

인지과학에서 빅데이터까지, 미리 보는 지식 세상 - 지식서비스공학

앨빈 토플러는 『권력이동』이란 책에서 "지식을 가진 자가 결국 지배력을 가진다"고 주장했다. 왜 그럴까.

지식이 '의사결정'과 관련되기 때문이다. 의사결정이란 상황을 판단하고, 몇 가지의 대안 중에서 무언가를 선택하는 행위다. 지식은 이런 의사결정의 연료이자 재료다. 충분한 지식이 없으면 제대로 된 의사결정을 할 수 없다. 아는 게 없는데 결정을 해야 한다면 주사위를 굴리거나 눈을 감고 운전하는 것과 다를 게 없다. 보다 많은 양질의 지식을 사용해야 더 나은 의사결정을 할 수 있다는 건 이제는 상식이다.

알아야 이긴다! 지식이 지배하는 세상

경쟁을 하는 사회라면 지식과 의사결정의 중요성은 더 커진다. 높은 사람은 더 크고 오래 영향을 미치는 결정을 한다. 이런 결정에는 더 많은 지식이 필요하다. 단순히 양적으로 많거나 세밀한 지식뿐만 아니라, 너무 광범위해서 때로는 연관이 없어 보이는 것도 알아야 한다.

오늘날은 지식이 많이 쌓여 있고 쉽게 찾을 수 있는 시대다. 예를 들어 인터넷에 흘러넘치는 지식은 누구나 접근할 수 있고 대부분 무료다. 지식을 얻는 데 아무 제약이 없을 것 같다. 하지만 그렇지 않다. 남들도 나와 똑같이 지식에 접근할 수 있기 때문이다. 이 속에서 '상대적으로' 남보다 잘 알기란 예전보다 더 어렵다.

문제는 두 가지다. 지식이 너무 많아 오히려 정말 쓸모 있는 지식을 찾아내기가 더 어려워졌다는 점이다. 또 지식이 많아도 내 머리로 효과적으로 활용하지 못하면 소용이 없다는 점이다. 구슬이 서 말이라도 꿰어야 보배인데, 꿰기가 쉽지 않다.

지식서비스는 바로 이런 문제를 해결하기 위해 탄생했다. '지식을 어떻게 발견하고 골라내고 정제할 것인가, 어떤 모습으로 만들어 어떤 시점에 누구에게 제공해야 가장 원활한 의사결정 활동을 할 것인가'를 고민하는 것이 지식서비스공학이다.

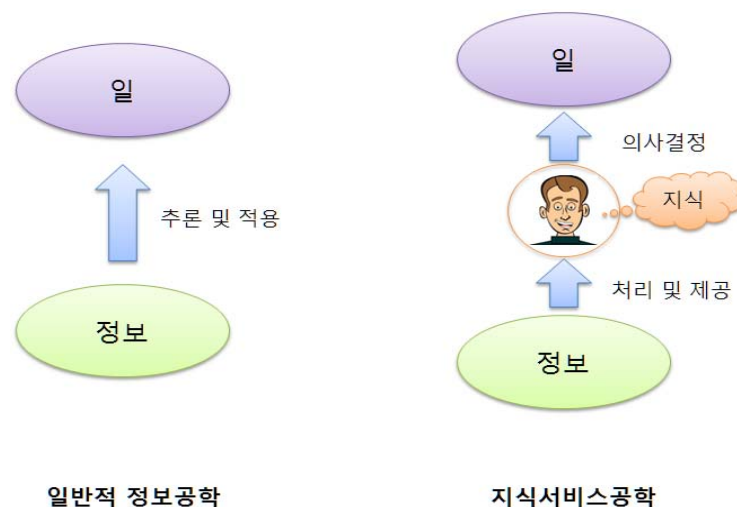


그림 1. 일반적 정보공학과 지식서비스공학의 관점 차이

인지과학: 인간의 지식활동을 파헤치다

현대 과학에서 가장 인상 깊은 장면 중 하나를 소개해 보겠다. 제2차 세계대전 후인 1948년 미국 캘리포니아공대(칼텍)에서 열린 '행동에 있어서의 뇌의 메커니즘'이라는 학술대회에서, 유명한 심리학자인 칼 래슬리(Karl Lashley)가 한편의 논문을 발표했다. 훗날 '인지심리학 혁명'이라 불리는 일대 전환의 서곡이었다.

1910년 이후 대다수 심리학자들은 행동주의 심리학에 서 있었다. 행동주의 심리학은 확 인할 길 없는 인간 내면의 움직임을 설명하려는 시도가 비과학적이라고 무시했다. 오직 관찰할 수 있는 인간의 행동과 그것을 유발한 환경 인자를 연결하는 것만이 실증적인 연구 방법이라고 믿었다. 그런데 1946년 최초의 컴퓨터 에니악(ENIAC)이 만들어졌고 1948년에는 폰 노이만에 의해서 소프트웨어의 개념이 생겨났다.

그렇다면 사람의 마음속에도 컴퓨터 프로그램 같은 '무엇'이 있어서 사람들로 하여금 복잡한 행동을 하게 할 수도 있지 않을까. 그 과정에 대해서 연구하는 것도 비과학적일 이유는 없지 않을까. 칼 래슬리의 논문은 사람이 언어를 구사할 때에도 일련의 계획이 미리 꾸며져서 우리가 쉽게 말을 할 수 있는 것이지, 결코 자극-반응-자극-반응 식으로 연결돼 말을 하는 것이 아니라고 지적했다. 이 학술대회는 당시의 심리학 분위기를 일거에 뒤집었고, 1950년대 중반 인지과학이 탄생하는 데 큰 영향을 끼쳤다.

