

# 1. 정보는 생각의 표현

2010 데이터로 표현하는 세상 요약본  
고려대학교 김현철 교수  
hkim64@gmail.com

## ‘데이터’와 ‘정보’와 ‘지식’의 차이

우리는 먼저 ‘정보’라는 것이 무엇인지를 생각해 볼 필요가 있다. 어떠한 것에 대한 정의를 내릴 때에는 그것과 비슷하다고 생각되는 것과의 차이점을 생각해 보면 보다 편리할 때가 많다. 따라서 우리는 ‘데이터’, ‘정보’, ‘지식’이라는 용어를 놓고 그 차이점들을 생각해 보도록 하자. 사실 그들을 명확하게 구분 짓는 것은 쉽지도 않고 별 의미도 없다. 왜냐하면 그들의 의미는 모두 인간의 생각에 따라 다르기 때문이다. 그럼에도 불구하고 다음의 예를 가지고 우리가 ‘정보’라는 용어를 어떻게 사용하고 있는지를 보도록 하자.

한 반에 15명은 남자, 나머지 15명은 여자로 구성된 30명의 학급이 있다. 이 학생들의 키를 측정하였다고 하면 30개의 숫자들이 나열되고, 그것은 “데이터”이다. 데이터는 단지 숫자들의 나열로, 그 30개 데이터를 ‘가공’하여 평균을 내보았을 때, 평균이 170cm 가 나왔다고 한다면 우리는 “이 반 아이들의 키가 이 정도이구나”라는 “정보”를 얻게 될 것이다. 이 정보는 책상의 크기를 정할 때에도 사용할 수 있고, 세면대의 높이를 정할 때에도 사용할 수 있듯이 다양하게 활용되는 유용한 “정보”이다. 나아가 남학생들 키의 평균은 173cm 이고, 여학생들 키의 평균은 165cm 하는 또 다른 “정보”를 얻어 낼 수도 있다. 이 정보는 또 다른 의미를 우리에게 줄 수 있다. 하지만, 이 정보는 단지 이 특정한 반 학생들에게만 적용될 수 있는 정보이다. 다른 학년의 학생들에게나, 혹은 독일 학생들에게도 적용될 수 있는 정보는 아니다. 하지만, 우리가 찾은 이 정보에서 우리는 또 다른 상위 레벨의 패턴을 찾아 낼 수 있다. “아, 남학생들의 키가 일반적으로 여학생들의 키보다 크구나”라는 패턴을 생각해 낼 수가 있다. 그리고 이것은 이 특정한 반 학생들에게뿐만 아니라 일반적으로 적용할 수 있는 것이라고 할 때 우리는 이것을 “지식”이라고 부른다.

이 반 학생들의 키 평균이 170cm 라고 하는 정보도 그 반 학생들이나 담임 선생님에게는 중요한, 그리고 의미를 가진 정보이지만, 과테말라의 초등학교 학생에게는 그것은 한낱 “데이터” 숫자에 불과할 것이다. 또한 반대로 어느 특정학생의 키가 177.5cm 이라는 것도 그 반 담임 선생님에게는 데이터이지만, 그 학생 본인에게는 귀중한 정보 혹은 지식도 될 수 있는 것이다. 저자의 이름인 “김현철”도 나를 아는 사람들에게는 의미 있는 정보이며 지식이 될 수도 있겠지만, 그렇지 않은 사람들에게는 그저 4천만 가지의 한국 이름들 중의 하나에 불과한 의미 없는 데이터에 불과할 것이다. 이렇듯 데이터가 누군가에게 ‘의미’를 가질 때 그것을 ‘정보’라고 할 수 있으며, 그 의미를 만들기 위해 데이터에 어떠한 가공을 할 수 있다. 그 ‘의미’는 사람마다 다르게 받아 들일 수 있기 때문에 데이터, 정보, 지식의 차이는 매우 상대적이고 개인적인 것이다. .

