

1 자료와 정보의 표현

들어가기 전, Self Test

- ① 단순한 관찰이나 측정을 통해 얻어지는 값이나 사실을 ()라고 한다.
- ② 컴퓨터에서 처리되는 자료 표현의 최소 단위를 ()라고 한다.

생각 깨우기

영화 '정글북'은 주인공 모글리를 제외하고, 우거진 숲과 맑은 강물이 흐르는 아름다운 정글은 실제 정글 사진을 촬영하여 *컴퓨터 그래픽으로 작업하였다. 또, 동물들의 실제 표정, 질감, 움직임, 빛에 대한 반응 모습 등은 *모션 캡처를 이용해 실제 동물 영상으로 구현하였다.

모션 캡처를 이용해 동물 캐릭터의 움직임을 가상의 데이터로 만들어 실사 촬영과 정글 그래픽의 상호 작용을 이루었다.



실제 정글 사진을 10만 장 이상 찍어 라이브러리를 구축하고 다면 촬영 기법을 이용해 공간의 깊이감을 살렸다.

* 컴퓨터 그래픽(CG, Computer Graphics): 컴퓨터를 이용해 실제 세계의 영상을 조작하거나 새로운 영상을 만들어내는 기술이다.

* 모션 캡처(motion capture): 사람, 동물 또는 기계 등의 사물에 센서를 달아 그 대상의 움직임 정보를 인식해 영상 속에 재현하는 기술을 말한다.

Q. 컴퓨터는 다양한 유형의 정보를 어떻게 저장할까?



1 디지털 표현

성취 기준

- 디지털 정보의 개념과 특징을 이해할 수 있다.
- 문자 정보와 수치 정보의 디지털 표현 원리를 이해할 수 있다.

▲ 학습 요소 아날로그, 디지털, 비트, 디지털 정보, 문자 정보, 수치 정보



1 자료와 정보의 표현

아날로그는 연속적으로 변화하는 값을 표현하고, 디지털은 연속적으로 변화하는 값을 일정한 간격으로 끊어 불연속적인 값으로 표현한다.

컴퓨터는 생활 속의 문자, 그림, 소리 등의 다양한 형태의 아날로그 자료를 이진(0과 1) 값으로 변환하여 디지털 정보를 생성한다. 이때 0(꺼짐)과 1(켜짐), 각각을 자료 표현의 최소 단위인 비트(bit)라고 부르며, 이러한 비트가 나열된 값을 디지털 정보라 한다. 1비트는 두 가지($=2^1$) 정보를 표현할 수 있으며, 2비트는 네 가지($=2^2$), 3비트는 여덟 가지($=2^3$) 정보를 표현할 수 있다.

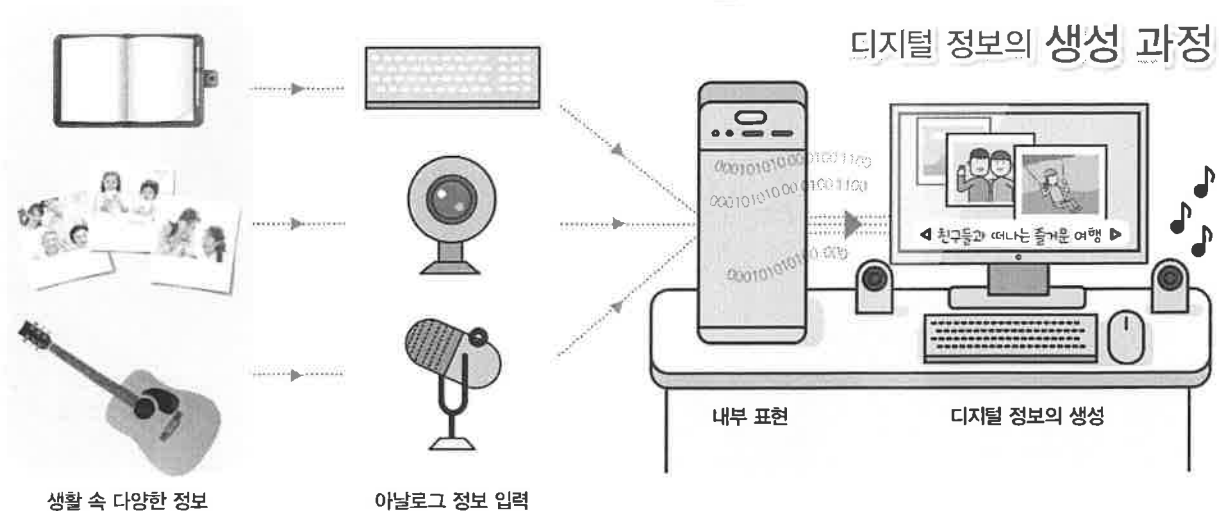
컴퓨터에서 비트 8개를 모아 놓은 것을 바이트(byte)라고 하고 영문자, 숫자, 기호 등 $2^8(=256)$ 개의 정보를 표현할 수 있다.