몇 년 뒤인 1955년 여름, 미국 다트머스 대학교의 젊은 수학자 존 매카티는 IBM에서 일을 하다가 28세 동갑인 마빈 민스키를 만났다. 둘은 곧 '지능적 기계'에 대한 아이디어를 서로 나누고 이듬해인 1956년, 다트머스 대학교에서 10명이 2개월간을 함께 지내며 '인공지능'을 연구하자는 제안을 했다. 인공지능이란 말이 처음 세계에 등장한 것이다. 여기에는 컴퓨터 언어인 어셈블리(Assembly)의 창시자이자 당시 IBM의 패턴인식과 정보이론 연구그룹을 이끌고 있던 너대니얼 로체스터와, 정보이론의 창시자이자 커뮤니케이션 이론의 대가인 미국 벨 연구소의 클로드 섀넌의 지원이 있었다.

인공지능 연구를 위해 다트머스 대학교에 모인 사람들 중에는 위의 4명 외에도 벨 연구소의 아서 사무엘, 허버트 사이먼, 앨런 뉴웰 등이 있었다. 모두가 세계 최고 수준의 학자가 됐다. 사이먼은 노벨 경제학상을 타고, 사무엘은 인공지능 역사에 한 획을 긋는 세계 최초의 자가 학습 프로그램을 만들었다. 당시 30세도 안 된 풋내기 청년이었던 맥카티, 민스키, 뉴웰 등은 그 후 긴 세월 동안 인공지능 분야를 지배하며 수많은 업적을 남겼다. 맥카티와 민스키는 MIT에 세계 최초의 인공지능 연구소를 만들었다. 몇 년 뒤 맥카티는 다시 스탠포드 대학교로 옮겨 또 하나의 인공지능연구소를 만들었는데, MIT와 함께 양대 산맥이 됐다.

뉴웰은 1955년에 인간처럼 체스를 두는 가상적인 프로그램에 대한 글을 썼다. 그것이 사이먼의 주의를 끌었고 둘은 곧 최초의 인공지능 프로그램을 만들었다. 이어서 지금도 모든 인공지능학도가 최우선으로 배우게 되는 사례가 되는 연구를 했다. 뉴웰은 이후 미국 카네기멜론 대학교에서 인지과학, 즉 인공지능과 인간의 지능행동을 연결하는 연구를 계속했고, 1983년, 인간-컴퓨터 상호작용(HCI) 분야에 지대한 영향을 끼친 『인간-컴퓨터 상호작용의 심리학』이라는 책을 썼다. 카네기멜론 대학교는 융합적 연구가 강한 HCI 분야의 중심지가 됐고, HCI 학과까지 가진 드문 학교가 됐다.

1956년은 특별한 해였다. 9월, 다트머스 대학교에서 얼마 떨어지지 않은 보스턴의 MIT에서도 '정보 이론에 대한 심포지움'이라는 모임이 열렸다. 사이먼과 뉴웰, 심리학자 조지 밀러, 그리고 현대 언어학의 창시자인 노엄 촘스키 등이 가세했다. 밀러는 훗날 이때의 감

회를 “실험심리학, 이론언어학, 인지과정에 대한 컴퓨터 시뮬레이션 등은 더 큰 전체의 조각들이며 앞으로 이들의 공통 관심사의 상호연결과 협력이 증가할 것임을 강하게 확신하며” 그 모임을 떠났다고 회고했다.

이 모임은 인지과학의 시초였다. 다트머스 때와 달리 컴퓨터가 아니라 인간의 인지 과정 즉 ‘생각’ 과정으로 주제가 확장됐다. 촘스키의 발표는 1년 후 논문으로 완성돼 이론언어학 분야에서 인지 혁명을 일으켰다. 촘스키는 MIT 교수로 재직하며 지금도 왕성한 활동을 하고 있고 ‘현대 언어학의 아버지’ ‘20세기 최고의 지성’으로 존경 받고 있다. 하지만 당시 그는 맥카티나 민스키, 뉴웰보다도 젊은 만26세의 청년이었다.

인간 기억의 한계 용량은 얼마인가

1956년의 MIT 모임에서 밀러는 과학논문 치고는 기묘한 제목의 논문을 발표했다. 「마술숫자 7 ± 2 」라는 논문이었다.

인간의 기억장치는 뇌다. 그러나 해부학적으로 뇌의 기억방식이나 용량을 알아낼 수는 없다. 컴퓨터를 뜯어서 소프트웨어를 읽어낼 수 없는 것과 같다. 그런데 밀러는 여러 가지 과거의 실험 데이터를 통해서 인간의 기억 중에 단기기억(모든 사고활동을 뒷받침하는 필수적이지만 유지 기간이 짧은 기억)의 한계 용량을 추정해 냈다.

생각해 보자. 눈앞에 쌀알이 흩어져 있을 때 한 번에 셀 수 있는 수는 몇일까. 한 번에 구분할 수 있는 음정의 수, 색상의 수 등에서 우리 뇌는 공통적인 한계를 느낀다. 이렇게 인간들의 지적 활동은 여러 방면에 걸쳐서 비슷한 가짓수의 제한을 받는다. 밀러는 실험적으로 인간의 정신 능력의 구조와 한계를 알아내고 설명했는데, 이렇게 할 수 있다는 것 자체가 인지과학의 앞길을 보여줬다는 평가를 받는다.

