

빅데이터 시각화

- 최광선(솔트룩스 전략사업본부)
- 함영경(솔트룩스 전략사업본부)
- 김선호(솔트룩스 전략사업본부)

I. 서론

최근 빅데이터 기술시장의 성장과 함께 시각화 기술도 중요한 기술요소로 자리잡고 있다. 또한 전통적인 시각화 기술은 주로 시스템 로그나 실험 분석 결과 등에 대한 통계정보를 그래프로 보여주는 방식인데 반해, 빅데이터의 시각화 측면에서는 모든 데이터를 살펴보는 것에 제약이 따르기 때문에, 시각화의 기술적인 요소와 더불어 데이터를 요약하고, 한 눈에 살펴볼 수 있도록 돕는 시각화 방법론적 요소의 중요성이 커지고 있다.

시각화 관련 연구들에서는 이를 데이터 자체에 대한 시각화 외에 정보 시각화, 지식 시각화 등으로 구분하고 있다.[1-3] 정보 시각화가 데이터 마이닝 등의 분석 기술 등을 사용하고, 그 결과를 탐색하고 검색하는 관점에서의 기술이라면, 지식 시각화는 시각화의 목적, 문제해결, 인사이트 제공 등에 초점을 갖는 스토리텔링 관점에서의 시각화를 지향하는 것으로 설명될 수 있다.[4]

즉, 빅데이터의 시각화는 정형데이터 뿐 아니라 비정형 데이터를 포함한 데이터의 특성과 해당 도메인에 대한 충분한 이해를 바탕으로, 사용자가 그 의미를 직관적으로 파악할 수 있도록 표현되어야 하며, 이를 위해 사용자가 원하는 정보를 탐색할 수 있도록 사용자 인터페이스(UI) 또는 사용자 경험(UX)의 기능 지원도 중요한 고려사항이 되고 있다.

본 기고문에서는 이러한 관점에서 빅데이터 시각화의 중요성과 관련 기술 동향을 살펴보고, 향후 빅데이터 시각화의 발전방향을 조망해보고자 한다.

II. 빅데이터 분석과 시각화

1. 빅데이터와 빅데이터 분석

위키피디아에 따르면, 빅데이터란 기존 데이터베이스 관리 도구로 데이터를 수집, 저장, 관리, 분석할 수 있는 역량을 넘어서는 대량의 정형 또는 비정형 데이터 집합 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술을 의미한다. 즉, 빅데이터는 단순히 많은 자료만이 아닌, 다양한 형태의 자료를 빠르게 처리해 필요한 정보를 얻어내는 과정 전체를 뜻한다.

지난 10년 간 인터넷과 컴퓨팅의 발전, 모바일 기기와 센서들의 진화, 페이스북과 트위터와 같은 소셜 네트워크의 출현 등으로 기업 내 데이터량의 폭증했는데, 웹 사이트의 사용으로 발생하는 시스템 로그나 전자상거래 목록, UCC를 비롯한 동영상 콘텐츠, 휴대전화와 SNS에서 생성되는 문자 및 콘텐츠, CCTV 영상 데이터 등을 빅데이터의 예로 들 수 있다. EMC에 따르면 다양한 디지털 콘텐츠는 2020년에는 2.8제타바이트에 달할 것으로 예상했다. [그림1]

빅데이터의 특징은 일반적으로 3V - 데이터의 양(Volume), 데이터의 생성 속도(Velocity), 형태의 다양성(Variety) - 로 설명되며, 최근에는 가치(Value)나 복잡성(Complexity) 등을 추가하기도 한다.[5]

빅데이터 분석은 이와 같은 특징을 갖는 데이터로부터 비교적 적은 비용으로 가치를 추출하고, 데이터의 초고속 수집, 발굴, 분석을 통해 미래 경쟁력의 우위를 좌우하는 중요한 자원으로 활용 될 수 있는 가능성을 갖고 있다.

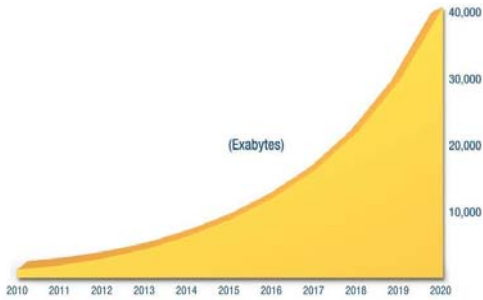


그림 1. 디지털 데이터량 증가, EMC

빅데이터 분석 방법론이나 기술에는 기존의 통계학과 전산학에서 사용되던 데이터 마이닝, 기계 학습, 자연 언어 처리, 패턴 인식 등이 활용될 수 있다. 특히 최근 소셜 미디어등 비정형 데이터의 증가로 인해 텍스트 마이닝, 오피니언 마이닝, 소셜네트워크 분석, 군집분석 등의 분석 기법도 주목 받고 있다.

빅데이터 분석은 정치, 경제, 문화, 과학기술 등 다양한 분야에서 활용되고 있으며, 근래에는 범죄예방, 재난예측, 기상예보 등 활용분야가 점점 더 확대되고 있다.

2. 시각화의 발전

빅데이터는 분석 자체도 중요하지만, 분석 결과와 가치를 서비스에 어떻게 적용할 것인지, 또 그 가치를 사용자에게 어떻게 보여줄 것인지가 더 중요하다. 단순한 숫자 정보나 텍스트로는 사람들에게 그 의미를 설득력있게 전달하기 어렵기 때문에, 분석된 결과를 적절한 차트나 이미지 등으로 시각화하는 것이 효과적이다.

