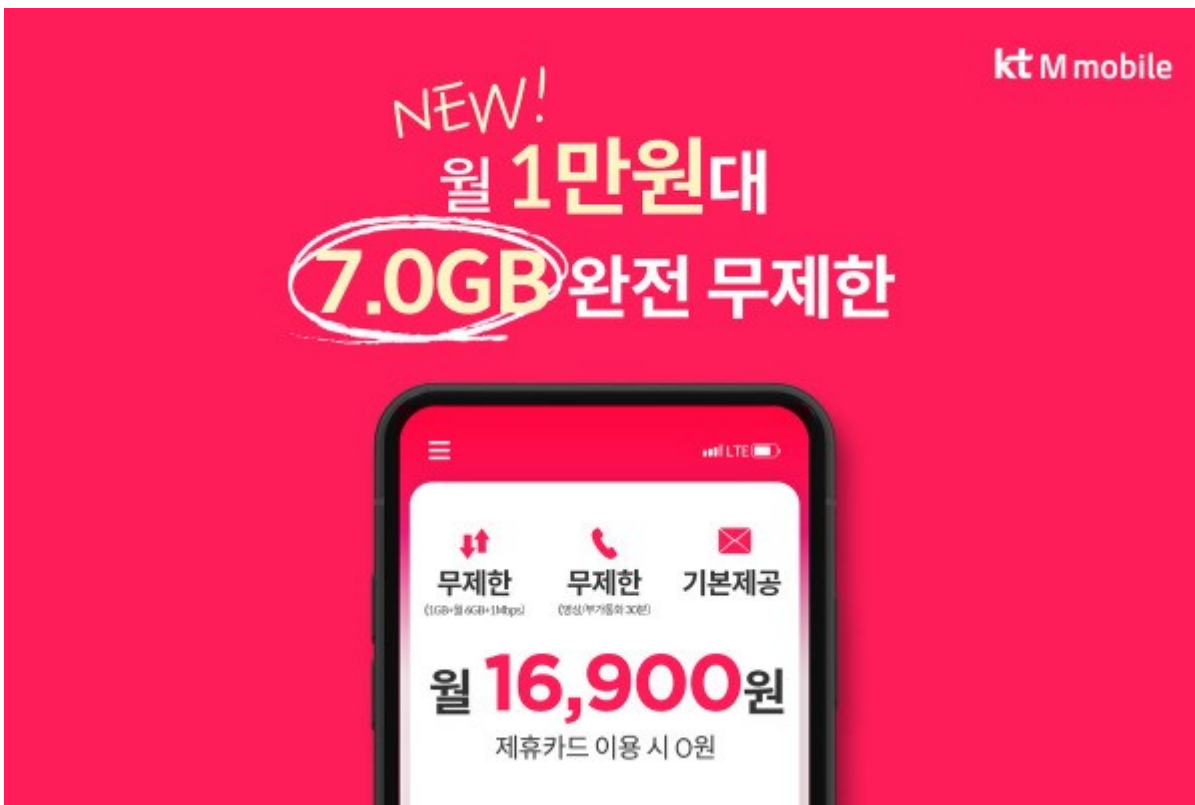


05 벚꽃이 피는 시기와 데이터

정경문

데이터에 대한 이해

(데이터 무제한 요금제



출처 : kt M 모

바일

01 [데이터 = 무제한?] 데이터란 무엇인가?

"데이터"라는 단어를 들으면 어떤 생각이 떠오르시나요? 아마도 "데이터 무제한", "데이터 선물" 이런 말들이 우리에게 가장 익숙하지 않을까 생각됩니다. 그렇다면 우리는 이 무제한의 데이터를 어디에 쓸까요? 유튜브 영상을 시청하거나, 네이버로 검색을 하거나 또는 친구들과 카카오톡으로 대화하고 사진을 보내는데 대부분의 데이터를 사용합니다. 그래서 우리에게 "데이터 = 사용하는 것"이라는 인식이 자연스레 자리 잡은 것 같습니다. 데이터는 이제 우리에게 공기와도 같습니다. 데이터가 없으면 사랑하는 가족과 연락도 어렵습니다. 또 한 은행에서 다른 은행으로 돈을 보내거나 받을 때도 데이터 없이는 직접 은행에 가야 하는 번거로움이 생깁니다. 데이터는 우리 삶을 윤택하게 해주는 소중한 것임에 틀림이 없습니다.

하지만 막상 데이터는 무엇인가요?라는 질문에 답하기란 쉽지 않을 겁니다.

데이터의 어원은 라틴어 "datum"이라는 단어에서 유래되었다고 합니다. 그리고 datum 은 '주다(to give)'는 뜻의 단어 dare의 과거분사형으로 '주어진 것'이라는 뜻입니다. 그 후 '데이터'라는 말은 사람

들 사이에서 보다 추상적인 개념으로 쓰여 왔습니다. 데이터는 "이론을 세우는 데 기초가 되는 사실, 자료"라는 사전적 의미를 가짐과 동시에 "추정 또는 예측을 위한 근거"라는 관용적 의미를 지니고 있습니다.