이렇게 컴퓨터과학, 인지심리학, 인공지능, 인지과학 등의 새로운 학문이 불과 10년 동안에 나타났다. 젊은 천재들이 새로운 학문에 뛰어든 덕분에 각 분야는 무서운 속도로 지식을 확장해 나갔다. 한편으로는 인간의 사고 과정에 대해서 다른 한편으로는 인간을 닮거나 필적할 컴퓨터 모형에 대해서 발견과 업적이 쌓였다. 1970년대가 끝나갈 무렵엔 인지심리학, 언어학, 인공지능 등에 기반을 둔 융합학문인 인지과학이 어엿한 학문으로 자리를 잡았다.

1980년대 초반에는 ‘인지공학’이란 단어가 탄생했다. 지금도 HCI 분야에서 최고의 권위자로서 활약하고 있는 카이스트의 겸임교수 도널드 노먼 교수가 1981년 「인지공학으로의 발걸음」이란 보고서를 쓴 것이 시초다. 인지공학이란 인간 정신에 대한 이해를 바탕으로 사물과 시스템을 설계하고 분석하고 평가하는 모든 공학적 활동을 말한다.

이후 인지공학은 컴퓨터와 전자기기를 주대상으로 한 HCI 분야와 복잡한 인간의사결정을 다루는 인지시스템공학 분야로 나뉘어 각기 발전했다.

1980년대 일본의 야심을 무너뜨린 애플의 충격

HCI(인간-컴퓨터 상호작용)에 대해 들어본 독자가 많을 것이다. 그만큼 최근 중요하게 여겨지는 과학 분야다. 오늘날 소위 ‘사용자 인터페이스’나 ‘UX 전쟁’이니 하는 것들이 다 HCI라는 분야에서 다루는 일이다. 예를 들어 스마트 폰의 애플리케이션을 쓴다고 해보자. 어떤 애플리케이션이 다른 것보다 쓰기 좋다고 할 때 “인터페이스 설계가 잘됐다.”고 한다. 이것을 ‘사용자 인터페이스(UI, User Interface)’라고 하는데 요즘은 기업의 중요한 경쟁력이 되었다. 하드웨어든 소프트웨어든 사용자 친화적으로 설계하지 않으면 외면받기 때문이다.

HCI는 1980년대부터 발전했는데, 개인용 컴퓨터의 발전과 연관이 많다. 당시 사람들은

기업을 중심으로 사회의 정보화가 한껏 이뤄졌다고 생각했다. 관계형 데이터베이스와 고급 언어가 보편화됐고, 컴퓨터 교과서는 “컴퓨터가 성숙단계인 제4세대로 접어들었다.”고 자평하고 있었다. 집에서 쓸 수 있는 컴퓨터까지 나오기 시작한 상황이었으니까.

이런 상황에서 1982년 당시 승승장구하던 경제 강국 일본이 국가 프로젝트로 제5세대 컴퓨터를 만들겠다고 선언해 세상을 긴장하게 했다. 문명학자들이 앞 다퉈 예언한 미래 지식 사회를 일본이 확실히 주도하겠다고 야심을 보인 것이다. 그 주안점은 하드웨어에서는 병렬 컴퓨터 기술을 통한 성능 향상이었고, 소프트웨어에서는 인공지능이었다. 그런데 1970년대 세계 전자시장을 석권하고 1980년대에 자동차 시장을 확대하며 자신감에 충만해 있던 일본에게는 불행하게도, 정보지식사회는 전혀 예기치 못했던 길로 다가오고 있었다.

1981년 미국 캘리포니아 팔로알토 시에 있는 제록스 사 PARC 연구소에서 새로운 유형의 컴퓨터 시스템이 개발되었다(그림 2좌). 아직 개념 설계 상태였던 이 컴퓨터를 눈썰미 좋은 애플 사가 받아들이고 마우스를 적극적으로 개선하여 채용했고 2년 뒤 리사(LISA)라는 컴퓨터가 돼 세상에 새로 나왔다. 애플은 다시 1984년 ‘매킨토시’라는 이름으로 이 개념을 대중적으로 보급시키는 데 성공했다(그림 2우). 우리가 오늘날 그래픽 사용자 인터페이스(GUI, Graphical User Interface)라고 부르는 것이 탄생한 것이다. 여기에는 윈도우, 아이콘, 메뉴 등이 있고 초기적인 마우스도 포함돼 있었다. 오늘날 우리가 PC의 윈도우, 스마트폰, 맥북에서 볼 수 있는 거의 모든 요소가 다 있었다!



그림 2. 초창기 GUI를 보이고 있는 Xerox 8010 Star 워크스테이션과 애플 매킨토시

매킨토시의 혁신은 대성공을 거뒀다. 컴컴한 화면에 투박한 녹색 문자만 보여주던 컴퓨터 스크린은 갑자기 원시적으로 보였다. 단순히 예쁜 모습이 다가 아니었다. 사람과 컴퓨터의 대화 방식을 혁명적으로 변화시켰기 때문이다. 컴퓨터가 독자적으로 사람처럼 똑똑해지는 대신, 사람과 상호작용하는 방법을 아주 인간답게 바꿨다. 애플이 선택한 길로 마이크로소프트사 등 세상이 다 따라갔다. 일본이 좌절한 것은 이 부분이었다. 10년에 걸쳐 4억 달러 이상을 쏟아 부은 제5세대 컴퓨터 프로젝트는 일반 기업의 상업용 컴퓨터의 발전에 추월당한 채 막을 내렸다. 요즘은 더 이상 컴퓨터의 세대를 세는 사람을 찾아볼 수 없다.

모니터 속 ‘휴지통’과 인터넷 브라우저의 HCI

사람과 컴퓨터 사이의 상호작용 방식이 하루아침에 바뀌다시피 하자 HCI 연구자들이 갑

자기 바빠졌다. 이전의 문자 기반의 화면에서는 메뉴나 명령어를 어떻게 구성하는지 등만 연구하면 됐다. 하지만 이제는 윈도우의 적절한 사용, 글자 폰트의 크기, 자유로워진 화면 디자인에 따른 시각적 배치와 차별화, 마우스의 사용 등 인간공학적 문제와 인지공학적 문제를 연구해야 했다. 연구 범위가 10배는 넓어졌다.

