

실제사례 통해 데이터를 보석으로 활용하기 위한 세 가지 요소 ❶ 시간

“데이터의 시간에 따른 변화를 이해하라”

은유적 도식화가 해법...다각적이고 효과적인 변화분석 가능

데이터는 모든 업무의 출발점이며, 과정이며, 종착점이다. 우리 주변에는 방대한 양의 데이터가 있으나 이러한 데이터의 본질을 제대로 이해하고 업무에 효과적으로 사용하는가에 대해서는 의문이 드는 경우가 많이 있다. 살아있는 데이터를 보석으로 만들어서 업무에 활용하기 위해서는 데이터를 그림으로 표현하는 게 효과적이다. 왜냐하면 그림을 보면 데이터가 말하려고 하는 내용이 들리기 때문이다. 그래서 전문가일수록, 어렵고 복잡한 개념일수록 그림으로 그리면서 설명하게 된다. 적절하게 데이터를 활용하기 위해서는 그림을 그리고, 이해하기 위한 세 가지 요소, 즉 시간, 조감, 상관에 대하여 그 특징을 이해해야 한다. 이번 호부터 총 3회에 걸쳐 이러한 특징들에 대해 알아보며, 각 요소는 실제로 업무에 적용됐던 사례를 중심으로 상세히 설명한다. 먼저 시간 요소에 대해 알아본다.



윤 태 성
KAIST 교수 · 윤츠 사장
yoon.taesung@kaist.ac.kr

필자는 KAIST 경영과학과와 기술경영전문대학원 겸직 교수이며 윤츠와 오픈날리지(동경) 사장이다. 저서에는 '오픈날리지 · 지식은 어떻게 비즈니스가 되는가(21세기복스)', '전문가 일수록 그림으로 데이터를 그림으로 설명하는 이유(매경출판)'(출판은 10월 10일경), '테크놀로지 로드맵 · 기술지식의 조감과 분석에 의한 신산업 창조(일본어, 공저)' 등 다수가 있다. 저자에의 연락은 untz.book@gmail.com이며, 이 원고의 참고 사이트는 <http://untz.co.kr>.

글 · 연 · 재 · 순 · 서

❶ 시간 → 데이터의 시간에 따른 변화를 이해하라

❷ 조감 → 데이터의 전체상을 조감하라(11월호)

❸ 상관 → 데이터간의 상관관계를 파악하라(12월호)

1 데이터 정의 · 특성

본격적인 논의에 앞서서 우선 데이터에 대해서 간단히 정의할 필요가 있다. 왜냐하면 우리가 데이터를 다루거나 데이터에 관한 논의를 할 때에 미리 정의를 해 놓으면 논의의 범위가 명확하게 지정되기 때문이다.

여기서는, 우리 눈앞에 있는 팩트(fact)를 데이터라고 정의한다. 데이터는 여러 가지 형식으로 표현되는데 예를 들어 숫자, 문자, 날짜, 기호 등으로 표현된다. 데이터는 한 가지 형식만으로 표현되기도 하지만 여러 가지 형식이 혼재되어 표현되기도 한다.

데이터 활용 따라 정보 · 지식으로 변환

만약 우리가 데이터를 사용하여 의사결정을 하게 된다면 이때는 '정보'라고 정의한다. 데이터를 사용하여 추론을 하게 된다면 이때는 '지식'이라고 부른다. 즉 데이터와 정보, 지식은 근본적으로 같은 형태를 가지고 있지만 각 개인이 이것을 어떻게 활용하는가에 따라서 데이터에서 정보나 지식으로 변환하게 되는 것이다. 그러나 이러한 변환은 각 개인에 의존하기 때문에 아무 것도 하지 않는 사람에게는 데이터이지만, 의사결정에 사용하는 사람에게는 정보가 되고, 추론에 사용하는 사람에게는 지식이 된다.

이러한 정의에 따르면, 예를 들어 우리 눈앞에 10년간의 신문이 놓여 있다면 이것은 데이터라고 한다. 전문서적이 100권 있어도 데이터이며, 웹 사이트의 게시판에 투고된 내용이 1,000건 있어도 데이터이다. 왜냐하면 이 상태에서는 아무런 의사결정도 없고 아무런 추론도 행하여 지지 않기 때문이다.

