해 기사 자체에 서술된 내용 이외의 새로운 스토리텔링을 탄생시킬 수 있다.

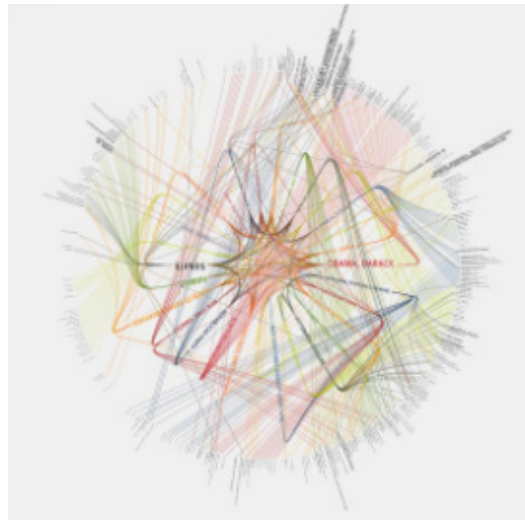


그림 10. NYTimes: 365/360(2009)/ Jer Thorp

또 다른 시각화 사례로 2009년 ACM Siggraph의 Information Aesthetics Showcase에 선정된 작품인

'Visual Genealogy: Mr. PARK, Myrang-Hwarok Clan'[12][13]을 들 수 있다[그림 11]. 이 작품은 한국의 밀양 박 씨 화록공 파의 족보를 시각화하여 나타낸 것으로, 추상적인 이미지에에도 불구하고, 한국의 역사와 제도, 시대상 등의 배경 지식을 도구삼아 다양한 스토리텔링을 가능하게 한다. 구체적 예로, 한 집안과 또 다른 집안의 결합을 이루는 결혼 제도에 관한 주제를 이끌어낼 수도 있고, 결혼을 통해 이루어지는 자손의 번창에 따라 중앙에서 방사형으로 뻗어나가는 전체적인 이미지 내에서 자손의 번식이 끊어지는 노드의 발생에 대하여 '전쟁의 발발'이라는 새로운 스토리텔링을 창출해 낼 수 있는 것이다.

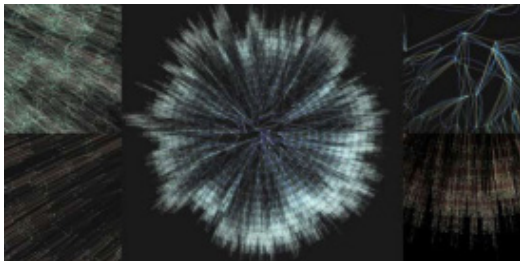


그림 11. Visual Genealogy: Mr. PARK, Myrang-Hwarok Clan(2009)/ Jin Wan Park

3. 예술적 표현의 재료

앞서 제시된 텍스트 시각화의 다양한 사례들이 정보의 전달이나 스토리텔링의 창출과 같은 일종의 디자인적 또는 도상해석학적(iconology) 접근 방식을 기반으로 하고 있다면, 이와 차별되게 완성된 예술 작품을 텍스트 시각화의 소재로 활용하여 또 다른 예술 장르로 치환하는 형태의 시각화 사례 또한 찾아볼 수 있다.

[그림 12]는 Stephanie Posavec의 'Writing without Words' 프로젝트[14]로서, 저명한 문학 저서들의 텍스트 데이터를 장(chapter)에서 절(paragraph)로, 각 절은 문장(sentence)으로, 문장은 단어들로 분리한 후 각 문장에 따른 단어의 수에 따라 조직화 하여 꽃잎의 형태로 표현한 것이다. 여기에 11가지의 주제별 색상을 할당하여 적용함으로써 각 시각화 결과 이미지가 임의의 소설에 대한 주제와 내용을 시각적으로 나타내고자 하

였다. 이에 따른 본 프로젝트의 최종 목표는 여러 문학 작품에 대한 주제와 성격 및, 작가에 따른 문장과 문체의 차이점 등을 시각적으로 느낄 수 있도록 하는 것이다. 그러나 이러한 원리로 탄생된 시각화 결과물은 각 문학 작품에 대해 전달하고자 하는 정보를 전혀 직관적으로 제공하고 있지 않으며, 그 보다는 하나의 시각 예술 작품에 가깝다. 즉, 하나의 완성된 예술품인 문학 서적을 데이터로 하여 또 다른 장르인 시각 예술의 작품이 재탄생 된 것이다.



그림 12. Writing without Words(2008) / Stephanie Posavec

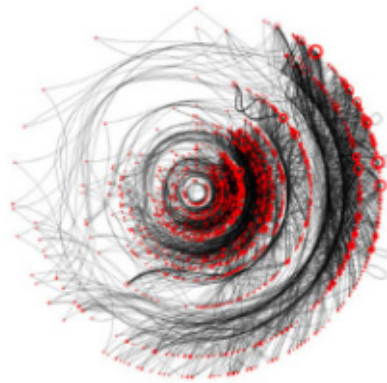


그림 13. Poetry on the Road(2006)/ Boris Muller

'Poetry on the Road'[15][그림 13] 또한 마찬가지로 경우에 속한다. 이것은 2006년 독일 브레만에서 개최된 국제 문학 페스티벌의 포스터로, 다양한 시(詩)를 해체하여 새로운 데이터 형태로 가공한 후, 이를 시각적으

로 재구성한 것이다. 이를 위해 작가는 각 시를 알파벳 단위로 분리하고, 각 알파벳에 고유의 숫자를 할당함으로써 시에 사용된 단어들을 단어가 갖는 알파벳 숫자의 합으로 치환하였다.

