

Computational Thinking



2017-2학기

2장-실 세계 정보는 어떻게 계산 가능한 데이터가 되는가?

박종혁 교수 (컴퓨터공학과)

jhpark1@seoultech.ac.kr

<http://www.parkjonghyuk.net>

➤ 2장 실세계 정보는 어떻게 계산 가능한 데이터가 되는가?

1. 정보와 데이터
2. 정보를 데이터로 전환하기
3. 데이터 용량
4. 데이터 유형과 데이터 부호화
5. 데이터 압축
6. 참고문헌

➤ 조별과제

- 다음 주 강의 시간 발표

➤ 조별발표

- 정보와 데이터라는 용어를 정의한다.
- 컴퓨팅 시스템 내에서 데이터가 어떻게 비트열로 부호화되는지 설명한다.
- 데이터 용량에 대한 측정과 어떤 유형의 실세계 정보를 저장하기 위해서 얼마나 많은 용량이 필요한지 정의한다.
- 위치기수법을 설명한다.
- 정수와 실수가 어떻게 부호화될 수 있는지 설명한다.
- 컴퓨터가 부정확해질 수 있다는 것을 보여준다.
- 텍스트, 색상, 그림, 소리와 같은 복잡한 정보가 어떻게 비트열로 부호화될 수 있는지를 설명한다.

❖ 정보 (information)

- 어떤 특정한 사실, 주제 또는 사건에 대해 전달되는 지식
- 알려지거나 회자되는 지식
- 지능, 뉴스
- 데이터를 처리해서 얻어지는 것

❖ 데이터 (data)

- 컴퓨터와 기타 자동장치에 의해서 연산이 수행되는 숫자, 문자 또는 심볼
- 전기적으로 저장되거나 전송될 수 있고 자기 테이프나 천공카드에 등에 기록되어 있는 것

연속적인 (continuous) 데이터

- 개별적인 자료가 무한개의 가능한 값을 가질 수 있다면 데이터는 연속적 물리적인 또는 실세계를 포함하는 측정값

이산적인 (discrete) 데이터

- 유한개의 가능한 값
- 셀 수 있는 것과 관련

아날로그 (analog)

- 연속적인 데이터를 부호화
- 초기 전자공학에서 사용

디지털 (digital)

- 이산적인 데이터를 부호화
- 현대적인 컴퓨팅 시스템에서 사용

❖ 이진수 (binary digit)/비트 (bit)

- 디지털 시스템에서 가장 작은 데이터 단위
- ON(1)인지 OFF(0)인지 표시

❖ 비트열 (bit string)

- 비트들의 긴 수열

비트열의 길이	패턴의 개수
1	$2^1 = 2$
2	$2^2 = 4$
3	$2^3 = 8$
4	$2^4 = 16$
5	$2^5 = 32$
8	$2^6 = 256$
N	2^N

그림 2.4 비트열이 생성하는 패턴의 수

❖ 데이터 용량 (data capacity)

- 시스템이 부호화할 수 있는 정보의 양
- 비트수

❖ 바이트 (byte)

- 8자리 비트열
- 256개의 패턴

정보 유형	값의 개수	비트수
동전 던지기	2	1
요일	7	3
월	12	4
일	31	5
키보드 기호	~ 104	7
1년 날짜	365	9

그림 2.5 다양한 유형의 정보를 저장하기 위해서 필요한 대략적인 비트수

❖ 워드 (word)

- 데이터 용량을 위한 또 하나의 측정 단위
- 하드웨어에 기반을 둔 데이터 용량 단위
- 처리기에 의해서 단일 항목으로 처리
- 하나의 워드를 구성하는 비트수는 8의배수
- 유형별 정보를 저장하기 위해서 필요한 메모리의 양이 다름

단위	기호	2의 거듭제곱	십진수
킬로(Kilo)	K	2^{10}	$\sim 10^3$
메가(Mega)	M	2^{20}	$\sim 10^6$
기가(Giga)	G	2^{30}	$\sim 10^9$
테라(Tera)	T	2^{40}	$\sim 10^{12}$
페타(Peta)	P	2^{50}	$\sim 10^{15}$

그림 2.6 데이터 용량 단위

2.4 데이터의 유형과 부호화 – 2.4.1 숫자들

2.4.1.1 기수법

기수법 (numeral system) : 기록된 형태로 수를 표현하는 방법

- ❖ 계수표시 (tally markings)
 - 정수는 수의 각 단위에 대하여 하나의 계수를 표시함으로써 표현
- ❖ 이진기수법 (binary numeral system)
 - 밑을 2로 하는 기법
 - 각 숫자의 위치는 2의 멍승을 표시하고 값은 배수를 나타냄
- ❖ 위치기수법 (positional numerical system)
 - 밑(base) / 기수(radix) : 특정 기수법에 얼마나 많은 숫자가 존재하는지 명시

