

1. Data Type

1-1. Character

1-2. Number

1-2-1. integer

1-2-2. float

1-3. Boolean

1. Data Type

데이터 타입은 전기 신호를 저장하고 있다. 컴퓨터를 구성하는 가장 작은 요소는 비트이다. 비트는 뭔가를 저장할 수 있는 공간이다. 0과 1을 저장할 수 있다. 그러나 0과 1로는 많은 경우의 수를 만들어낼 수 없다. 비트에서 자릿수를 늘려서 바이트 단위를 만들었다. 컴퓨터가 처리하는 가장 작은 단위를 바이트라고 한다. 1바이트는 8비트이다. CPU나 버스에 데이터가 이동하거나 저장할 때 바이트로 처리된다.

사람과 컴퓨터가 소통하려면 경우의 수가 많아야 한다. 상대방과 번호가 같아야 소통을 할 수 있다. 사람과 소통할 수 있는 언어를 컴퓨터가 받아들이고 우리에게 다시 보여줄 수 있어야 한다. 모든 컴퓨터에 들어가는 프로그램에 0과 1로 들어갔다가, 이제는 자연어를 직접 입력할 수 있게 되었다. 우리가 가장 기본적으로 컴퓨터와의 소통을 위해 필요한 것은 컴퓨터가 글자를 알아야 한다는 것이다. 컴퓨터가 모든 글자에 일련번호를 매겼는데, 이것이 아스키코드이다. 그리고 외국어까지 모두 포함하는 유니코드가 생겼다.

“자료형은 자료가 어떤 속성을 가졌는지 컴퓨터에 미리 알려주는 역할을 하는 것으로, 정해진 범위 이외의 값을 사용할 경우에는 오류로 인식하게 된다.”[두산백과]

1-1. Character

문자를 표현하기 위해 내부에 저장할 때 Character 타입(문자 타입)을 쓴다. 타입은 무엇으로 결정되나면, 크기가 얼마인지 혹은 내부에 어떤 형태로 구성되어 있는지로 결정한다.

크기와 내부 형태에 따라 컴퓨터의 데이터 타입을 결정한다. 우리가 아는 모든 글자를 일련번호로 만들어놓은 것이 Character 타입이다. Character 타입은 크기가 1바이트에서 4바이트까지 쓴다. 아스키코드는 0~127까지 쓴다. 표준 아스키코드가 8비트를 쓰면 무조건 첫 번째 비트가 0이다. 첫 번째 비트가 1이면 외국어이다. 1이면 어떤 코드를 쓰고 어떤 코드표를 사용해야 하는지 찾기 시작한다. 내가 폰트를 찍을 때 코드표를 찾아서 폰트를 보여주는 것이다.

예) 16진법으로 61은 소문자 a, 20은 스페이스, 0은 30, 41은 대문자 A - 언어적 문자를 집어넣기 시작했고 Character 타입, 문자라고 한다.

Character 타입은 문제가 있다. 컴퓨터가 계산할 때, 0이 코드표(문자)의 0인지 값의 0인지 헷갈린다는 것이다. 아스키코드의 문자는 0은 무조건 따옴표(')를 친다.

1-2. Number

데이터 타입 중에 숫자가 있어야 컴퓨터가 계산할 수 있다. 데이터 타입은 Character,

Number, Boolean 이렇게 3가지 타입이 있다. 로직 연산자의 피연산자는 true, false를 대응시켜놓았다. Relation 유닛은 뿔셈으로 해결할 수 있기 때문에 CPU 내부에 존재하지 않는다. 초기 컴퓨터는 계산만 하다가, 프로그램이 저장되고 흐름이 제어되면서 우리의 말을 알아듣게 된다. 즉 우리의 오감과 서로 상호작용을 시작한 것이다.

숫자는 종류가 크게 두 가지로 나뉜다. 정수와 실수, integer와 float이다. 둘은 내부적으로 표현방법이 다르다. 즉 비트를 저장하는 방법, 값을 저장하는 방법이 다르다.

1-2-1. integer

정수는 기본적으로 2진법으로 크기를 저장하고 있다. 코드가 아니기 때문에 값만 저장하고 있다. 기본이 4바이트이고, 8바이트도 있다. 2의 32승까지의 경우의 수를 표현할 수 있다. 마이너스까지 표현하면 40억 개의 경우의 수를 표현할 수 있다. (-20억부터 +20억까지)

메모리 크기는 byte는 1바이트, short는 2바이트, int는 4바이트, long은 8바이트이다.

1-2-2. float

실수 타입은 float 타입과 double 타입이 있다. “하나의 수를 고정 소수점 부분을 나타내는 부분(가수)과 고정 소수점 위치를 나타내는 부분(지수)으로 나누어 표현하는 표기법인 ‘부동소수점’ 표기법을 사용한다.”[두산백과] float는 4바이트(32비트)의 크기이며 그 중 1비트는 부호, 8비트는 지수, 23비트가 가수 부분이다. double은 8바이트(64비트)의 크기이며 그 중 1비트는 부호, 11비트는 지수, 52비트 가수 부분이다.

소수점이 보통 계산보다 시간이 오래 걸린다. 그래서 실시간으로 빠르게 요구되는 곳은 정수로 만들어서 계산하게 만든다. 정수 계산이 훨씬 빠르기 때문이다.

1-3. Boolean

“불리언(Boolean) 타입은 ‘논리 자료형’이라고도 하며, 참과 거짓을 나타내는 데 쓰인다. 주로 참은 1, 거짓은 0에 대응하나 언어마다 차이가 있다. 숫자를 쓰지 않고 참과 거짓을 나타내는 true와 false를 쓰기도 한다.” [위키백과] 1바이트(8비트)의 크기이다.