하지만 이 신문 중 특정한 기사를 읽고 의사결정을 내렸다면 이 기사는 정보가 된다. 마찬가지로 전문 서적 중 특정한 내용이나 특정 게시판의 투고 내용도 정보가 된다. 더 나아가 만약 특정한 신문 기사를 읽고 난 후 어떤 추론을 하게 되었다면 이 기사는 지식이 된다. 전문서적이거나 투고 내용도 마찬가지다.

시간에 따라 변화하는 생명력 가진 데이터

지금 살아가는 모든 사람들은 방대한 양의 데이터를 보존하고 있다. 만약 개인적으로 수십 기가바이트의 데이터를 보존하고 있대도 누구도 놀라거나 의아해 하지 않는다. 오히려 데이터를 거의 가지고 있지 않는 사람을 보면 위화감을 느낀다. 데이터를 보존하기 위한 비용은 점차 줄어들지만 방대한 양의 데이터를 단순히 보존만 하고 있기에는 너무 아까운 내용이 많다.

그래서 누구라도 데이터를 효과적으로 활용해 의사결정이나 추론에 활용하길 원한다. 다시 말하면 누구나 데이터를 정보나 지식으로서 변환하고 싶어 한다. '데이터를 보석으로 활용한다'는 표현은 정확하게 '데이터를 정보나 지식으로서 활용하는 것'을 지향한다.

데이터의 본질을 이해하고 업무에 효과적으로 사용하려면 "데이터는 살아있는 생물과 같다"는 점을 인식해야 한다. 자연계에 있는 생물을 생각해 보면 시간이 흘러도 전혀 변하지 않는 것은 없다. 강산조차도 10년이면 변한다고 하지 않는가. 살아있기 때문에 시간이 경과하면 어떠한 형태로든지 변화하게 된다.

일상생활에서 접할 수 있는 데이터는 거의 대부분이 시간의 경과에 따라 변화한다. 예를 들어, 주식시장의 주가, 웹 사이트의 접속 건수, 생산 현장의 불량품 발생률, 차량 통행량, 매출액, 예금 입출금 액수, 예금 금리, 특허 출원 건수, 사원의 출퇴근 시각, 주유소별의 기름 값, 개인의 체중 등등 일일이 나열할 수가 없을 정도다.

2 시간에 따른 데이터 변화 이해 해법

데이터가 시간의 경과에 따라서 변화하는 특징을 이해하려면 데이터가 변화하는 모양을 사람이 잘 이해할 수 있는 형태로 그림을 그리는 게 좋다. 왜냐하면 그림을 봄으로써 우리는 데이터가 가지는 의미를 시각적으로 분석해 결과적으로 데이터의 본질을 이해할 수가

있기 때문이다.

그러나 전통적인 방법, 즉 꺾은선그래프나 막대그래프 정도의 표현력만으로는 불충분하다. 이러한 방법은 데이터의 어느 한 가지 속성을 기준으로 그림을 그리기 때문에 나머지 다른 속성을 알기 위해선 또 다른 그림을 그려야 한다. 만약 데이터의 속성이 열 가지라면 열 개의 그래프를 그려서 모은 후 분석해야만 데이터가 어떻게 변화하는지를 이해할 수 있다. 그렇다면 어떤 방법을 사용해 한 가지 그림에 몇 개의 속성을 동시에 나타내면서 시간의 경과에 따른 데이터의 변화를 이해할 수가 있을까? 대답은 자연계에 있는 식물에 있다.

은유적 식물변화 표현 가능

데이터의 변화를 식물의 변화에 비유하여 나타내는 방법, 특히 꽃은 일생의 주기가 짧아서 인간에게도 이해하기가 쉽기 때문이다. 일반적으로 꽃이 성장하는 모양을 관찰해 보면 먼저 뿌리가 생기고 나서 줄기가 성장하고, 거기에 맞추어서 잎도 성장하고 꽃도 피어난다. 실제로 꽃이 성장해 가는 모양을 보면서 시간이 흘러가고 계절이 바뀌는 것을 실감하는 사람도 많이 있을 것이다.

만약 꽃의 각 부분을 데이터의 각 속성과 대응시키면 데이터의 변화를 꽃의 성장에 비유하여 나타낼 수 있게 된다. 꽃의 성장과정은 누구나 알고 있기 때문에, 꽃의 성장을 통하여 데이터의 변화를 이해하기가 매우 쉬워진다.