그중에서도 가장 중요한 것은 사용자 인터페이스였다. 매킨토시의 ‘윈도우’는 그냥 문서를 여러 화면에 나누어 겹쳐 놓은 게 아니었다. 마치 책상 위에 놓은 문서들 같았다. 거기엔 또 폴더가 있었고 폴더를 클릭하면 폴더가 열리며 새로운 윈도우가 됐다. 한 폴더에서 다른 폴더로 파일을 옮기려면 예전에는 이동(Move) 등의 명령어를 써야 했지만, GUI에서는 그냥 마우스로 집어 여기에서 저기로 옮겨 넣으면 됐다. 게다가 한 편에는 ‘휴지통’이 있었다! 휴지통에 파일을 끌어넣으면 없어졌다.

이런 비유(메타포)의 등장은 사람의 마음 속에 있는 일상의 기본 지식을 응용해야 한다는 사실을 말해주고 있었다. 즉 정신 모형(멘탈모델)까지 HCI에서 다뤄야 하는 것이다 이제야 인지과학과 인지공학을 배경으로 한 HCI가 그동안 쌓은 전문성과 능력을 제대로 발휘해 볼 기회를 얻게 됐다.

그러나 사람들은 곧 컴퓨터가 ‘정보를 담고 처리하는 기계’가 아니라 ‘정보의 창문’이었음을 깨닫게 됐다. 네트워크의 등장 때문이었다. 인터넷의 전신인 알파넷(ARPANET)은 미국 국방연구소인 다르파(DARPA)의 연구용으로 1969년부터 존재했다. 1982년에 알파넷의 TCP/IP가 표준이 되면서 인터넷이라는 공용망이 탄생했다. 그러나 네트워크의 폭발은 아직 무언가를 기다려야 했다.

인터넷의 폭발적인 성장은 1990년대 중반에 있었다. 유럽입자물리연구소(CERN)의 한 연구원 덕분이다. 팀 버너스-리 연구원은 연구소 내의 자료가 어느 컴퓨터에 있는 하이퍼텍스트의 개념으로 서로 참조해볼 수 있도록 하기 위해 1991년 최초의 서버와 최초의 사이트, 최초의 브라우저를 만들었다. 1992년 11월 서버는 26대가 됐고, 1993년 10월엔 200개로 늘었다. 1993년 미국 일리노이 대학에서 윈도우와 맥OS에서 사용할 수 있는 ‘모자이크’라는 웹 브라우저를 개발했다. 요즈음 우리가 보는 웹 브라우저와 거의 비슷한 브라우저다. 이 브라우저로 당시에 적은 수지만 박물관 같은 사이트를 방문해 본 사람들의 놀라움은 엄청난 것이었다. 그리고 그야말로 폭발이 일어났고 1995년엔 거의 모든 나라의 신문들이 새로운 세상 월드와이드웹의 이야기를 다루게 됐다. 2013년 현재 전 세계적으로 5000만 대 이상의 서버가 가동되고 있는 것으로 추산된다.

인터넷의 초기 역사 또한 지식서비스와 인간-컴퓨터 상호작용(HCI)이 얼마나 결정적인 역할을 하는지를 잘 보여주는 또 하나의 사건이다. 학문이나 일 때문에 컴퓨터를 쓰는 사람들은 이미 1980년대 중반부터 인터넷을 통해 메일을 주고받고 토론을 즐기고 있었다. 하지만 그들만의 이야기였고, 혁명은 일어나지 않았다. 아무도 인터넷을 하려고 컴퓨터를 사는 때가 아니었으니, 그 외에 ‘컴퓨터를 쓸 일’이 있는 교수, 대학원생, 연구원, 프로그래머 등 극소수만이 이메일을 교환했다. 그런데 하이퍼텍스트라고 하는 지식 조직을 적용하자 전에 없던 유용성이 발생했다. 그리고 하이퍼텍스트 공간을 돌아다닐 수 있는 GUI 기반의 사용자 친화형 브라우저가 나타나자 인터넷은 누구나 접근할 수 있고 누구나 쓰고 싶어 하는 공간이 됐다. 비로소 PC가 누구나에게 필요한 전화 같은 존재가 되었고 네트워크에는 빅뱅이 일어났다.

가장 가까운 HCI 변화는 손안에!

1990년대 중반에는 또 하나의 큰 변화가 발생했다. 휴대전화의 급격한 보편화다. 통신시스

템 전체가 변했다. 휴대전화들은 단순한 전화가 아니었다. 메시지 교환, 일정 관리, 카메라 등 많은 부수 기능을 가졌다. 1990년대 중반의 HCI 전문가들은 너무나 바빠졌다. 어떻게 하면 보다 사용하기 좋은 웹페이지를 만들 수 있는 것인가, 어떻게 쓰기 쉬운 휴대전화를 설계할 것인가 연구를 해야 했다. 폭주하는 수요를 따라 많은 사람들이 UI(사용자 인터페이스) 연구자 내지 전문가가 됐고, 웹페이지를 개발하던 프로그래머들도 UI와 HCI 분야의 전문 용어를 듣고 쓰며 살아가는 반전문가가 됐다. HCI의 전성시대가 열린 것이다.