## ‘정보’ 단어 풀이

‘정보’를 한자로 쓰면 情報라고 쓴다 ‘정(情)’은 마음속의 의미라는 것이고 ‘보(報)’는 다른 사람에게 알린다는 것이다. 즉, 내 마음속의 생각을 (의미를) 다른 사람에게 알린다는 뜻으로, 사람들끼리 그들의 생각을 의사소통 한다는 의미이다. 하지만 중국에서는 ‘정보’라고 쓰면 그것은 스파이(spy)와 같은 뜻으로 받아들인다고 한다. 실제로 CIA도 중앙 정보국이고, 우리나라도 국가정보원 이라고 하지 않는가. 중국에서는 대신 ‘신식(信息)’이라는 용어를 쓴 한다고 한다. ‘신(信)’자는 신호 할 때의 그 신자이다. 즉 부호화(encoding)된 문자(letter)를 의미한다. ‘식(息)’자는 휴식, 숨쉬 식 이라는 뜻으로, 마음 심(心) 위에 코(自)가 붙어 있는 모습이다. 즉 마음 속의 숨이 코를 통하여 오가는 것을 의미하는데, 이것은 내 마음의 의미, 즉 생각이 내 몸 밖으로 오가는데 그것이 신호 심불(즉, 信)을 통해서 오간다는 의미라고 할 수 있다. 情報에서도 報, 즉 남에게 알려지면 생각이 어떠한 신호 형태, 즉 언어나 문자, 혹은 동작 등으로 표현이 되어야 한다. 마음속의 생각을 다른 사람과 어떤 심볼을 사용하여 의사 소통하는 것을 뜻한다. 그래서 여기서 중요한 두 가지 요소는 ‘의미’와 그것을 남에게 전달하기 위한 ‘기호표현’이라고 할 수 있다.

또한 정보는 영어로는 information이라고 쓰는데 그 의미는 어떨까? 어원을 보면 뒤의 ‘formation’을 주목할 필요가 있다. Formation이 사용되는 information, deformation, conformation, transformation으로, 거기서 사용되는 formation은 ‘form’에서 나온 말이다. 즉 모양을 만들어 낸다는 의미이다. 그렇다면 information은 모양이 없는 존재에 모양을 주입 infusion한다는 것

을 의미한다. 그렇다면 그 모양은 과연 무엇인가. 중국어 ‘신식(信息)’에서 나온 말처럼 머릿속의 생각, 개념, 의미 등을 머리 밖으로 표현하기 위하여 어떤 모양으로 encoding한다. 그 모양이 문자, 숫자와 같은 심볼 기호도 될 수 있고 이미지도 되고, 다른 모양, 예를 들면 패턴, 구조, 배열도 될 수 있을 것이다.

한스 크리스천 폰 베이어가 쓴 ‘과학의 새로운 언어, 정보’(전대호역, 승산 2006)에서는 ‘정보’를 특정 종류의 메시지가 담고 있는 의미와 전문적인 영역에서 메시지를 전달하는 기호(심볼)의 2가지 의미를 가지고 있으며, 정보기술(IT)은 두 번째 의미라고 하였다. 하지만, 둘은 별개의 뜻이 아니며 마음속의 생각이나 의미를 다른 사람에게 전달하기 위하여, 혹은 내가 기록하기 위하여 기호/심볼/형태 등을 사용하여 표현된 것을 의미한다고 생각되며 이것은 information, 정보, 신식 이라는 용어에서 모두 다 표현되고 있다. ‘자연과학(Natural Science)’이 자연을 대상으로 관찰하여 거기서 법칙, 패턴 등을 찾아내려고 하는 것이라고 한다면 ‘정보과학(Information Science)’는 인간의 머릿속의 생각과 의미를 대상으로 그것을 표현하고 가공하며, 숨겨진 패턴과 법칙을 찾아내려는 분야라고 할 수 있다. 정보기술(IT: Information Technology)은 정보통신기술(ICT: Information Communication Technology)라고도 하는데, 이는 정보의 습득, 처리, 저장, 전달 등을 컴퓨터와 통신 기술을 이용하여 다루는 것을 의미한다.

**나에게 의미 있는 것은 정보, 나에게 의미 없는 것은 데이터**

우리 인간의 뇌 바깥에서 관찰되는 것은 대부분 ‘데이터’라고 할 수 있으며, 그러한 우리 주변의 그 수많은 데이터는 우리의 눈, 귀, 코 등을 통하여 우리의 뇌 속에 전달 된다. 이 과정에서 뇌는 우리에게 의미 있을 만한 것만 골라서 전달하며 나머지는 모두 필터링 시켜버린다. 혹은 조합 혹은 가공을 하여 정보의 형태로 받아들이기도 한다. 우리가 오늘 아침 지하철(혹은 버스, 혹은 자전거)로 출근할 때 우리 눈으로 본 사람이 몇 명이나 될까. 그 중 지금 기억하는 사람은 몇 명이나 될까. 아주 아름다운 여자를 봤거나, 아주 우스꽝스런 모습의 사람을 봤거나, 혹은 친했던 고등학교 동창을 만났다면 당신은 아마 그들을 기억하고 있을 것이다. 오늘 마주쳤을 200명의 사람들 중에서 단 세 명만이 당신의 머릿속에 남아 있다면 아마도 그것은 그 세 명이 당신에게 어떤 ‘의미’가 있었기 때문일 것이다. 우리는 수많은 데이터들 중에서 나에게 의미 있는 것만 받아들이고 나머지는 걸러내 버린다. 그래서 우리의 머릿속에 남아 있는 것은 모두 어떤 의미를 지니고 있는 ‘정보’이다.