데이터를 꽃에 비유해 표현하는 방법은 데이터의 은유적인 표현 방법이라고 할 수 있다. 데이터를 막대그래프의 높이에 직접 대응시켜서 그래프를 그리거나, 원그래프의 각도에 대응시켜서 그래프를 그렸던 방법은 직접적인 표현방법이다. 지금까지는 데이터를 가시화하려면 직접적인 표현방법을 주로 사용하였는데, 이는 직접적인 표현방법이 완벽하다거나 은유적인 표현방법이 적절하지 않아서가 아니다. 데이터를 그림으로 표현하는 방법이 그리 많지 않기 때문에 데이터를 은유적으로 표현하기가 어려웠다.

하지만 누구나가 보유하고 있는 데이터의 양이 비약적으로 증가하고, 데이터로 나타내고 싶은 내용이 점점 복잡해질수록, 데이터를 은유적으로 표현하는 경우가 늘어날 것이다. 사람들은 일상생활에서 은유적인 표현을 많이 사용하고 있기 때문에 데이터를 은유적으로 표현하는 것에 어떤 위화감도 느끼지 않으며, 특히 복잡한 데이터일수록 은유적인 표현을 선호한다.

3 은유적 데이터 표현 실례

은유적으로 데이터를 표현하는 실례로 약 3,000 건의 탄소 나노 튜브 관련 특허를 분석한 데이터를 통해 살펴본다. 특허를 분석하는 방법에는 기계적으로 대량의 특허를 분석하는 경우와 전문가를 동원하여 깊이 있게 분석하는 경우가 있다. 그러나 탄소 나노 튜브가 발명된 초기라면 기계적 분석은 기술 해석상의 한계가 있으므로 각 분야의 전문가를 동원해 한 건 한 건 내용을 심도 깊게 분석했다. 그러나 아무리 전문가라 하여도 최첨단의 기술 분야에 관한 특허 3,000 건을 모두 다 이해하기는 어렵기 때문에 수십 명의 전문가를 동원하여 특허를 분석했다.

이러한 분석 작업을 거쳐서 생성된 특허 분석 데이터를 어떻게 표현해야 경영자, 관료, 정치가 등이 이해할 수 있도록 설명할 수 있을까? 그들은 특허 내용의 기술적인 측면은 모르지만 투자나 정책입안과 같은 경영적인 아니면 정책적인 의사결정을 하여야만 한다. 이때 필자가 사용한 것이 <그림 1>에 있는 꽃 모델이다.

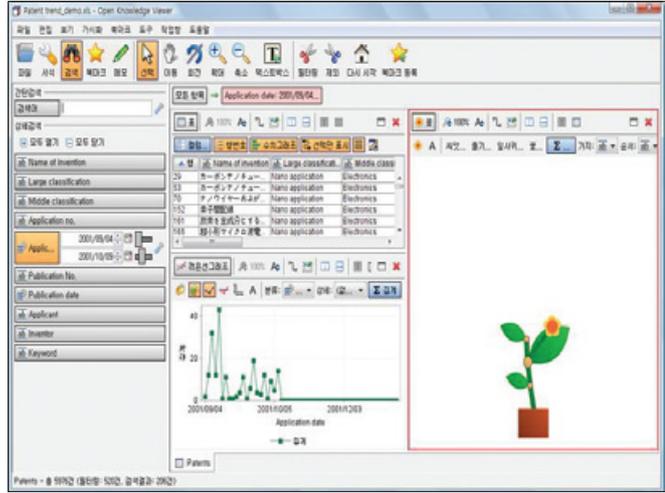
꽃 모델

<그림 1>은 꽃이 성장하는 그림에 비유하여 특허가 시간의 경과에 따라서 변화하는 특징을 보여주고 있다. 이 그림은 '오픈놀리지 뷰어'를 사용하여 생성한 그림인데, 이 소프트웨어는 필자가 개발한 데이터 가시화를 위한 범용 패키지 소프트웨어다. 이 그림을 보면 왼쪽 위에는 원래의 데이터가 표 형식으로 정리되어 있다. 왼쪽 아래에는 이 표를 사용하여 작성한 꺾은선그래프가 있다. 그러나 이 그래프만 봐선 데이터의 단면밖에 볼 수가 없다.