2007년 6월, 스티브 잡스의 애플은 세계가 아직도 너무 느리게 움직이고 있다는 듯 또 하나의 사건을 세상을 향해 던졌다. 바로 스마트폰 시대를 가져온 아이폰의 출시다. 그 정도의 기능을 가진 PDA가 이미 많이 나와 있었지만 보편화되지 못했다. 그런데 아이폰은 전혀 다른 파장을 일으켰다. 많은 사람들은 그 차이가 무엇인지 잘 몰랐고 단지 큰 변화의 시작이라는 사실만 희미하게 감지했다.

바로 이듬해 애플이 도입한 앱 스토어는 또 다른 충격을 가져왔다. 아이폰은 작은 컴퓨터 기능을 담은 잘 설계된 기계가 아니었다. 애플은 생태계를 만들었고 아이폰은 그 창구였다. 요컨대 데이터뿐 아니라 소프트웨어까지도 네트워크 세상에서 사용자가 자유롭게 선택하게 된 것이다. 매킨토시 때부터 ‘휴먼 인터페이스 가이드라인’을 통해 강력하게 사용자 인터페이스를 통일시켜온 애플의 철학은 아이폰에서 더욱 강화됐다. 중심 기능에 충실하도록 단순화해, 사용자가 아이폰 안에서라면 어떤 새로운 소프트웨어도 두려움 없이 써 볼 수 있게 만들어 놓았다. 프로그램이 마치 데이터처럼 자유롭게 선택되고 선택받는 장이 설 수 있게 한 것이다. 생태계가 발명된 셈이다.

원전 사고와 인지시스템공학

1979년 3월 28일 새벽 4시, 미국 펜실베이니아 주의 스리마일 섬 원자력발전소(이하 TMI)가 고장을 일으켜 냉각수 상당량이 유실됐다. 게다가 운전원이 상황을 오판해 자동 비상급수 시스템을 수동으로 꺼버렸다. 이 때문에 결국 원전의 노심이 녹아내리는 미국 원전 사상 최악의 사고가 일어났고, 이후 30년 동안 미국이 다시 원전을 만들지 못하게 할 만큼 큰 파장을 몰고 왔다.

그런데 그 운전원의 실수는 미숙한 훈련 탓도 있지만, 그보다는 발전소의 인터페이스 설계에 있는 몇 가지의 결함 때문인 것으로 판정이 났다. 모호한 신호나 잘못된 계기 배치 등의 HCI적인 문제도 있었지만, 더 큰 문제가 있었다. 발전소 제어실의 위쪽에는 많은 수의 수저통만한 크기의 경보창이 매트릭스 형태로 배치돼 있었다. 한 가지 문제가 있을 때마다 그 문제가 간략히 표기된 창에 불이 켜진다. 유사시에 대비해 정보를 주는 것이다. 그런데 TMI의 상황에서 어떻게 됐을까. 사고 후 첫 1분 만에 무려 500개의 경보창에 불이 들어왔다. 2분이 지나자 800개의 경보가 추가됐다. 운전원이 상황을 파악할 수 있었을까? 운전원은 오히려 공황 상태에 빠져버렸다.

이것은 인간의 정신적 능력이 순간적으로 얼마나 많은 정보를 처리해 낼 수 있는지 고민하지 않고 시스템을 만들었기 때문이다. 이 사건은 인간이 운전하는 대형 시스템의 설계 방법에 대해 다시 생각하게 하였다. 인간중심적인 시스템 설계가 필요한 것이다. 이 분야를 연구하는 분야가 인지시스템공학이라는 분야다.

인지시스템공학은 주로 몸집이 큰 시스템을 대상을 다룬다. 사람이 그 시스템의 일부가 돼 참여하는 분야로서 원자력발전소의 제어실이나 항공기 조종실이 그 예다. 이 분야에서는 주어진 바로 ‘그 일들’을 하는데 도움이 안 되는 인터페이스는 어떤 이유로도 정당화될 수

없다. 그래서 작업을 면밀히 분석해야 하고, 어떻게 하면 최고의 성과를 낼 수 있는 시스템을 설계할 수 있는지 연구해야 한다. 특히 대형 시스템에서의 인간의 실수는 원자력 발전소의 고장이나 비행기 사고와 같은 큰 재앙으로 이어지므로 어떻게 인간의 실수를 예방하도록 시스템을 설계하고 사람을 훈련하고 안전성을 평가하는가 하는 문제 역시 비중이 크다.

실수만이 문제가 아니다. 인간의 정보 입수, 종합, 판단 능력에 관해서도 잘 알아야 한다. 모두 인지심리학과 인지과학이 연구해온 과제다. 이를 실제 상황과 결합해 공학적으로 응용해야 한다. 또 사람의 시스템에 대한 지식, 시스템의 설계 문제 등을 고려해야 한다.

이 분야는 덴마크 RISO 연구소의 앤스 라스무센(Jens Rasmussen)이 세운 '직무영역분석'을 통해 발전소, 병원, 비행기, 함선 등에 응용되고 있다. 인간의 실수에 대해서도 실수 과정을 분석하고 방지하며, 실수에 영향을 받지 않는 시스템을 설계하는 등의 연구가 있었다. 한편 정보통신 기술의 발달, 인터넷과 인트라넷의 발달로 기업 역시 변화해 왔는데, 결국 인간의 의사결정, 판단, 상황 진단 등을 중요시하는 거대 시스템이 되어가고 있다. 인간의 의사결정이 중요한 거대 시스템은 지식서비스공학의 대상이 된다. 그러니 지식서비스공학은 정보-지식 기반의 기업 시스템에 적용되는 인지시스템공학의 역할을 하고 있다고 말할 수 있다.