현재까지는 분석결과를 관계형 데이터베이스에 저장하고, 웹사이트에서는 테이블 형태로 데이터를 조회하거나 검색하는 방식이 주된 방식이었다면, 최근에는 빅데이터를 직관적으로 살펴볼 수 있도록 돕는 시각화 기술이 중요해지고 있다. 가트너는 2015년까지 빅 데이터 관련 일자리가 440만 개로 늘어나게 될 것이며, 이들 중 대부분은 데이터 시각화와 같이 이전에는 요구되지 않았던 새로운 기술들을 필요로 할 것으로 예측했다.[6]

데이터 시각화는 정보를 시각적으로 표현하는 기술로서, 아직은 초기 단계라 정확한 정의, 역할 등을 명시하기는 어려우나, 보통 정보 시각화, 과학적 시각화, 인포그래픽 등으로 나눌 수 있다.

정보 시각화 (Information Visualization)는 보통 대규모 데이터를 색채, 통계(도표, 그래프 등), 이미지 등을 활용해 요약적으로 표현하는 것을 의미한다.



그림 2. 정보시각화 예

과학적 시각화(Scientific Visualization)는 실험결과나 시뮬레이션 데이터 등 복잡한 데이터를 쉽게 탐색할 수 있도록 3차원 그래픽 기술 등을 활용하여 시각화하는 기술이다.

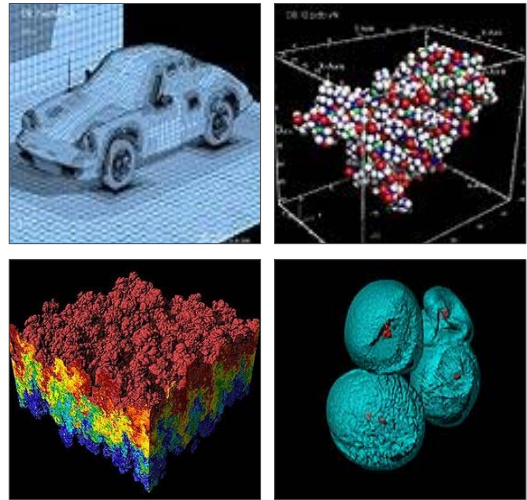


그림 3. 과학적 시각화 예

인포그래픽스는 인포메이션과 그래픽의 합성어로, 복잡한 수치나 글로 표현되어 있는 다량의 정보를 차트, 지도, 다이어그램, 로고, 일러스트레이션 등을 활용하여 한눈에 파악할 수 있도록 하는 디자인이다.

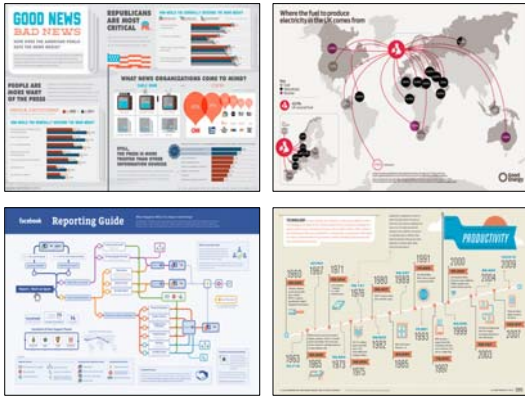


그림 4. 정보 그래픽 예

3. 시각적 분석

앞서 언급된 시각화 기법들은 정보 제공자가 시각화 대상 정보를 일반적으로 선택하고 사용하는 시각화 결과물을 수동적으로 제공받는 것과 달리, '시각적 분석은' '사용자가 능동적으로 시각화 도구를 활용하여 원하는 정보를 추출하고 분석하는 방법과 그 과정을 의미한다.

전통적인 시각화 도구는 SAS, SPSS와 같은 데이터마이닝 도구들이 있다. 최근에는 허둠(Hadoop)을 기반으로 하여 빅데이터 분석을 지원하는 데이터미어(Datameer), 타블로(Tableau), 스플렁크(Splunk)와 같은 시각적 분석을 돕는 도구들이 개발되고 있다.

이러한 시각적 분석도구들은 분석 데이터 선택, 필터링, 분석, 시각화, 리포팅, 공유와 같은 전반적인 분석 과정 전반을 지원한다. 즉, 대화용 GUI를 통해 사용자가 데이터를 정렬, 재배열, 필터링하거나, 기본/고급 통계, 상관성 분석, 시계열 분석 등 다양한 분석 기법들을 통해 사용자가 원하는 정보를 쉽게 분석할 수 있도록 지원한다. 또한 다양한 그래프나 보고서 템플릿 등을 통해 분석결과를 쉽게 시각화하고 배포할 수 있도록 지원하고 있다.

전통적인 데이터마이닝 도구는 데이터 분석가들을 대상으로 하는 설치형 어플리케이션이 대부분이었으나, 최근의 도구들은 일반 사용자들의 접근성을 높일 수 있도록 웹을 기반으로 하는 경우가 많고, 상대적으로 이해하기 쉬운 인터페이스를 사용하며, 간단한 질의문 작성을 지원하기도 한다.

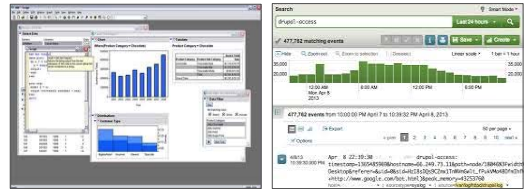


그림 5. (좌) SAS, (우) Splunk

III. 시각화 기술 동향

1. 시각화 기법

가장 일반적으로 사용되는 시각화 기법은 점이나 선을 활용한 차트를 활용하는 방법이다. 주로 통계적 데이터에 대한 시각화 기법이라고 할 수 있으며, MS Excel, Numbers 등의 스프레드시트 도구를 사용하거나, SPSS, SAS 등과 같은 전문적 통계분석 도구를 사용해서 쉽게 데이터를 선그래프, 막대그래프 등의 시각화 결과물을 얻을 수 있다. 한스 로슬링은 TED에서 버블 차트를 활용해 인구통계학 데이터를 간결하게 시각화함으로써 인상적인 강연을 한 바 있다.[7]