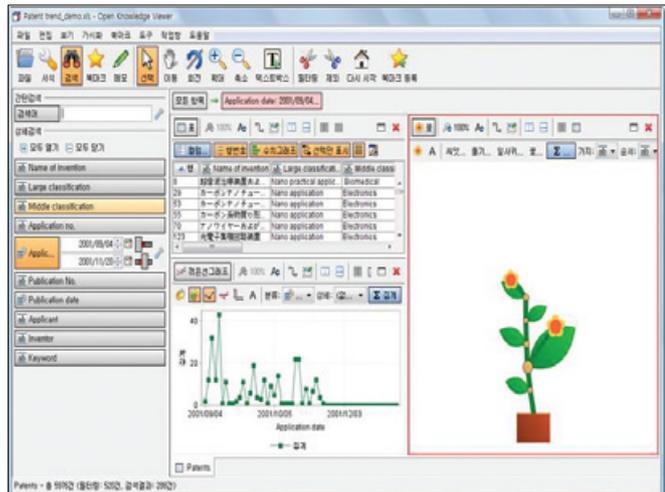
한편 그림의 오른쪽에는 꽃이 그려져 있는데 꽃의 각 부분, 즉 흙, 줄기, 잎, 꽃에 해당하는 부분은 데이터의 각 속성과 대응하고 있다. 화면의 가장 왼쪽에 있는 컨트롤 패널에서 기간을 조정하면 꽃이 성장하는 모습으로 변화한다. <그림 2>는 <그림 1>에 대하여 기간을 좀 더 연장하여 나타낸 그림이다. <그림 2>를 보면 <그림 1>보다 몇 가지 분야에서는 많은 성장이 있는 것을 한 눈에 수 있다.

데이터를 꽃으로 그리면 검색 결과도 꽃으로 나타낼 수가 있다. <그림 3>은 그림에 대하여 직접 가시화 검색을 한 결과이다. 왼쪽 위에 있는 검색창에 검색어로 'sony'를 입력하면 검색 결과는 표, 꺾은선그래프, 꽃의 형상에 직접적으로 반영

■ 그림 1 일정기간중의 특허 출원 건수의 추이를 꽃 모델로 표현



■ 그림 2 그림 1의 기간을 연장한 결과



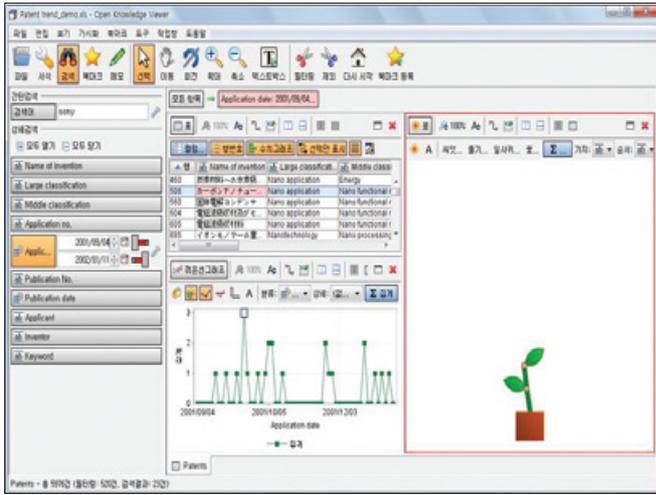
되어 소니사만의 특허가 나타난다.

<그림 1>과 <그림 2>를 연속적으로 표현하면 특허 출원 데이터를 마치 한편의 영화를 보는 것처럼 데이터의 변화를 확인할 수 있다. 이 경우 흙은 재료, 줄기는 가공 기술, 잎은 분야, 꽃은 실용화와 대응시켜서 그린 것이다. 탄소 나노 튜브에 전혀 지식이 없는 사람이라도 이 그림을 연속적으로 보면 관련 특허출원의 특징을 쉽게 알 수 있다. 즉 재료에 관한 특허가 가장 먼저 나오기 시작한 후 이어서 가공기술에 관한 특허가 출원됐다.