데이터 뭉치에서 보석을 캐내는 비즈니스 인텔리전스(BI)와 데이터마이닝

기업에서는 경영자의 의사결정이나 기획업무를 제대로 뒷받침하기 위한 시스템의 연구가 1980년대에 유행했다. 그때까지의 경영정보시스템은 기업 내 데이터의 분류-저장-사용에 머물면서 정작 경영 책임자들의 의사결정을 돕지 못했다. 그래서 의사결정지원시스템(Decision Support System) 개념이 발전했다. 이런 시스템들은 컴퓨터에 통계학이나 산업 공학적인 데이터 분석도구들을 켜겨 넣은 것인데, 의사결정자가 정보를 충분히 얻고 보다 나은 결론을 낼 수 있도록 도와줬다.

이후 비지니스 인텔리전스(BI)라는 용어가 1989년 등장했고, 1990년대 후반부터 널리 쓰였다. BI는 의사결정을 돕는다는 면에서는 인지공학과 목적이 같지만, 수단이 다르다. BI의 수단은 데이터다. '데이터를 더 잘 다룰 수 있는 도구를 가지면 당연히 더 나은 결정을 할 기회를 가진다'는 생각에서 나왔기 때문이다.

실시간으로 데이터를 분석해 유용한 정보를 얻어내는 온라인분석처리(OLAP) 시스템이나 데이터를 여러 차원으로 다룰 수 있는 자료저장과 대화형 사용방법이 주요 연구 주제다. 다양한 차원의 데이터 저장 기술을 다루는 데이터웨어하우징, 데이터 통합, 데이터 품질, 텍스트와 콘텐츠 분석 기법 등도 BI에서 데이터의 효용을 높이기 위한 연구 주제가 된다.

문제는 기업에서 이런 일에 활용할 만한 데이터의 양이 점점 많아지고, 형태도 제각각이라는 점이다. 예를 들어 한 기업이 가지고 있는 디자인 관련 서류, 또는 부품 관련 서류를 생각해 보자. 아무도 그 전체를 쓰임새에 따라 분류하고 저장 관리할 엄두를 내지 못한다. 하지만 귀중한 자료이기 때문에 보관은 한다. 이것을 어떻게 활용할 수 있을까.

오랫동안 축적해 놓은 '데이터 덩어리'로부터 효용이 있는 정보를 얻어내는 방법이 필요하다. 예를 들어, 병철이의 키가 180센티라는 것은 단순 데이터이지만 병철이의 동갑내기들의 키를 모아 평균을 내면 한국의 18세 남자의 키라는 의미를 가지게 되고 정보나 지식이 된다. 또 그것이 5년 전 18세들의 평균 키보다 2센티가 커졌다고 하면 이는 세상의 변화를 알려주는 근거가 된다.

이렇게 단순한 통계적 정보도 시공간적 맥락에 의해서 또는 다른 유형의 지식과 융합하여

지식으로 승화된다. BI가 당면한 문제는 애초에 이런 목적에 맞추어 모아지지 않은 데이터로부터도 도움 되는 정보를 뽑을 수 있어야 한다는 것이다. 많은 양의 데이터에서 체계적으로 의미 있는 정보를 뽑아내는 작업을 ‘데이터마이닝’이라 한다. 정보를 마치 지하자원처럼 캐어 올린다고 해서 붙여진 이름이다. ‘지식 발견’이라고도 한다.

데이터마이닝의 요점은 단지 데이터를 통계 처리하는 것이 아니라 숨어 있는 지식을 발견해내는 기술이다. 데이터마이닝을 하려면 우선 데이터 중의 불순물에 해당하는 비정상 데이터를 찾아 제거한다. 그 뒤 다음 작업들을 한다.

- 연관성 발견 (Association Rule Learning): 변수 사이에 동시적인 움직임 패턴을 발견.
- 군집화 (Clustering): 많은 데이터 중 유사한 것들을 찾아 모음.
- 분류 (Classification): 각 데이터가 미리 정해진 유형 중 어디에 속하는지 판별.
- 연속패턴 발견 (Sequential Pattern Finding): 시간적 동향에서의 패턴을 발견.
- 회귀 (Regression): 여러 변수의 움직임 간의 영향 관계를 설명할 수 있는 구조 발견.

연관성 발견은 우리가 가지고 있는 데이터의 항목들 사이에 혹시 우리가 모르는 연관성이 있는가 찾는 것이다. 만일 내과의 환자 기록 안에 나타나는 증상 50개 중에 서로 동시에 나타나는 증상 집합이 있는지 알아본다고 하자. 경우의 수는 $2^{50}-1 = 1.125 \times 10^{15}$ 라는 엄청난 수가 되고, 컴퓨터로 한 경우를 따져보는 데 0.001초만 소요된다고 해도, 이 모든 가능성을 타진하는 데 3만 5000년이 걸린다.

그러나 잘 생각해 보면, A와 B가 이미 같이 동시 발생하는 경향이 없는데 A, B, C의 조합이 동시발생 경향이 높다고 나오기는 어렵다. 이런 식으로 배제할 자료를 빼고 순서를 잘 정해서 논리적으로 해 나가면 시간을 크게 줄일 수 있다. 간단한 예지만, 데이터마이닝은 이런 식으로 이뤄진다.

이런 방식은 활용할 곳이 많다. 은행에서 많은 대출자들의 과거 기록을 가지고 데이터마이닝을 하면 어떤 특징이 있는 대출자가 가까운 미래에 파산을 하게 되는지 패턴을 알아낼 수 있을 것이고, 대출심사를 보다 정확하게 할 수 있을 것이다.