이 작품 또한 하나의 완성된 예술 장르인 ‘시’를 시각화의 재료로 삼아 이를 해체한 후 재가공함으로써 시각 예술이라는 또 다른 장르의 예술 작품으로 승화시켰다는 점에서, 데이터 시각화를 통한 예술 장르의 치환이라는 관점을 성립시킬 수 있다.

V. 결론

텍스트 데이터는 인류가 문자를 사용하기 시작했던 고대시대부터 시작하여 현재의 facebook, twitter 등으로 대표되는 SNS 시대에 이르기까지 인간과 인간의 생활을 반영하는 무한에 가까우리만큼 거대하고 방대한 데이터이다. 특히 인터넷을 통한 데이터의 공개에 따라 텍스트 데이터의 시각화는 사회 현상 분석이나 인문학적 접근을 넘어 공학자들의 새로운 시도와 예술가들의 창의 원천 등 매우 다양하게 활용될 수 있다.

이러한 시점에서 텍스트 데이터의 표현 재료와 접근 방식에 관한 체계적인 고찰은 필수적이다. 텍스트 데이터 시각화의 재료가 되는 텍스트 데이터의 본질 및 속성에 관한 이해는 메타 데이터의 개인화 과정을 거쳐 제작자에게 데이터를 바라보는 관점을 제시해줌으로써 시각화의 의도 및 주제 선택을 가능하게 한다. 또한 다양한 시각화 표현 및 접근 방식에 대한 심도 깊은 관찰을 통해 시각화의 목적에 최적화된 표현 방법론을 도출해낼 수 있는 것이다.

최근 소셜 미디어 이용 인구의 폭발적 증가로 사회적 거대 데이터에 대한 관심 또한 증폭됨에 따라 인간의 능력으로는 처리 불가능한 빅 데이터(big data)에 관한 다양한 연구 및 시도들이 이어지고 있다. 소셜 미디어 데이터를 활용한 다양한 인포그래픽(Infographic)의 활성화가 그 한 예이다. 데이터는 점차 복잡, 다양, 거대해지고 이를 해석 및 활용하려는 노력은 끊임없이 이어지고 있다. 그럼에도 불구하고 텍스트 데이터를 비롯한

데이터 시각화에 관한 이론적 고찰과 접근 방식에 대한 체계적 분석은 거의 이루어지지 않고 있는 실정이다. 데이터는 이해와 해석의 대상이며, 가공 및 접근에 따라 무한한 정보와 가능성을 갖는다. 따라서 데이터의 선별부터 시각화, 그리고 그 결과물을 해석하는 과정까지 인문, 사회, 공학, 예술의 협업을 필요로 한다. 데이터의 중요성이 점차 증가하고 있는 현 시점에서 데이터의 이해와 해석에서 출발하는 시각화라는 융합 학문적 연구 분야는 미래 인간 사회에서 데이터가 차지할 위상을 고려할 때, 보다 체계적인 연구와 이론적 축적을 필요로 한다.

참고 문헌

- [1] N. E. Miller, 'TOPIC ISLANDSTM - A Wavelet-Based Text Visualization System', Proceedings of the IEEE Visualization 98, 1998.
- [2] <http://www.chrisharrison.net/index.php/Visualizations/BibleViz>.
- [3] <http://tinyurl.com/2orwut>.
- [4] H. Kim and J. W. Park, "Textual Visualization based on Readability," Siggraph Asia 2011, ACM, 2011.
- [5] <http://www.itsbeenreal.co.uk/index.php?on-going/about/>.
- [6] <http://www.textarc.org/>.
- [7] T. Legan and L. Baker, "Visualizing the Text of Phillip Pullman's Trilogy 'His Dark Materials,'" Proceedings of NordiCHI 2010, 2010.
- [8] <http://www.eigenfactor.org/map/maps.htm>.
- [9] <http://christopherbaker.net/projects/mymap>.
- [10] Rosvall and Bergstrom, "Maps of Random Walks on Complex Networks Reveal Community Structure", Proceedings of the National Academy of Sciences(PNAS) USA 105, no. 4, 2008.

- [11] <http://blog.blprnt.com/blog/blprnt/nyt-this-was-1984>.
- [12] J. W. Park, "Visual Genealogy: Mr. PARK, Myrang-Hwarok Clan," Art Gallery in ACM Siggraph 2009, 2009.
- [13] J. W. Park, "Information Aesthetics with Visual Genealogy Project," Leonardo Transaction Research Announcemnets, 2009.
- [14] <http://www.itsbeenreal.co.uk>.

저 자 소 개

김 효 영(Hyoyoung Kim)

정회원



- 2006년 8월 : 성신여자대학교 미디어정보학부(공학사)
- 2010년 8월 : 중앙대학교 첨단영상대학원 영상학과(영상학 석사)
- 2010년 9월 ~ 현재: 중앙대학교

첨단영상대학원 영상학과 박사수료

<관심분야> : 데이터 시각화, 정보 시각화

박 진 완(Jin Wan Park)

정회원



- 1995년 2월 : 중앙대학교 컴퓨터 공학과(공학사)
- 1998년 : Pratt CGIM Computer Media(MFA)
- 2003년 3월 ~ 현재 : 중앙대학교 첨단영상대학원 교수

<관심분야> : Art&Technology, Procedural Animation