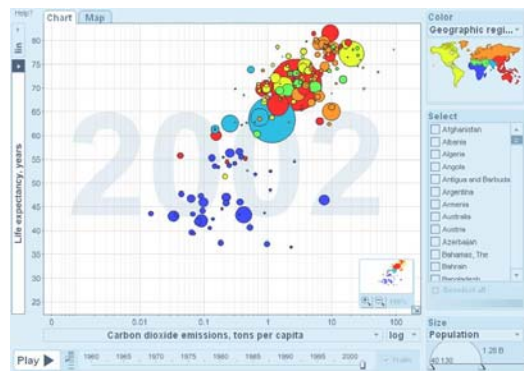


그림 6. 갱마인더 - 한스로슬링 강연 중 사용된 시각화 도구

시각화 대상데이터가 통계적 데이터 외에 위치나 크기 등의 공간적 속성 또는 데이터 분류 정보 등의 메타적 속성을 갖고 있는 경우 활용할 수 있는 기법으로는 열지도(Heat map), 하이퍼볼리 트리(Hyperbolic tree)와 같은 시각화를 예로 들 수 있으며, 차트 형태의 시각화와 더불어 일반적으로

많이 사용되는 기법이라 할 수 있다. 이 기법은 차트가 표현할 수 없는 메타정보를 직관적으로 살펴볼 수 있도록 도와주는 유용성을 제공한다.

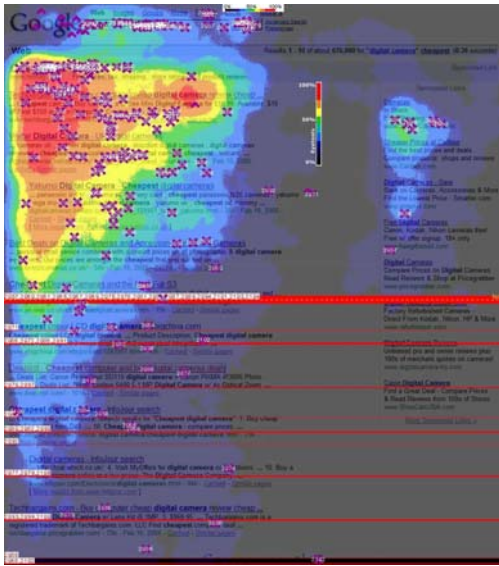


그림 7. 구글웹페이지를 살펴보는 사용자의 Eye tracking 정보를 열지도로 시각화한 예

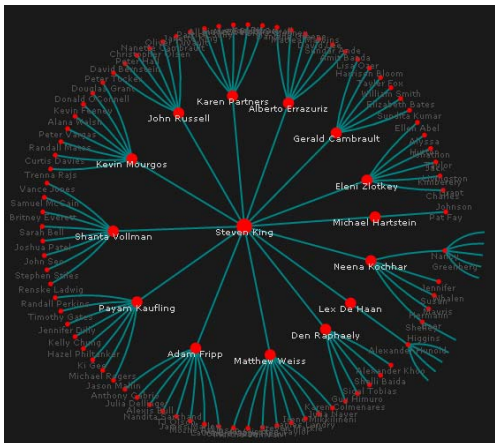


그림 8. 하이퍼볼릭 트리

소셜미디어가 각광을 받으면서 계층이나 네트워크 등의 관계정보를 표현하는 기법도 많이 활용되고 있다. 이러한 시각

화 기법은 소셜미디어 뿐 아니라 뇌 기능, 원소 정보 등 복잡한 연관관계를 표현하기 위해서 사용되기도 한다.

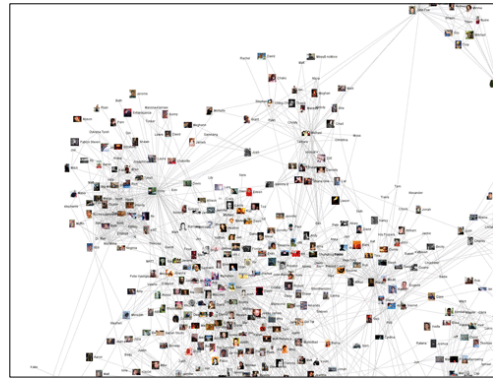


그림 9. 소셜미디어 네트워크 정보 시각화

소개된 기법들은 주로 시각화 도구나 알고리즘 등으로 구현할 수 있는 일반적인 데이터 시각화에 가까운 기법이라 할 수 있다.

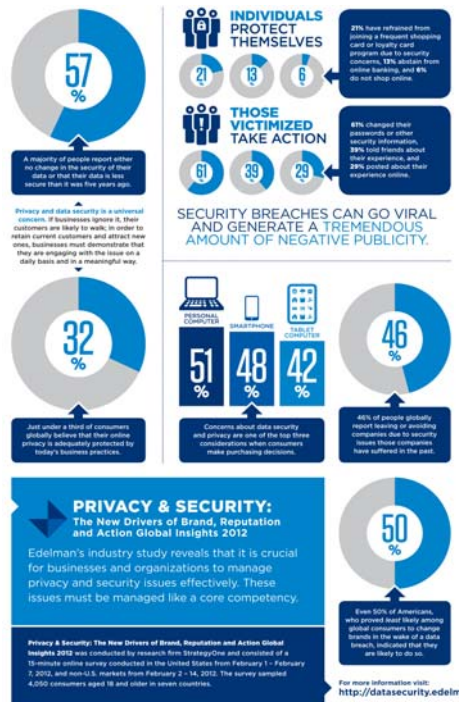


그림 10. 데이터 시각화와 인포그래픽스가 혼합된 사례

인포그래픽스(Infographics)의 경우는 좀 더 직관적인 정보 전달을 위한 디자인을 필요로 하지만 기본적으로 데이터 시각화 결과물을 재가공하여 두 시각화 기법이 혼합되어 사용되기도 한다.