여기서 잠깐 TMI 사고를 떠올려 보자. 정보만 많다고 사람이 항상 더 좋은 의사결정을 하지는 않는다. 정보가 사람과 일에 맞지 않으면 효과가 없거나 역효과가 난다. 의사결정자의 인지적 문제를 간과하면 안 된다. 너무 적은 데이터 보다는 당연히 더 많은 데이터가 낫다. 그러나 곧 정보란 점점 많아지기 마련이고, 결국 정보의 과부하가 찾아온다.

데이터를 다루고 지식을 뽑아내는 기술이 좀 더 성숙한 뒤에는 지식을 어떻게 관리하고 전달해야 실제로 도움이 될지 고민해야 한다. 지식이 그냥 정보 또는 데이터와 다른 것은 유용하다는 뜻을 포함하는 것이다. 조직되지 않고 표현되지 않은 지식은 유용할 수 없다. 데이터에서 지식을 찾아내는 것을 지식 추출과 데이터마이닝이라고 한다면, 그것을 조직하고 사용할 수 있게 하는 것은 지식모형과 지식표현의 문제가 된다. 최근 지식조직을 다루는 온톨로지와 세만틱 웹의 연구가 크게 발전하고 있다. 또한 지식의 시각화를 통하여 직관적으로 사람에게 전달하는 방법들도 연구되고 있다. 모두 지식을 서비스하기 위한 필수적인 연구들이다. 이젠 지식의 양이 아니라 언제 어떻게 어떤 형태로 사람에게 맞게 지식을 전달하는가 하는 것이 문제가 되기 때문이다. 이렇게 항상 인간 중심적인 문제의식과 관점을 유지하는 것이 지식서비스공학의 특징이다.

빅데이터와 데이터사이언스의 한판 승부

인터넷 시대에, 세상에는 온갖 형태의 데이터가 널려 있다. 혹시 이들을 모아서 정렬하면 의사결정에 힘을 보태주지 않을까? 마침 데이터마이닝 기술도 발전되지 않았는가? 이런 문제 의식에서 나온 주제가 최근 유행어가 된 빅데이터다.

빅데이터 이야기를 이해하려면 일단 평소 안 쓰던 단위를 좀 알아야 한다. 1기가바이트란 단위는 이제 일상적으로 친숙하며 영화 한편 정도를 대략 볼 만한 크기로 압축했을 때 거론되는 단위다. 요즘 휴대용 외장하드는 그 1000배인 1테라바이트를 담는 것이 주종을 이룬다. 그 1000배는 페타(Peta)바이트, 백만배는 엑사(Exa)바이트, 10억배는 제타(Zetta)바이트라 한다.

2007년 전 인류가 가지고 있던 모든 지식을 다 합치면 약 295엑사바이트 정도다. 그런데 2009년에는 그 2배가 넘는 790엑사바이트가 됐고, 2020년엔 35제타바이트 즉 350억개의 외장하드 용량이 될 것으로 예측하고 있다. 멀티미디어 정보를 비롯하여 SNS나 각종 센서, 전자상거래 사이트, 스마트폰의 GPS, 기업 프로세스, 웹 로그 등이 쏟아내는 정보는 어마어마한 속도로 누적되고 있다. 이런 빅데이터는 단지 양만 많은 것이 아니고, 그 형태가 비정형적인 경우가 많다. 또 생성-유통-소비까지의 전 주기가 짧고, 서로 연관성이 높은 등 다루기 까다롭다.

그러나 인간의 기술도 만만치는 않다. 빅데이터를 이미 실제 생활에 활용하고 있다. 여기에는 구글의 활약이 컸다. 빅데이터 기술의 시작은 다량의 데이터를 저장하는 일과 처리하는 일이다. 저장에서는 수많은 하드디스크에 분산 저장하는 기술이, 연산 처리에서는 많은 CPU로 나누어 병렬 처리하는 것이 기본적인 전략이다. 구글에서 개발한 맵리듀스(MapReduce)가 대표적 기술로, 수백억 페이지에 달하는 방대한 양의 웹 페이지의 색인을 구축하는데 쓰이고 있다. 이 기술을 바탕으로 ‘하둡(Hadoop)’이라는 오픈소스 프로그램이 개발됐는데, 빅데이터를 다룰 때 널리 쓰이고 있다. 오라클이나 IBM 등 기존의 데이터베이스 개발사들도 기존 프로그램을 하둡에서 동작하게 할 정도다. 빅데이터에 혼한 비정형화된 데이터 처리를 위해서 NoSQL이라는 데이터베이스 형태가 개발됐다. 기존의 데이터베이스와 달리 위치배열이 고정되지 않고 키와 값을 연결해 저장하는 방식인데, 역시 구글이 효시다.

그럼 실제로 빅데이터를 써서 구체적인 어떤 일을 할 수 있을까? 슈퍼마켓이나 온라인 서점에서 사람들의 구매 동향을 분석하면 누구에게 어떤 상품을 추천하면 효과적일지를 알 수 있다. 세계 최대의 온라인 서점인 아마존은 1억 2000명 이상의 고객 정보와 230만 권 이상의 서적 정보를 보유하고 있다. 고객의 과거 구매 정보를 분석해 특정 제품과 함께 자주 구매하는 제품을 찾아 추천해 주면 판매를 촉진할 수 있다.

자동차 제조사인 Volvo는 고객이 자동차를 운전하는 과정에서 수집되는 수많은 데이터를 본사 분석 시스템에 전송하여 고객의 운전 패턴 뿐 아니라, 차체의 결함이나 잠재적인 소비자 욕구를 찾아내는 데 활용하고 있다. 미국 국세청은 대용량 데이터와 다양한 기술을 결합한 탈세 및 사기 범죄 예방 시스템을 구축했다. 지능형 교통안내 시스템은 GPS로부터 자동차의 주행 스피드를 계산하여 교통 정보를 수집하고, 수집된 교통정보를 바탕으로 실시간으로 최적의 교통 안내 서비스를 제공한다. 미국 슈퍼마켓 체인인 타겟은 고객들의 구매 물품을 분석해 고객이 미래에 무엇을 구매할지 예측하고, 그 예측에 따라 고객에게 해당 물품의 쿠폰과 전단지 등을 발송하여 매출을 높이고 있다.