2022년 4월 인터넷에서 "데이터"와 연관된 단어로 만들어진 워드 클라우드(Word Cloud)

조금만 관심을 가지고 본다면 우리는 데이터를 사방에서 볼 수 있습니다. TV 뉴스에서부터, 신문기사, 연구논문, 회사 경영분석 리포트, 그리고 주식과 부동산, SNS까지 데이터는 우리 생활과 산업의 전반적인 과정에서 핵심적인 역할을 수행합니다. 뉴스 보도가 사실이라고 신뢰를 줄 때 어김없이 데이터가 등장합니다. 또 논문에서 핵심적인 부분은 데이터로 증명합니다. 예를 들면 COVID-19 백신에 대한 효과 입증은 실험군과 대조군의 데이터를 확인하고 검증했기 때문에 신뢰할 수 있게 됩니다.

우리는 데이터, 정보, 지식에 대해 종종 혼동하기도 합니다. 뉴스에서 보도하는 내용은 데이터일까요? 정보일까요? 또는 학교에서 배우는 교과서는 데이터일까요 아니면 지식일까요? 이 물음에 대한 답변은 다음을 통해 설명드리겠습니다.

02 [DIKW 피라미드] 데이터는 모든 것의 근원이다.

DIKW 피라미드는 아래에서부터 **Data(자료)**, **Information(정보)**, **Knowledge(지식)**, **Wisdom(지혜)**를 의미합니다. 이 이론은 미국 펜실베이니아 대학의 러셀 L. 애코프에 의해 처음 언급되었습니다.



가장 아래층부터 살펴보겠습니다.

자료(Data)는 관찰과 실험에 의해 얻은 사실입니다. 이는 데이터가 가공되기 전 있는 그대로를 의미합니다. 이렇게 얻어진 데이터는 어떠한 결론, 판단, 예측을 내리는데 기초가 됩니다.

예를 들어, "올해 벚꽃이 핀 날짜는 4월 10일이다"라고 관찰한 사실은 데이터입니다.

정보(Information)는 이러한 데이터를 가공하여 원하는 형태로 얻어낸 규칙을 의미합니다. 앞서 얻은 다양한 데이터 중에 우리에게 필요한 내용을 얻어내는 것이죠. 통상 관찰된 수치나 값에 대한 평균이나 분류 값 등이 여기에 해당됩니다.

위의 벚꽃 예를 본다면, "올해 지역별 벚꽃이 핀 날짜를 지도상에 표시하면 이렇다"라는 것은 정보입니다.

지식(Knowledge)은 앞서 가공한 정보를 일반화 또는 체계화한 패턴입니다. 우리는 이 지식을 현실에 적용 또는 활용할 수 있습니다. 이 때는 다양한 방면의 정보들이 서로 비교되거나 융합되기도 합니다.

벚꽃의 경우, 작년 개화시기보다 올해 개화시기가 더 빠르다는 비교를 통해 "우리나라 개화시기가 빨라지고 있다"라고 일반화할 수 있겠습니다.

마지막으로 **지혜(Wisdom)**는 지식의 근본적인 원리를 주어진 상황과 맥락에 맞게 적용하는 것을 말합니다. 또는 개별적인/개인적인 문제에 맞게 그 지식을 적용하는 것을 의미하죠. 따라서 지혜는 원리에 대한 깊은 이해를 통해 미래를 예측, 문제에 대한 해결 및 판단을 하는 것입니다.

올해는 늦게 예약을 해서 벚꽃여행 시기를 놓쳤습니다(주어진 상황). 그렇다면 "내년에는 올해보다 좀 더 일찍 경주 벚꽃여행을 예약해야겠다."(벚꽃이 더욱 빨리 피므로) 예측 및 판단합니다.



데이터, 정

보, 지식, 지혜 예시(벚꽃 개화 시기)

03 [GIGO] 쓰레기를 넣으면 쓰레기가 나온다.

"가비지 인, 가비지 아웃(Garbage In, Garbage Out)"이라는 말을 들어보셨나요?

바로 "쓰레기를 넣으면 쓰레기가 나온다."는 뜻으로 IT, 빅데이터 인공지능 분야에서 데이터의 중요성을 의미하는 문장입니다. 위 DIKW 피라미드에서 살펴보셨듯이 정보, 지식, 지혜의 근간이 되는 데이터는 매우 중요합니다. 제 아무리 훌륭한 데이터 사이언티스트, 그리고 인공지능이라도 쓰레기 데이터를 넣으면 쓰레기 데이터 분석 결과가 나오고, 인공지능도 의도했던 목적으로 작동하지 않습니다.

이 용어는 인공지능이나 데이터 분석에서 사용하지만, 실제 우리 일상생활에서도 마찬가지입니다. 정확한 데이터로 결정을 내리지 못하면 잘못된 결과를 초래할 수 있기는 인간도 마찬가지입니다. 그래서 왜곡된 정보를 바탕으로 정보를 내리지 않도록 우리는 여러 사람의 의견에 귀를 기울이는 행동을 하곤 합니다.

컴퓨터의 경우에는 완전히 입력된 데이터를 가지고 결과물을 출력하기 때문에 부정확한 데이터로는 부정확한 결과를 낼 수밖에 없습니다. 즉 출력물의 품질은 입력 물의 품질보다 나을 수 없다는 말입니다.