결국, 시각화 기법은 시각화 대상의 데이터 중에서도 부각하고자 하는 정보의 종류, 특성, 즉 시각화 목적에 부합하는 적절한 기법을 선택하는 것이 중요하다고 할 수 있다. 또한 최근에는 모바일 디바이스나 터치 인터페이스 등이 널리 사용되고 있고, HTML5 등의 기술이 발전하면서 확대, 멀티터치 등의 사용자 인터랙션도 시각화 기법 선택의 고려사항이 되고 있다.

2. 시각화 플랫폼

‘시각화 플랫폼(Visualization Platform)’은 시각화를 위해 사용되는 솔루션이나 오픈소스 등을 통칭하는데 사용되고 있고 있을 뿐 명확한 정의가 내려져 있지 않다. 본 기고문에서 사용하는 시각화 플랫폼은 ‘시각화를 위해 활용되는 기반 기술’로써 시각화 도구나 솔루션 등을 구동하기 위한 소프트웨어 아키텍처, 프레임워크 등을 포괄하는 의미로 사용한다.

이러한 의미로 살펴본다면 초기의 시각화 플랫폼 마이크로소프트 윈도우즈(Windows) 운영체제에서 사용하는 GDI(Graphics Device Interface)를 예로 들 수 있다. GDI를 통해 윈도우즈 어플리케이션 개발자들은 다양한 시각화를 시도할 수 있게 되었고, 그 이후에는 DirectX, OpenGL 등 보다 자유로운 시각화 지원 기술 등과 결합하여 3D 시각화, 사용자 상호작용 지원 등도 가능하게 되었다.

일반 운영체제 외에 브라우저에서 동작하는 웹기반 시각화 기술로는 ActiveX나 자바에플릿 등을 예로 들 수 있다. 시각화만을 위해 개발된 기술은 아니지만, HTML 등의 웹기반 언어의 제약을 벗어나 다양한 시각화를 시도할 수 있는 기반을 제공했기 때문이다. 특히, ActiveX의 경우는 마이크로소프트의 IE(Internet Explorer)에서만 동작하는 한계가 있었으나 자바에플릿의 경우 대부분의 운영체제 및 브라우저에서 호환이 가능했기 때문에, 차트 등 통계데이터 시각화, 과학적 시각화 뿐 아니라 OpenGL등의 기술과 결합하여 3D 시각화 등 다양한 형태로 연구 개발 분야에서 널리 사용되었다.

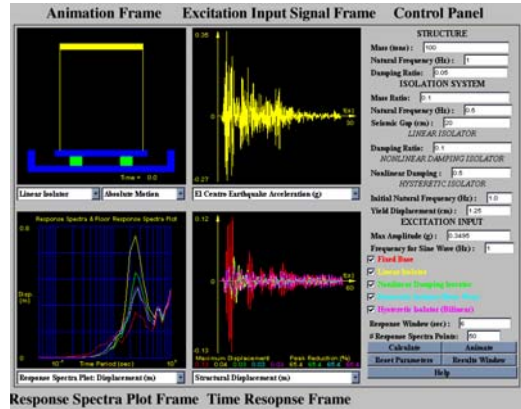


그림 11. 자바 애플릿 기반 시각화 사례

웹기반 시각화 플랫폼에서 빼놓을 수 없는 것은 어도비(Adobe)사의 플래시(Flash)이다. 플래시는 ActiveX나 자바에플릿 등과 달리 속도가 빠르면서도 경량의 시각화를 가능하게 했고, 벡터 기반으로 이미지를 처리하기 때문에 해상도의 제약 없이 자연스러운 이미지 표현이 가능했으며, 다양한 사용자 상호작용 지원이 가능했기 때문에 웹사이트 뿐 아니라 응용 어플리케이션 개발에도 널리 사용되었다.

최근에는 HTML5 기술이 새로운 웹 인터페이스 기술로 주목을 받으면서 웹기반 시각화 플랫폼 기술로 자리 잡아가고 있다. HTML5는 특정 플랫폼에 종속적이지 않으면서도 기존의 HTML의 한계를 극복하여 다양한 시각화 및 동작을 지원할 수 있는 등 기존 기술들의 장점을 대부분 수용한 형태로 발전하였고, 특히 모바일 디바이스 등에서도 지원이 가능한 표준 언어이기 때문에 시각화 플랫폼으로서의 큰 발전 가능성을 갖고 있다.

3. 주요 시각화 솔루션

시각화 솔루션은 주로 비즈니스 인텔리전스(Business Intelligence) 분야에서 활용된다. 대표적으로 IBM의 코그노스 인사이트(Cognos Insight), 마이크로소프트의 파워피벗(PowerPivot), 파워뷰(PowerView), SAS 엔터프라이즈 비즈니스 인텔리전스(SAS Enterprise Business Intelligence) 등을 들 수 있으며, 이들 솔루션은 데이터에 대한 다차원적 분석 결과에 대한 시각화를 제공하고, 그 결과를 보고서로 생성하는 기능 등을 제공한다.



그림 12. HTML5 기반 시각화의 예

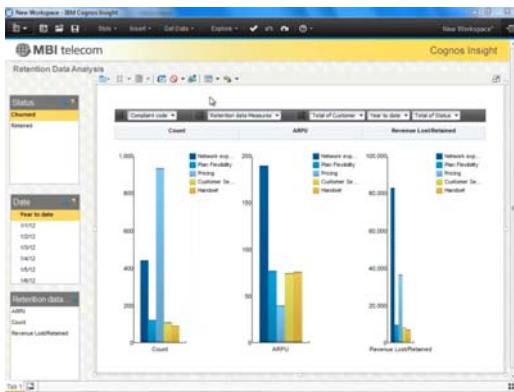


그림 13. IBM 코그노스 인사이트 [8]

또 타블로(Tableau), 팀코 스폿파이어 애널리틱스(Tibco Spotfire Analytics) 시각적 분석 솔루션으로써, 앞서 소개했던 솔루션들처럼 광범위한 기능을 제공하는 비즈니스 인텔리전스 전용 솔루션은 아니지만, 사용자가 원하는 정보를 다양한 측면에서 검색하고 분석해볼 수 있는 도구를 제공하고 있다.