막대한 양의 데이터를 수집, 저장, 처리하는 기초 기술과 데이터 처리 방법이 있다고 해서 자동적으로 유용한 데이터를 얻거나 잘 활용할 수 있는 것은 아니다. 그 데이터의 정보를 유용한 결과와 연결 지어 생각할 수 있는 능력, 그리고 데이터 처리의 결과를 정확히 판

별하고 해석하는 기술은 인간의 몫이다. 그래서 '데이터 사이언티스트(데이터 과학자)'라는 신종 직업이 생겨났다. 데이터 사이언티스트는 통계, 수학, 인공지능 등의 지식 위에 통찰력과 분석력을 통하여 데이터를 모형화하고 각 분야 의사결정에 유용한 결론을 이끌어내는 기술자다. 이는 지식서비스공학의 가장 중요한 축이 되었다. 최근인 2012년 10월엔 하버드 비즈니스 리뷰에 21세기에 가장 '섹시'한 직업이 바로 데이터 사이언티스트라는 소개가 실려 화제가 되기도 했다. 빅데이터가 야생마라면 데이터 사이언티스트는 로데오 선수와 같다.

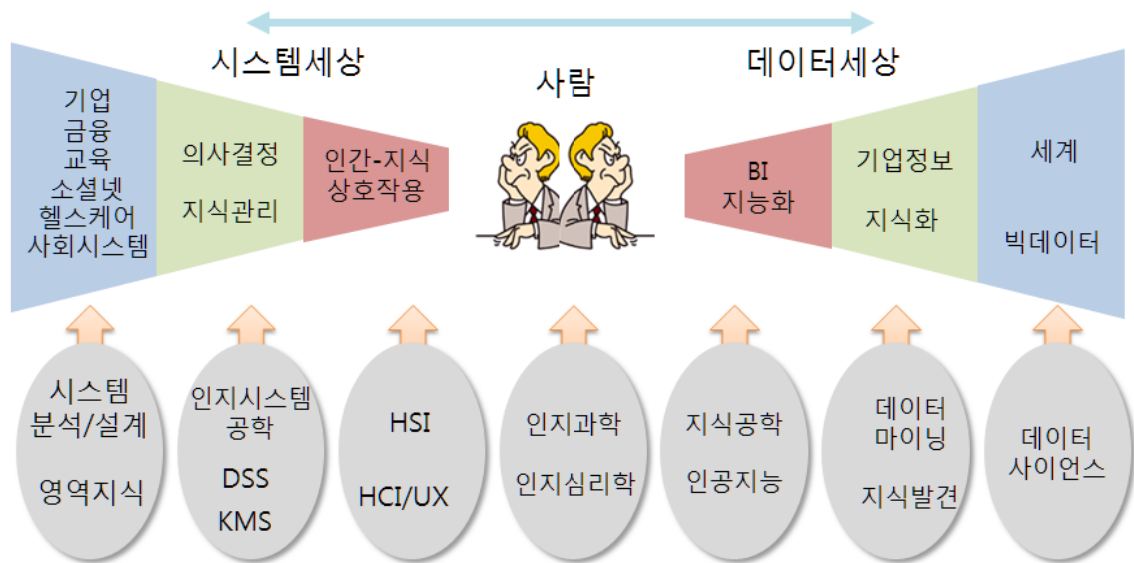


그림 3. 지식서비스공학의 내용 전개 개념도

지식의 도도한 흐름을 탈 미래의 지식서비스공학자

그림 3은 지금까지 다룬 것을 포함하여 지식서비스공학의 내용을 한 눈에 파악하게 펼쳐본 것이다. 우리는 기업 활동을 위시한 시스템의 세상과 데이터와 지식으로 이루어진 정보의 세상을 연결하여야 한다. 그리고 그 가운데에는 사람이 있다. 궁극적으로는 사람이 해야 하고 사람을 통하여야 한다. 인공지능도 데이터사이언스도 HCI도 모두 사람의 능력으로 통일되어야 하며, 이것을 서비스라 한다. 이것을 아우르려는 것이 지식서비스공학이다. 정보기기를 만드는 전자회사도 재화를 다루는 금융기업도 경영을 지원하는 컨설팅 회사도 모두 점점 더 지식서비스공학을 필요로 한다.

지식서비스공학은 마치 시대의 조류와 바람을 제대로 탄 범선과 같다. 배는 조류나 바람을 거스를 수 없지만 제대로 흐름에 올라타면 그보다 더 빨리 배를 목적지에 데려다주는 것도 없다. 배의 운명은 조류와 풍향에 의해 결정된다.

오늘의 사회가 지식사회라는 것은 문명학자들이 예언한 시대의 조류요 풍향이다. 아무도 그것을 부인하지 못한다. 지식서비스공학은 바로 그 조류 위에 떠 있는 배다. 그리고 지식서비스공학을 전공하는 사람들은 시대의 흐름의 맨 앞을 용감하게 향해하는 선장이다. 그 앞길은 도도하게 흐르는 시대의 조류가 열어 줄 것이다. 그들이 개척자로서 해낼 일도 시대가 차례로 제시해 줄 것이다. 세상을 바꿀 장쾌한 항해에 참여하고 싶지 않은가. 지식 사회의 망망대해가 당신 앞에 펼쳐져 있다.