기수법	밑	숫자
이진(Binary)	2	0, 1
팔진(Octal)	8	0, 1, 2, 3, 4, 5, 6, 7
십진(Decimal)	10	0, 1, 2, 3, 4, 5, 6, 7, 8, 9
십육진(Hexadecimal)	16	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F

그림 2.10 일반적인 위치기수법

- ❖ 예) 3개의 숫자들



- 로마기수법: 1, 2, 3 10진기수법: 1, 11, 111 2진기수법: 1, 3, 7

- ❖ 숫자 열에서 각각의 숫자의 위치는 밑의 거듭제곱과 연관됨
- ❖ 각각의 숫자는 밑에 대해 승수가 됨
- ❖ 첨자표기를 사용하여 밑을 표현함
- ❖ 101_{10} 에서 첨자₁₀는 십진수로 표현되어 있음을 나타냄

정수	→	9	2	5	밑	→	10
위치	→	2	1	0			
밑의 거듭제곱	→	10^2	10^1	10^0			
의미	→	$9 \times 10^2 + 2 \times 10^1 + 5 \times 10^0$					

그림 2.11 십진기수법을 사용한 예

❖ 인간의 발명품 중에서 컴퓨터보다 더 많이 세상을 변화시킨 것이 있을까?

정수	→	1	0	1	밑	→	10
위치	→	2	1	0			
밑의 거듭제곱	→	10^2	10^1	10^0			
의미	→	$1 \times 10^2 + 0 \times 10^1 + 1 \times 10^0$					

그림 2.13 101_{10} 의 해석

정수	→	1	0	1	밑	→	2
위치	→	2	1	0			
밑의 거듭제곱	→	2^2	2^1	2^0			
의미	→	$1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0$					

그림 2.14 5_{10} 와 일치하는 101_2 의 해석

❖ 길이 N인 임의의 비트열

- 0부터 2^N-1 까지의 수로 부호화 가능

비트열	숫자값	비트열	숫자값
00000000	0	00000100	4
00000001	1	...	...
00000010	2	11111110	254
00000011	3	11111111	255

그림 2.15 이진 비트열의 이해

❖ 소수 위치를 이용하면 실수를 이진 비트열로 표현 가능

$$\begin{aligned} 1.101_2 &= 1 \times 2^0 + 1 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3} \\ &= 1 \times 1 + 1 \times \frac{1}{2} + 0 \times \frac{1}{4} + 1 \times \frac{1}{8} \\ &= 1 + \frac{1}{2} + \frac{1}{8} \\ &= 1.625 \end{aligned}$$

- ❖ 완벽하게 정확한 표현을 위해 필요한 만큼 많은 숫자를 쓰는 것이 불가능
 - 근사값 사용
 - 반올림 오류 발생

- ❖ 정밀도 (precision)
 - 저장된 숫자의 정확성
 - 사용할 수 있는 비트들의 수로 측정

언더플로우 (underflow)

- 사용가능한 비트수로 부호화 하기에 너무 작은 값 일때
Ex) 거의 0에 가까운 값을 가질 때

오버플로우 (overflow)

- 사용가능한 비트수로 부호화 하기에 너무 큰 값을 가질 때
- 8비트 이진 비트열은 단지 0부터 255까지의 값만 가질 수 있음
Ex) $255 + 1 = ?? \rightarrow$ 오류발생, 0

2.4.2 텍스트 문자

- ❖ 텍스트 문자는 부호화 방법을 사용하여 정수로 부호화
- ❖ 정수와 텍스트 문자 사이의 연관 → ASCII (American Standard Code for Information Interchange)
- ❖ ASCII표의 일부 텍스트 문자는 실제로 그림으로 나타나지 않고 하나의 명령으로 사용됨

Ex) 백스페이스

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
													!	"	#	\$	%	&	'
20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
(	)	*	+	,	-	.	/	0	1	2	3	4	5	6	7	8	9	:	;
40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59
<	=	>	?	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_	`	a	b	c
80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99
d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w
100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119
x	y	z	{		}	~													
120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139