기존의 비즈니스 인텔리전스 솔루션들은 주로 데이터 분석, 마이닝 등의 기법을 통해 일정한 방식의 결과 레포트를 생성하기 위해 시각화 기술을 활용했다면, 이들 시각적 분석 솔루션은 빅데이터의 시각화를 통해 사용자가 다양한 관점에서 인사이트를 얻을 수 있도록 지원함으로써, ‘지식 시각화’ 관점에서의 기능을 제공한다고 볼 수 있다. 특히 타블로의 경우 2011년에는 94% 수준의 매출 신장률을 기록함으로써 그 활

용성을 인정받았는데, 이러한 추세는 시각화가 단순히 정보를 보여주기 위해서만 활용되는 것이 아니라 정보를 탐색하고 지식을 발견하는 도구로 활용되고 있기 때문에, 사용자와의 상호작용이 점점 더 중요해지고 있다는 점을 시사하고 있다.

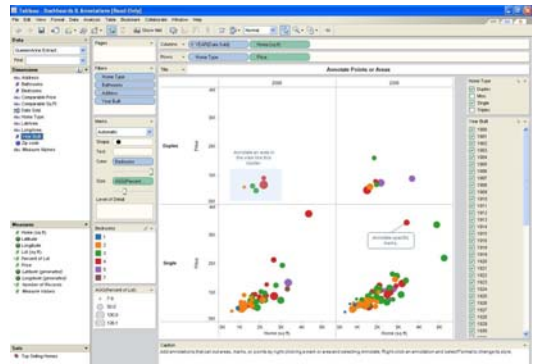


그림 14. 타블로(Tableau) 인터페이스 [9]

4. 오픈 소스 시각화 기술

시각화 기술이 점차 다양한 연구 및 서비스에서 활용됨에 따라 시각화 기술에 대한 요구사항이 점점 늘어나고 있다. 특히 최근에는 웹사이트에서도 데이터 시각화 뿐 아니라 인포그래픽스를 활용하는 경우도 증가하고 있으며, 이러한 사용자 요구사항과 함께 오픈소스 시각화 기술, 관련 시각화 도구 등의 활용이 빠르게 확산되고 있다. 이러한 시각화 기술은 소스를 모두 공개하는 프로젝트 또는 라이브러리 형태로 배포되기도 하고, 무료 시각화 생성 도구의 형태로 서비스되기도 한다.

일반적으로 이러한 기술들은 특정 환경에 대한 종속성 없이도 사용가능한 플렉스(Flex), 자바스크립트(Java Script), HTML5 등의 기술을 기반으로 개발 및 발전되어가는 추세이며, 특히 최근에는 모바일 기기들에 탑재된 웹브라우저의 성능이 크게 향상됨에 따라 자바스크립트, HTML5 기반의 시각화 기술이 각광을 받고 있다. 대표적인 시각화 라이브러리로 D3.js를 들 수 있다. D3.js는 다양한 차트 시각화를 지원하는 자바스크립트 기반 프로젝트인 Protovis를 개발하던 팀이 기존 프로젝트를 중단하고, HTML, SVG, CSS 등을 지원할 수 있도록 확장한 시각화 패키지이다. 매우 다양한 형태의 시각화를 지원하기 때문에 사용자가 커스터마이징하여 사용하기에 편리하다.



그림 15. D3.js가 지원하는 시각화 종류 [10]

유명한 자바스크립트 라이브러리인 jQuery 기반 또는 플러그인으로 개발된 시각화 프로젝트들로는 jQuery Visualize, Flot, jqPlot 등이 있다. 일반적으로 웹사이트에서 jQuery를 사용하는 경우가 많이 때문에, 이들 프로젝트들은 기존 웹사이트에서 사용하기에 개발편의성 측면에서 장점을 갖는다고 할 수 있다.

Visual.ly, VisualizeFree 등은 사전에 제작된 템플릿을 기반으로 인포그래픽스를 생성하는 도구를 제공한다. 예를 들어 로그인한 사용자 또는 페이지와 관계된 활동, 관계정보 등을 데이터시각화 또는 인포그래픽스 결과물로 변환해준다.

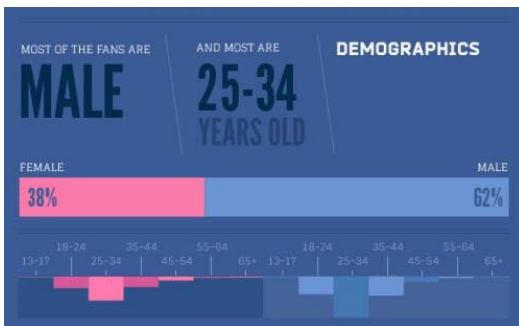


그림 16. Visual.ly에서 생성한 인포그래픽 예 [11]

이 밖에도 Google Fusion Tables나 Leaflet를 사용하면 지도를 기반으로 위치기반 데이터를 시각화할 수 있으며, Timeline, Dipity 등은 데이터를 시간축을 기준으로 시각화 해주는 등 다양한 시각화 관련 프로젝트들이 개발되고 있다.

IV. 시각화 절차와 기법

1. 시각화 절차와 기법 연구 동향

기존의 시각화 시스템 구축 절차와 관련한 연구들을 먼저 살펴보면, 연구자에 따라 다소 차이가 있기는 하지만 공통된 요소들을 제시하고 있다.