그러면 우리의 머리 속에서는 도대체 어떻게 눈, 귀, 코를 통하여 들어오는 수많은 데이터들을 자동으로 ‘인식’하여 그 중에서 ‘의미’ 있는 정보를 골라내며, 그러한 정보들을 기반으로 어떻게 그 복잡한 사고와 계산, 그리고 의사결정 과정을 하는 것이며, 그리고 그러한 의미 정보를 어떻게 효과적으로 표현하여 남에게 전달 하는 것일까. 우리의 모든 두뇌 활동은 결국 정보처리 과정이라고 할 수 있으며, 이것에 대해서는 그 동안 수백 년 전부터 많은 심리학자, 인지과학자, 철학자, 정보과학자들이 지금까지도 연구하고 있다.

인간의 머리 속에 있는 정보를 처리하기 위해 먼저 생각해 볼 수 있는 것은 우리의 뇌 속에 존재하는 단편적인 정보의 명시적 표현이다. 내 머릿속에 있는 이 생각을 어떻게 밖으로 표현하여 다른 사람에게 전달 할까이다. 이것을 “정보표현”이라고 말한다. 표현을 위해서는 표현의 도구가 필요한데 일반적으로 “언어”, 즉 말과 글이 여기에 해당될 수 있다. 실제로 우리는 우리의 생각을 말이나 글, 숫자, 그림, 소리, 행위로 표현한다. 좀 더 나아가 표(table), 그래프(graph)뿐 아니라, 서울시 지하철 노선도와 같은 것도 효율적인 정보 표현을 위한 노력이라고 이해할 수 있다.

두 번째로 생각할 수 있는 것은 우리 뇌 속에서 진행되는 정보 처리의 절차이다. 단편적인 생각은 정보에 표현에 가깝지만, 그 단편적인 정보들의 관계, 정보의 처리 방법, 절차 등은 어떻게 표현해 낼 것인가는 어려운 문제이다. “업무 매뉴얼, 장난감 조립순서, 프로그램”은 이러한 정보의 처리 절차를 명시적으로 표현하려고 한 노력의 결과물이라고 할 수 있다. 그리고 그것은 대부분 “정보표현”들의 순차적 연결이라는 방법을 사용하고 있다.

**컴퓨터는 인간의 정보 처리 과정을 자동화한 기계**

20세기를 산업시대라고 말한다. 산업시대의 핵심 키워드는 ‘기계(machine)’이며, ‘기계’의 핵심 특징은 ‘반복’과 ‘자동’이다. 기계만 가진다면 똑 같은 것을 자동으로 계속 만들 수가 있다. 먼 옛날 누군가는 내 머릿속에 가지고 있는 어떠한 정보 처리 과정을 다른 사람에게 전달하여 그 사람이 그대로 하게 하면 좋겠다라는 생각을 했을지도 모른다. 그것이 정보처리과정을 표현한 것이며, 그 과정을 자동으로 처리하는 기계가 바로 ‘컴퓨터’라는 이름으로 등장한 것이다. 이 ‘반복’과 ‘자동’의 의미를 다음의 예를 가지고 다시 한번 생각해보자.

A씨는 은행원이다. 그의 업무는 은행에서 고객의 계좌 잔액에 따라 이율을 매겨 이자를 계산하는 일이다. 잔액에 따라 이자를 계산하는 과정은 그가 이미 잘 알고 있고, 고객이 몇 백 명 정도 밖에 되지 않았을 때에는 A씨는 수작업으로 그 계산 업무

를 처리하는 것이 그리 어려운 일은 아니었을 것이다. 세월이 흘러, 그 은행의 고객의 수가 백만 명이 되었다고 생각해보자. 이자를 계산하는 작업 과정은 변하지 않았으며 간단한 절차이지만, 그 간단한 절차를 “반복”하여 백만 명에게 똑같이 계속 적용하여 일을 처리하는 것은 결코 쉬운 일이 아니다. 한 명의 고객에 대한 계산에 1분의 시간이 걸린다고 하더라도 1백만 분이 걸리며 이것은 쉬지 않고 계속한다고 하더라도 1백만분/60분/24시간/356일 = 약 2년 정도의 시간이 걸리기 때문이다. A씨는 그 간단한 계산 업무를 처리해 줄 수 있는 기계가 필요하겠다고 생각했을 것이다. “기계”의 핵심은 “반복과 자동”이다. 컴퓨터는 그러한 인간의 뇌 속에서의 정보 처리 과정을 자동으로 반복해 처리할 수 있는 기계이다. A씨는 먼저 자신이 이자 계산을 하는 절차를 명시적으로 정확하게 표현하여야 하며, 그것을 프로그램으로 만들어 컴퓨터라는 기계에 넣으면, 그 컴퓨터는 그 프로그램대로 빠른 속도로 반복해서 백만 명의 이자계산을 일분 안에 해 줄 수 있는 것이다. 그 컴퓨터는 A씨의 머릿속에 있는 정보 처리 절차를 “대신” 해주고 있는 것이다. A씨의 머릿속 생각이 명시적으로 표현되어 컴퓨터에 넣어지지 않는다면, 컴퓨터는 아무것도 하지 못하는 바로 고철 덩어리에 불과한 것이다.

