

정보 -> 코드(discrete, 빈도수, entropy, 효율적 코드를 구성하자)  
미리 약속 된 코드들을 이용한 것이 디지털 코딩이라고 한다.

아날로그는 무한 개의 심볼을 사용한다.

사실 디지털 코딩으로 하면 아날로그에 비해서는 정보를 잃어버리는 것이다.

정보를 잃음에도 디지털 코딩으로 하는 이유는 효율 때문이다.

아날로그 정보 - 표현, code, 변환 -> 유한개의 심볼을 사용

시간(무한정보) -----> 10개의 심볼 (0~9)한정된 정보

3.156초 이 과정에서 정보가 사라짐      0.156초가 사라짐

디지털 -> 여기서는 바이너리가 아닌, 문자, 숫자, 심볼을 의미한다.

무한 정보를 디지털로 표현하는 이유?

단점 : 정보를 정확하게 표현하기가 힘들다. 정보를 잃을 수 있다.

장점 : 보관, 기록, 처리, 전달 등이 편리

(이것이 너무 편하기에 정보를 좀 잃더라도 디지털 코드를 사용하는 것이다)

코드를 어떻게 설계해야 장점을 살리면서 정보를 덜 잃을까?

컴퓨터 : 저장장치. 바이너리만 처리한다. (컴퓨터가 전기를 이용하기에 이진법을 쓴다.)

문자, 숫자, 심볼을 바이너리로 변환가능하다.

문자, 숫자, 심볼은 광의적 디지털이라고 본다.

디지털은 갑작스러운 변화가 아니다.

무한 -> 유한 (이진 심볼, 2개의 심볼, (0과 1))

shannon

0과1 -> 전기전자적 장치를 사용하였음

언어를 생각해보자, 26개의 문자들의 조합으로 만들어진 한정된 개수의 단어들을 생각해보자  
 $26^5 = 11881376$ 이다.

정확성과 효율성 사이에서 우리는 어느정도 타협을 해야한다.

컴퓨터는 전기를 사용하기에 바이너리를 사용하게 된 것이다.

아날로그 —그림필름—>디지털(0,1 어떻게 이 정보를 저장, 전달, 처리하는가?)

샤논은 모든 세상의 정보를 '비트 단위'로 분해가 가능하고,

이 비트 정보가 정보의 최소 단위인 것이다.

아날로그 정보 -> 코드화되지 않은 정보

a, b, c, d, e -> 바이너리 코드(이것이 코드표이다.)

a|000

b|001

c|010

d|011

e|100 -> 2.xxx -> 3bits가 필요하다.

아날로그 — 약속된 코드표로 변환 —> 디지털 정보

주로 초등학교에서 이런 약속된 코드표들을 배우는 것이다.

데이터 인코딩

데이터를 1과 0으로 변환해주는 과정을 데이터 인코딩이라고 한다.

사실 인코딩에는 정답이 없다.

문자, 이미지, 소리, 숫자를 어떻게 디지털 코딩, 데이터 인코딩을 했을까?

정보 —변환—> 문자이미지소리 —> 이진 인코딩

언어는 기본적으로 기호들의 순서적 나열이다.

영어 문자를 보면 대문자 26, 소문자 26, 숫자 10개, 기타기호들 로 이루어져 있다.

문자 -> 단어 -> 문장

12개 약속들 이 필요하다.

그래서 처음으로 나온 것이 ASCII code이다.

8bits에서 하나는 check bit이고 나머지 7bits로 문자를 표현한다.

비 영미권 국가에서도 문자를 표현할 필요성이 필요했다.

16bits를 사용하는 유니코드가 있다.

그런데 영어를 쓸 때는 8개만 사용해도 되는데, 이는 비효율적이지 않은가?

정보 -> 넓은 디지털 -> 좁은 디지털(0과 1)

약속표로 인코딩, 디코딩을 하는 것이다.

한글의 약속표에 대하여 조합형과 완성형의 전쟁이 있었다.

지금 유니코드에는 완성형과 조합형을 둘다 사용할 수 있게 해 놓았다.