그림 17. Ben Fry의 연구

Ben Fry의 연구에서는 데이터의 획득부터 해석, 정제, 의미, 표현, 정제를 거쳐 상호작용에까지 이르는 데이터 수집과 분석을 포함한 전체 과정에서의 시각화 절차를 제안하였다.[12]

표 2. ETRI의 시각화 절차의 단계 분류

| 단계 | 설명 |
|-----------|---|
| 정보 조직화 단계 | - 사용자의 정보 인지에 관여 - 혼돈의 상태로 존재하는 데이터를 분류하고 배열하고 조직화하여 질서를 부여 |
| 정보 시각화 단계 | - 사용자의 정보 지각에 관여 - 보다 효율적으로 정보 전달을 위해 시각, 청각, 촉각, 미각, 후각의 감각 기관에 최적의 자극을 제시하는 방법 제시 |
| 상호작용 단계 | - 정보와 사용자 간의 상호작용 측면의 사용자 경험을 디자인 - 정보의 인지적 요인뿐만 아니라 지각적 요인을 함께 활용 - 정보 시각화 단계와 밀접하게 연동되면서 동시에 입력 기술의 특성도 함께 고려 |

ETRI(신희숙 외)의 연구에서는 시각화 과정의 각 절차들을 3단계로 나누어 각 단계별 절차와 기법을 제시하였다.[13]

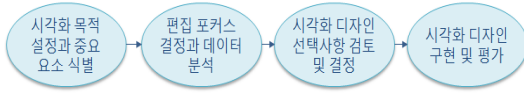


그림 18. Andy Kirk의 연구

Andy Kirk의 연구에서는 조금 더 그래픽 디자인 적인 요소를 중심으로 절차를 정의하였다.[14] 시각화를 위해 수행해야 할 절차로써 1)시각화 목적 설정과 중요 요소 식별, 2) 편집 포커스 결정과 데이터 분석, 3) 시각화 디자인 선택사항 검토 및 결정, 4)시각화 디자인 구현 및 평가 등의 절차를 제시하고 있다. 또한, 시각화 디자인을 위한 역할로서, 발기자, 데이터 과학자, 저널리스트, 컴퓨터 과학자, 디자이너, 인지 과학자, 커뮤니케이터, 프로젝트 관리자 등 8가지의 역할을 제시하고 있다.

오병근의 연구에서는 단순한 정보 시각화가 아닌 지식 시각화 관점에서 지식에 대한 사용자의 이해과정과 그에 따른 지식 시각화의 대상구분, 지식 아키텍처를 위한 요소와 정립 절차 및 시각적 매핑의 요소와 요인, 전달효과와 방법에 대한 개념과 작용하는 요인을 제시하였다.[3]

또한, 시각화에서 가장 중요한 과정인 시각화 디자인은 매우 심미적이며, 사용자 인터페이스와 직접적인 관계를 갖는다. 전통적인 사용자 인터페이스 관점과 관련하여 Norman의 연구 [15]에서는 사용자 인터페이스 설계의 중요한 고려사항으로 행동유도성(affordance), 대응(mapping), 심성모형(mental model), 실행주기(execution-evaluation cycle), 인터페이스 메타포(metaphors)등을 강조하고 있다. 이와 관련하여 Jacob Neilson의 연구에서는 휴리스틱에 의한 10가지의 사용자 인터페이스의 이용성 평가 기준을 제시하기도 하였다.

2. 빅데이터 시각화 절차

다양한 선행 연구들에 의해 시각화 관련 기술이 발전해왔으나, 시각화 시스템 구축 과정은 공학적 노력 뿐 아니라 심미적 노력도 함께 수행되어야 한다. 이러한 측면에서, 본 기고문에서는 앞선 선행연구의 공통적인 요소들을 중심으로 다음과 같은 빅데이터 시각화 시스템 절차를 제안한다.

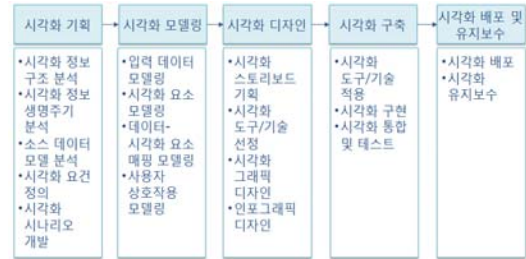


그림 19. 빅데이터 시각화 시스템 구축 절차

1) 시각화 기획

시각화 기획에 있어서 시각화의 목적을 명확히 하는 것이 무엇보다 중요하다. 또한, 시각화 대상 정보의 구조와 생명주기 및 소스 데이터 모델 분석을 통해 시각화하고자 하는 데이터의 물리적, 의미적 그리고 시간적 특성을 파악해야 한다. 시각화 대상 정보의 구조 분석은 시각화를 위한 정보 혹은 지식체계의 구조를 파악하는 일이다. 또한, 시각화 대상 정보가 제공되는 시간적, 절차적 구조인 생명주기를 파악함으로써, 시각화 정보와 시각화 도구간의 상호작용의 기준을 파악해야 한다. 마지막으로 시각화 정보를 구성하기 위한 소스 데이터의 모델을 분석하여 구체적으로 어떤 원천 정보들이 시각화에 활용될 수 있는지를 파악해야 한다.

이와 관련하여 오병근의 연구에서는 시각화의 대상을 가시적인 대상과 비가시적인 대상으로 구분하여 설명하였다. 여기서 가시적인 대상은 눈에 보이는 현상과 기기를 통해 보이는 현상 등을 포함하고 있다. 또한, 비가시적인 대상은 보이지 않는 현상에 대한 측정 데이터, 보이지 않는 현상의 추상적 데이터, 가정에 의한 현상, 개념의 실체화를 포함하고 있다.

이처럼 시각화 대상이 항상 물리적이고 가시적이지만은 않기 때문에, 시각화 결과에 대한 합의를 이끌어 내는 일은 쉬운 일이 아니다. 그러므로 시각화의 목적과 대상의 특성이 파악이 되면 시각화 시스템의 요건(중요 요구사항)을 정의하고 이에 적합한 대표 시나리오(사용자 시나리오에 따른 프로토타이핑) 개발을 통해 이해당사자들과 시각화에 대한 목적과 목표 및 결과물에 대한 공감대를 형성해야 한다.