컴퓨터는 주어진 정보 처리 절차를 자동적으로 반복 처리해 주는 기계이다. 정보 처리 절차를 만들어서 컴퓨터에게 주어주는 것은 사람이 해야 하는 일이다. 다음 예를 읽고 생각해 보자.

B씨는 문서 철자법 교정 전문가이다. 어떠한 문서이든지 최종 인쇄되기 전에 B씨가 최종적으로 철자의 오류를 수정한 후에 인쇄소로 넘겨진다. B씨는 수십 년 동안 이 일을 해왔기 때문에 철자오류를 정확하게 찾아내고 수정할 수 있는 능력을 가지고 있다. 또 다른 C씨는 그러한 능력을 갖고 있지 못하기 때문에 매번 B씨에게 그 일을 맡겨서 처리했다. 하지만, B씨도 짧은 시간 안에 너무나 많은 문서를 검토해야 하는 경우에는 종종 실수를 범하기도 하고, 시간도 많이 걸린다. B씨의 문서를 읽고 잘못된 철자를 찾아서 수정할 수 있는 능력을 그대로 “반복과 자동”의 능력을 가진 컴퓨터에게 옮겨 놓는다면, 빠르게 자동으로 문서 철자 교정을 할 수 있을 것이다. 그것을 위하여서는 먼저, B씨는 자신의 뇌 속에서의 정보 처리 과정, 즉 철자오류를 발견하고 수정하는 과정을 명시적으로 뇌 밖으로 표현해 낼 수 있어야 한다. 그러나 내 머리 속에서 아무 어려움 없이 정보를 처리했던 일을 명시적으로 표현해 컴퓨터에게 알려주는 것은 무척이나 어려운 일이다. 그것을 정확하게 똑같이 표현해 내는 것은 몇 달이 걸릴 수도 있고 몇 년이 걸릴 수도 있다. 아무튼 B씨는 각고의 노력 끝에 그것을 표현해 내었고 (몇 가지 규칙으로 만들었다) 그것을 컴퓨터 프로그래머에게 주어서 프로그램 코드로 만들어 컴퓨터에 집어 넣었다. 그것은 아마도 우리가 사용하고 있는 문서편집기 워드 프로세서 에 들어 있는 맞춤법 검사 프로그램과 같은 형태 일 것이다. 우리가 컴퓨터에서 그 맞춤법 검사 프로그램을 사용한다는 것은 B씨의 머릿속 생각을 사용하고 있는 것과 마찬가지이다. C씨가 그 컴퓨터를 직접 사용할 수 있다는 것은 다시 말하면 B씨의 머리를 직접 사용하고 있는 것과 같다. 소프트웨어를 사용한다는 것은 사람들의 머릿속의 업무 처리 절차를 사용한다는 것과 같다.

위의 두 가지 예를 보게 되면, 다음과 같은 동일한 절차가 진행되어 감을 알 수 있을 것이다. 즉, 어느 전문가의 머릿속에 들어 있는 특정 업무 처리 절차 즉 절차 정보, 절차 지식을 먼저 명시적으로 표현해 내야 한다. 이렇게 표현된 절차 지식을 일반적으로 “알고리즘”이라고 부른다. 그 알고리즘은 그대로도 다른 사람들과 공유될 수 있다. 업무 매뉴얼이 그런 것이고, 요리 책에 나와 있는 탕수육 요리법이 그런 것이다. 단위작업들이 순서대로 절차대로 정의가 되어 있으며 그대로만 따라 하면 문제가 해결될 수 있기 때문이다. 그 다음은, 그 알고리즘을 반복적으로 자동으로 빠르게 처리하려고 한다면 컴퓨터라는 기계에 그 알고리즘을 전달해야 한다. 그러기 위해서는 그 알고리즘을 컴퓨터프로그램으로 프로그래밍하여 전달해야 한다. 성공적으로 전달이 되면, 컴퓨터는 그 정보처리의 기계, 즉 사람의 생각 기계가 되는 것이다.

우리 머릿속의 생각을 ‘반복, 자동 기계’인 컴퓨터에서 대신 수행할 수 있는 형태로 명시적으로 표현해 내는 과정을 모두 합쳐서 일반적으로 컴퓨팅이라고 부른다. 보다 정확한 의미와 정의는 다음 장에서 살펴보도록 하자.

.끝.