포도 모델

이어서 각 분야별로 특허가 나오면서 2005년경에는 많은

■그림 3 특정의 회사가 출원한 특허만 검색한 결과



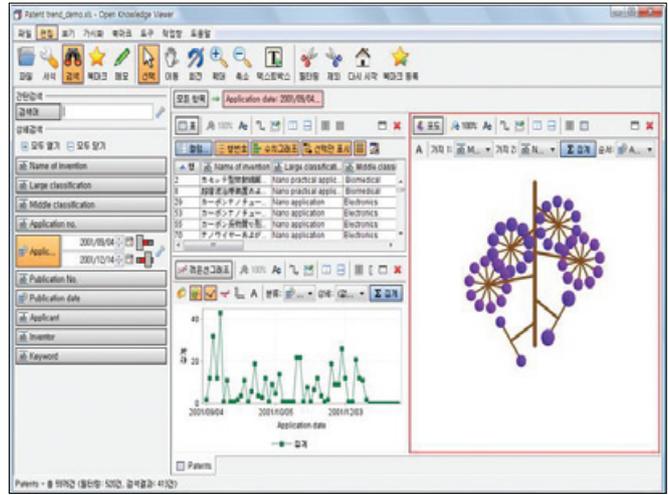
분야에 대한 특허가 출원됐고 실용화도 활발히 진행되는 것을 알 수 있다. 만일 특정한 분야의 특허에 대해 더욱 상세한 분석이 필요하다면 그 부분만 선택하여 분석할 수가 있다. 만약 데이터가 몇 개의 그룹으로 분류되어 있는 경우에는 분류와 시간에 따른 경과를 동시에 나타내어야 한다. 이 경우에는 포도형태로 데이터를 그리는 게 효과적이다. 즉 포도 줄기는 1차 분류를 나타내며 줄기에서 또 다시 분기된 작은 줄기는 2차 분류를 나타낸다. 꽃 모델은 아래에서 위를 향하여 성장하지만 포도 모델은 위에서 아래를 향하여 성장한다. <그림 4>는 포도 모델을 사용하여 데이터가 시간의 경과에 따라서 어떻게 변화하는지를 나타내고 있다.

꽃 모델 적용 시 가변적 데이터 속성 확인

또 다른 예로서 회사의 매출 데이터의 경우를 생각해 보자. 만약 매출 데이터를 구성하는 속성은 기간, 주요 상품, 주요 상품별 매출액, 고객 회사 수라고 한다면, 꽃 모델에서는 줄기는 기간을 나타내고, 잎은 주요 상품을, 꽃은 주요 상품별 매출액을, 흙은 고객 회사 수에 대응시킬 수 있다.

각 속성을 구성하는 데이터는 시간의 경과에 따라서 변화하므로, 속성에 대응하는 꽃의 부분도 시간의 경과에 따라서 변화하게 된다. 회사의 매출 데이터를 한번만 꽃 모델과 대응시키면 그 후로 각 속성을 기준으로 데이터가 변화하는 모양을 꽃의 모양이 바뀌는 것으로 비유해서 볼 수가 있다. 만약 기간을 좁히거나 넓히면 줄기의 길이가 변화할 것이며, 주요 상품별 매출액의 변화를 살펴보면 마치 꽃이 피거나 지는 것과 같은 형상을 나타낼 것이다.

■그림 4 포도 모델로 표현한 특허 출원 건수의 추이



실제로 꽃 모델을 데이터의 속성과 대응시킬 경우에는 꽃 모델의 각 부분과 데이터 속성간의 대응관계는 고정적이지 않으며 가변적이다. 만약 데이터를 꽃 모델로 표현하기 위한 데이터의 속성이 네 가지이며 꽃의 각 부분이 네 가지라면, 각각을 대응시키는 방법은 적어도 16가지 이상이 된다. 똑같은 데이터라도 대응관계를 바꾸게 되면 시간경과에 따라 꽃의 성장 모양이 달라질 것이다. 성장 모양이 달라지면 그 모양을 보고 해석하는 게 달라질 가능성이 높다.

이러한 시각적인 특징을 이용하면 새로운 비즈니스 모델의 설계가 가능해진다. 즉 데이터를 그림으로 그리는 서비스와 그림을 해석하는 서비스가 가능해 지는 것이다. 이러한 비즈니스 모델은 데이터 센터의 운용 시에 소프트웨어나 플랫폼의 이용이라고 하는 비즈니스에 더하여, 데이터의 가시화 서비스나 데이터의 해석 서비스와 같이 세분화된 비즈니스 모델로 개발될 수 있다.

데이터를 그림으로 나타내기 위하여 꽃 모델이나 포도 모델을 사용하는 목적은 자연계에 있는 식물의 성장과정을 참고로 하여 시간의 경과에 따른 데이터의 변화를 은유적으로 나타내고 이해하려는 데 있다.

일반적으로 사용하는 꺾은선그래프를 사용하면 데이터의 한 가지 속성을 나타낼 수는 있지만 여러 개의 속성을 동시에 나타내기는 어렵다.

따라서 데이터가 시간의 경과에 따라서 어떻게 변화하는지를 다각도로 분석하여야 할 경우에는 꽃이나 포도와 같이 성장주기를 알기 쉬운 모양으로 비유하여 표현하고 해석하는 게 효과적이다. 🌿