2) 시각화 모델링

시각화 하고자 하는 대상 정보 혹은 지식의 구조가 파악되

면, 실질적으로 시스템을 통해 입력되는 입력 데이터의 구조를 정의해야 한다. 최근의 시각화 도구들인 예전과 달리 웹 환경에서 비동기적 상호작용(예: AJAX와 HTML5)을 통해 작동하는 방식이 주류를 이루고 있다. 그러므로 시스템적인 입력 데이터 모델은 이러한 작동방식들을 고려해서 결정되어야 한다.

또한, 시각화하는 방식에 따라서 시각화를 위한 정보요소들의 설계도 매우 중요하다. 앞서 합의된 시각화 시나리오들을 만족하는 구체적인 시각화 요소들을 도출하고 정의해야 한다.

하지만 시각화를 위한 입력 데이터의 구조가 시각화 요소들과 항상 일치하는 것은 아니다. 실세계의 다차원적 정보들은 시각화 방식에 따라 1차원적, 2차원적, 3차원적 혹은 그래프나 네트워크 형태의 시각화 도구들에 사상되어야 한다. 이를 해결하기 위해서는 입력 데이터 모델과 시각화 요소간의 매핑 및 변환 모델을 설계해야 한다.

마지막으로 대상 정보들에 대한 단순한 조치가 아닌 탐색적 시각화 도구들을 위해서는 사용자와 시각화 도구간의 상호작용의 구조와 방식을 식별하고 설계해야 한다.

3) 시각화 디자인

사용자에게 제공되는 시각화의 디자인은 사용자 인터페이스 디자인의 일종이다. 물론 잘 패키징된 솔루션을 활용하여 시각화를 구현하는 경우에는 솔루션에서 제공하는 시각화 도구를 배치하는 정도로 디자인이 진행되지만, 시각화 자체가 단순한 분석결과를 제공하는 도구가 아닌 사용자와 상호작용을 하는 서비스의 요소로 디자인이 되는 경우에는 사용자 인터페이스 관점에서 디자인이 진행되어야 한다. 최근에는 사용자 인터페이스보다 더 발전된 사용자 경험 관점에서 접근되기도 한다.

의사결정을 위해 분석결과와 시각화를 제공하고자 하는 경우에는 분석의 논리 구조에 따른 시각화의 전개가 필요하다. 한편 데이터 탐색이나 검색, 혹은 이와 연계된 실시간적인 시각적 분석 기능을 제공하기 위해서는 사용자와 상호작용 흐름을 사용성 관점에서 구성하는 것이 필요하다. 이러한 과정을 디자인하기 위해서는 사용자 인터페이스 디자인에서 흔히 사용되는 스토리보드의 기획이 필요하다.

특히 이러한 사용자 인터페이스 디자인에서 Nielsen의 연

구에서 제시한 이용성 평가 방법인 휴리스틱을 고려하는 것은 매우 유용하다. Nielsen은 다음과 같은 10개 항목의 휴리스틱을 제시하고 있다.[16]

표 3. Nielsen의 휴리스틱 (이지연 재인용)

| 항목 | 내용 |
|----|---|
| 1 | 지주 이용되는 필요한 정보만으로 구성된 단순한 인터페이스인가? |
| 2 | 사용자가 이해하기 쉬운 단어와 문장으로 표현되었는가? |
| 3 | 필요한 모든 기능이 쉽게 눈에 띄도록 구성되었는가? |
| 4 | 같은 단어, 상황 등이 일관성있게 사용되었는가? |
| 5 | 시스템이 사용자 요구를 처리 중이라는 표시를 해 주고 있는가? |
| 6 | 사용자가 길을 잃었을 때 원점으로 바로 전환할 수 있도록 해주고 있는가? |
| 7 | 초보자, 전문가 등 사용자 수준에 맞추어 바로가기 등의 기능을 지원하고 있는가? |
| 8 | 사용자가 이해하기 쉬운 표현으로 오류에 대한 정보를 제공하고 처리방안을 제공하고 있는가? |
| 9 | 오류를 처음부터 방지할 수 있는 체계를 갖추었나? |
| 10 | 사용자의 작업을 지원하도록 적절한 표현과 양의 문서를 제공하는가? |

시각화 스토리보드를 통해 시각화의 기능적이고 구조적인 구성이 디자인되면, 각각의 구성요소들을 위해 분석 솔루션이나 시각화 전문 솔루션에서 제공하는 구성요소들을 선정하거나 목적에 맞는 시각화를 구현하기 위한 기반 기술을 선정한다.

시각화 도구나 기술이 선정되면, 각각의 도구와 기술들의 그래픽 요소들을 디자인 한다. 이러한 디자인 요소들은 단순한 스타일 요소인 색상, 폰트, 레이아웃 등을 선택하는 일로부터, 의도한 시각화 표현을 위해 모든 것을 새롭게 디자인하는 과정일 수도 있다.

마지막으로 기하학적인 시각화 표현이 아닌, 사용자 친화적인 시각화 표현을 위해서는 인포그래픽 기법이 사용될 수도 있다. 물론 앞서 말한 모든 시각화 디자인들이 인포그래픽의 한 종류이지만, 여기서는 특별히 사용자에게 친밀한 상징을 이용한 일러스트레이션에 가까운 표현을 의미한다.

4) 시각화 구축

시각화 구축은 완성된 시각화 디자인에 따라 시각화 도구와 기술을 적용하여 시각화를 제공하는 기능을 구현하고 시스템에 통합하는 작업이다.

시각화 구현은 전통적인 어플리케이션 UI 개발과 동일한 작업이지만, 독자적인 시각화 표현을 적용하는 경우에는 원천적인 컴퓨터 그래픽 기술을 적용하여 새로운 형태의 시각화 알고리즘을 개발하는 것을 포함하기도 한다. 최근에는 D3.js와 같은 HTML5 기반의 오픈 소스 시각화 기술을 응용하는 사례가 많다.