그림 2.17 정수와 텍스트 문자사이의 연관을 나타내는 ASCII 표

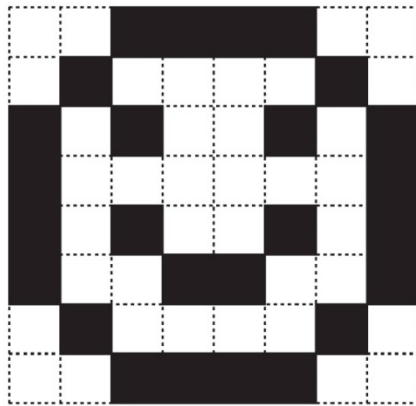
RGB 색상 모델

- 빨강색(R), 녹색(G), 파랑색(B)을 주요 색상으로 사용
- R,G,B 각각의 양을 다르게 배합하여 여러가지 색깔을 만들어 냄
- 색상(color)은 보통 0에서 255사이의 값을 가지는 정수 3개로 부호화
- R,G,B 각각 길이 8의 비트열로 표현
- 각 색상은 길이 24의 비트열(8비트*3가지 색상)로 부호화

Color	Name	Bits			Decimal
		red	green	blue	
	red	11111111	00000000	00000000	(255,0,0)
	green	00000000	11111111	00000000	(0,255,0)
	yellow	11111111	11111111	00000000	(255,255,0)

그림 2.18 이진 비트열로 색상을 부호화하는 일반적인 방법

- ❖ 픽셀 (pixel)
 - 색상을 2차원 격자로 부호화 한것
- ❖ 이미지는 일련의 픽셀들이고 각 픽셀들은 단일 색상임
- ❖ 색상은 보통 24 비트열로 부호화
- ❖ 따라서 디지털 영상을 부호화하기 위해서 필요한 전체 비트수는 영상에 포함되는 픽셀들의 수의 24배



0	0	1	1	1	1	0	0
0	1	0	0	0	0	1	0
1	0	1	0	0	1	0	1
1	0	0	0	0	0	0	1
1	0	1	0	0	1	0	1
1	0	0	1	1	0	0	1
0	1	0	0	0	0	1	0
0	0	1	1	1	1	0	0

그림 2.19 흑백 영상은 오른쪽에 나타난 비트들로 부호화될 수 있다.

❖ 소리 (sound)

- 마이크로폰 사용 - 소리파형을 아날로그 전기 신호로 변환
- 샘플링 - 변화하는 신호의 세기를 일정한 간격으로 측정해서 기록하는 과정
- 음파는 일련의 숫자값으로 변환

❖ 주파수 (frequency) – 음파가 변화하는 비율

- 단위 : 헤르츠 (Hz)
- 초당 바뀌는 사이클의 개수

❖ 나이퀴스트-새넌 샘플링 이론 (Nyquist-Shannon sampling theorem)

- 샘플링 비율은 측정된 신호의 가장 높은 주파수보다 최소 두 배가 되어야 한다는 법칙

❖ 데이터 압축 (Data Compression)

- 더 적은 비트수를 사용하여 실세계 정보를 부호화 하기
- 불필요한 비트를 부호화 기법에서 제거

❖ 데이터 압축시에 비용 발생

❖ 컴퓨터가 압축을 해제할 때 비용이 발생

❖ 실행 길이 부호화 (Run-length encoding)