시각화 구현이 끝나면 구현된 시각화 결과물을 서비스하고자 하는 시스템에 통합하고 기능, 사용성 및 성능 등을 시험한다.

이 때, 빅데이터 시각화 관점에서 꼭 점검해야 하는 사항으로는 데이터의 규모와 범위가 매우 방대하거나 실시간적인 반응을 요구할 수 있으므로, 개발된 시각화 사용성과 성능이 이러한 방대한 데이터를 수용하는지, 혹은 실시간적인 작동에 적합한지를 필수적으로 점검해야 한다.

5) 시각화 배포 및 유지보수

시각화 구축도 시스템 구축의 한 과정이므로 시각화 구축의 결과물도 당연히 일반적인 시스템과 동일하게 배포되고 유지보수 되어야 한다.

V. 결론

서론에서 언급한 바와 같이, 빅데이터의 시각화는 시각화를 통해서 전달할 지식을 중심으로 한 스토리텔링 요소가 중요하다. 이를 위해서는 시각화하여 표현하는 과정 뿐 아니라 시각화 대상 정보를 선별하고 조직화하는 과정에서도 데이터 분석 도구 등을 통해 사전에 데이터를 충분히 고찰하는 것이 필요하다고 할 수 있다. 이미 시각적 분석 도구 등 이러한 과정을 돕는 솔루션이나 프로젝트들이 개발되어 있으며, 이 과정에서 사용자와의 상호작용을 통해 다양한 관점에서 정보를 획득할 수 있도록 지원하고 있다.

특히 시각화 결과물이 데이터 분석 전문가에게만 국한되지 않고 일반 사용자에게 지식을 전달하기 위한 도구로 확대됨

에 따라, HTML5 기술 등을 활용하여 소프트웨어적, 하드웨어적 종속성 없이 시각화 결과물을 조회할 수 있도록 제작하는 기술도 중요해졌다. 관련한 오픈소스 프로젝트들을 포함한 기술적 요소가 중요함은 물론이다.

빅데이터 시각화는 더 이상 특정 도구에만 의존하거나 디자이너에게 디자인 업무를 맡기는 방식이 아닌 기획, 분석, 구현 각 분야의 담당자들이 함께 협력하여 완성해야 하는 종합적 기술에 가까워지고 있다.

또한 최근 인포그래픽스가 이슈화 되면서 국내에서도 관련 연구와 시도들이 생겨나고 있으나, 아직까지는 빅데이터 시각화 방법론에 대한 인식 부족이나 기술 인력 부족 등으로 아직까지는 의미있는 결과물이 나오지 못하고 있다. 그러나 최근 시각화 기술이 HTML5를 중심으로 발전함에 따라 휴대용 스마트 기기의 높은 보급률, 네트워크 안정성 등의 강점을 기반으로 다양한 시각화 기법을 활용할 수 있는 잠재 가능성과 기회를 갖고 있기 때문에, 충분한 투자를 통해 관련 기술을 발전시키고 보급하는 노력이 필요하다.

참고문헌

- [1] M. Zeiller, "A Case Study Based Approach to Knowledge Visualization", Proceedings of the Ninth International Conference on Information Visualization, 2005
- [2] 장석현, 이주엽, 이경원, "정보시각화와 지식시각화의 비교 분석을 통한 표현방법 연구", 한국HCI학회 학술대회, 1242-1248쪽, 2008년 2월.
- [3] 오병근, "지식시각화 모형체시를 통한 지식의 디자인체계 연구", 디자인융복합연구, 38호, 219-233쪽, 2013년 2월.
- [4] 이지수, 정겨운, 이경원, "스토리텔링의 시각화를 위한 개념적 맵들의 특성분석", 한국HCI학회 학술대회, 1214-1219쪽, 2008년 2월.
- [5] 정용찬, "빅데이터 이해총서", 커뮤니케이션북스, 2013년.
- [6] Gartner, "2012년 세계 IT 성장률 보고서", 2012년.
- [7] TED, Hans Rosling, <http://www.ted.com/>
- [8] Cognos Insight, <http://www-03.ibm.com/software/products/us/en/cognos/insi/>
- [9] Tableau, <http://www.tableausoftware.com>

- [10] D3.js, <http://d3js.org>
- [11] Visual.ly, <http://visual.ly>
- [12] Ben Fry, "Visualizing Data", O'Reilly, pp. 7-15, 2008.
- [13] 신희숙, 임정목, 박준석, "정보 시각화 기술과 시각장애인을 위한 정보 표현 기술", 전자통신동향분석, 제 28권, 제 1호, 2013년 2월.
- [14] Andy Kirk, "Data Visualization: a successful design process", PACKT,
- [15] Norman, D.A. "The Design of Everyday Things", New York: Basic Books, 1988.
- [16] Nielson, J. "Usability Engineering", Morgan San Francisco:Kaufmann, 1988.

저 자 소 개



최 광 선

1996: 인하대학교
자원공학과 공학사

2005: 숭실대학교
소프트웨어공학과 공학석사

2008: 숭실대학교
컴퓨터학과 공학박사 수료

현 재: 솔트룩스
전략사업본부 본부장

관심분야: 인공지능, 빅데이터,
지식베이스, 시맨틱 웹



함 영 경

2001: 상지대학교
전자계산학과 이학사.

2004: 숭실대학교
컴퓨터학과 공학석사.

현 재: 솔트룩스
전략사업본부
분석기술팀 팀장

관심분야: 인공지능, 빅데이터,
추천시스템,
시맨틱 웹



김 선 호

2005: 경희대학교
한방시스템공학과 공학사

2007: 경희대학교
동서의료공학과 공학석사

2009: 경희대학교
동서의료공학과
공학박사 수료

현 재: 솔트룩스 전략사업본부
분석기술팀 대리

관심분야: 빅데이터, 시맨틱웹,
링크드데이터,
추론시스템