- 영상과 텍스트에서 사용할 수 있는 가장 단순한 형태의 압축기술
- 행에서의 각 실행의 길이를 기록하여 적은 비트 수를 사용

예) 각 픽셀이 0 (검정) / 1 (흰색)인 단일 비트로서 부호화됨

2.5.1 실행 길이 부호화

예) 하트의 실행길이 부호화
- 18개의 열과 14개의 행으로 구성

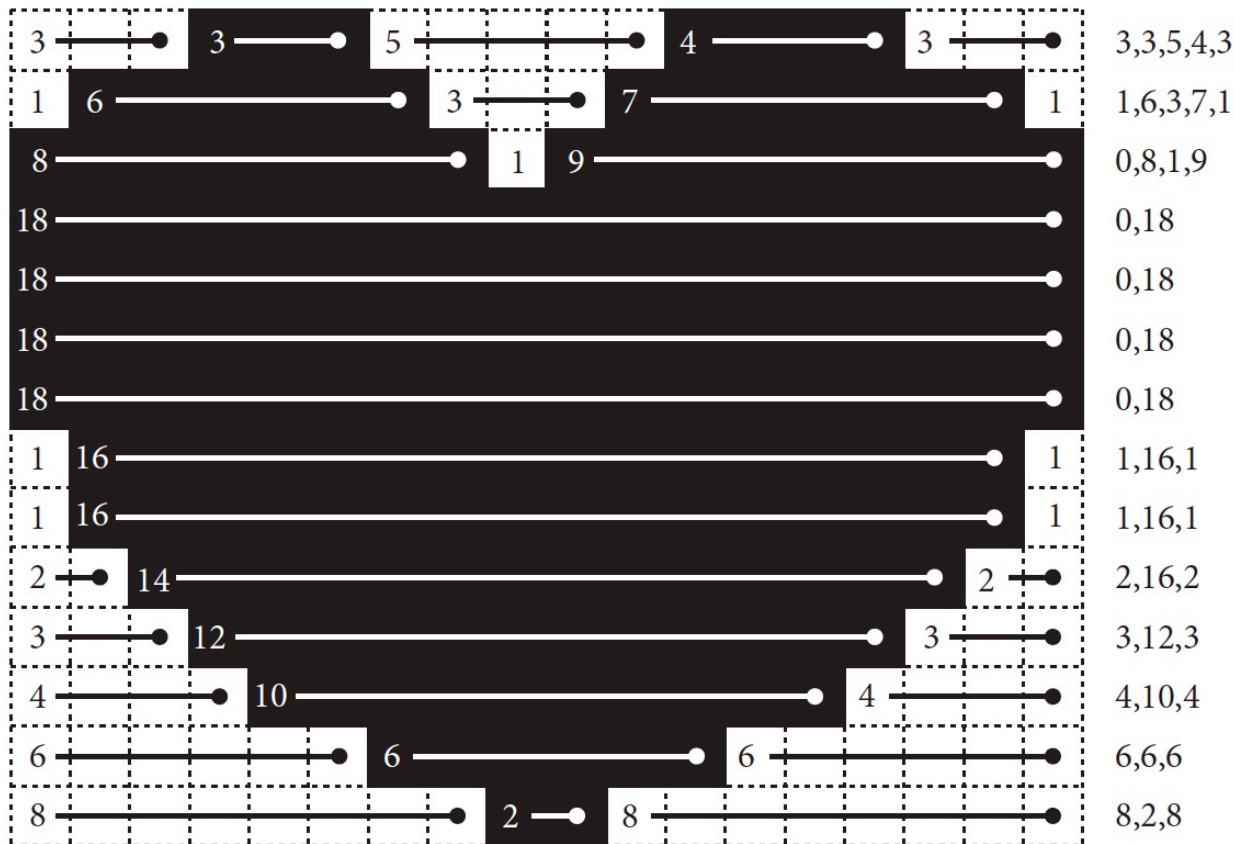


그림 2.21 흑백 영상은 오른쪽에 나타난 숫자들로 부호화된다.

손실 있는(lossy) 압축

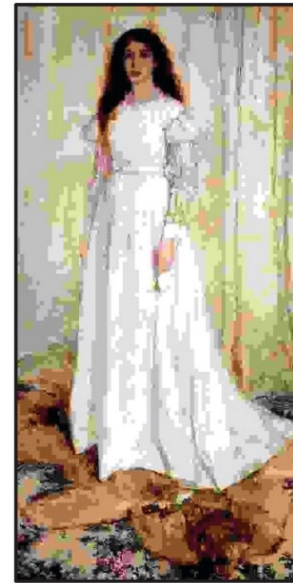
- 근사적인 표현을 부호화
- 별로 중요하지 않은 정보들을 폐기

무손실(lossless) 압축

- 원본 데이터의 정확한 표현을 부호화



압축되지 않은 영상
8,224,416 비트



JPEG(손실있는):
55384 비트



PNG(무손실):
5,668,304 비트

그림 2.22 손실있는 영상 압축과 무손실 영상압축

2.6 참고문헌

- 생각하는 프로그래밍, 윤성준,조상민 역, 인사이트, 2014
- 벤츠 타는 프로그래머, 정금호 저, 제이펍, 2013
- 컴퓨터과학이 여는 세계, 이광근 저, 인사이트, 2015
- 소프트웨어 전쟁, 백일승 저, 더하기북스, 2015
- 김대수, "소프트웨어와 컴퓨팅 사고", 생능출판, 2017
- 천인국, "어서와 파이썬은 처음이지!", 인피니티북스, 2016

- 발표방식: ppt 자료 활용 2개 조 각각 10분 발표 (질의응답 포함)
- 문제 1
 - 모든 컴퓨터는 0과 1 두 가지의 숫자를 가지고 있는 이진수를 기반으로 동작한다. 다양한 기수법들이 존재 하는데 왜 컴퓨터는 2진 비트를 사용하게 되고, 이러한 이진수를 컴퓨터가 어떻게 구현하고 있는지 조사하여 발표하라.
- 문제 2
 - 우리가 정보와 데이터라는 용어를 자주 혼용하고 있고 명확히 구분하지 못하는 경우가 많다. 이러한 정보와 데이터는 사실 정보의 진화단계에 포함되어 있다. 정보의 진화단계는 지혜, 지식, 정보, 데이터로 이루어져 있는데 각 단계의 차이점이 무엇인지 조사하라.