디지털 정보는 기존의 아날로그 정보보다 정보의 편집·가공·저장이 쉬우며, 다양한 디지털 정보 기기와 호환되어 정보를 처리할 수 있다.

비트와 바이트의 정보 표현 개수

1bit	2bit	3bit
0	00	000
1	01	001
	10	010
	11	011
		100
		101
		110
		111
2가지	4가지	8가지

• 1바이트 = 8비트 = $2^8(=256)$ 개



② 같은 정보, 서로 다른 표현

1. 문자 정보

문자 정보는 컴퓨터 키보드에서 직접 입력되는 한글, 영문자, 숫자, 기호 등으로 표현된 정보를 말한다. 컴퓨터에서 문자 정보를 표현하려면 컴퓨터가 이해할 수 있는 이진수 형태로 변환해야 한다. 현재 대표적으로 많이 사용하는 문자 코드 체계는 아스키코드, 유니코드 등이 있다.



※ 아스키코드와 유니코드(UTF-16BE) 비교

진수 표현

수의 진법은 밑(base)으로 표현하며, 이진수의 밑은 2이고, 십진수의 밑은 10이다. 일반적으로 십진수의 밑은 생략한다.

예 111₍₂₎
→ 밑(base)

이진수

0과 1로 나타낸 수이다.

십육진수

0에서 9까지의 10개 숫자와 A, B, C, D, E, F를 사용하여 나타낸 수이다.

0001 1110₍₂₎ = 1E₍₁₆₎

10진수	2진수	16진수
0	0	0
1	1	1
2	10	2
3	11	3
4	100	4
5	101	5
6	110	6
7	111	7
8	1000	8
9	1001	9
10	1010	A
11	1011	B
12	1100	C
13	1101	D
14	1110	E
15	1111	F
16	10000	10

2. 수치 정보

수치 정보는 컴퓨터에서 직접 연산이 가능한 숫자 정보로 정보의 양이나 크기를 표현한다. 컴퓨터는 사람이 사용하는 십진수를 0과 1의 이진수로 변환하여 저장 및 연산하고, 처리 결과는 다시 십진수로 변환하여 출력한다.

▶ 십진수를 이진수로 변환하기

$$\begin{array}{r}
 2 \overline{) 12} \\
 2 \overline{) 6} \quad \cdot 0 \\
 2 \overline{) 3} \quad \cdot 0 \\
 1 \quad \cdot 1
 \end{array}
 \quad
 \begin{array}{l}
 \uparrow \\
 12_{(10)} = 1100_{(2)}
 \end{array}$$

▶ 이진수를 십진수로 변환하기

$$\begin{aligned}
 &1100_{(2)} \\
 &= 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 0 \times 2^0 \\
 &= 12_{(10)}
 \end{aligned}$$

※ 십진수와 이진수의 변환 방법

컴퓨터는 0과 1의 나열인 이진수 표현의 이해를 돕기 위해 십육진수를 사용한다. 이진수 오른쪽 끝자리 숫자부터 시작하여 4자리씩 묶어 십육진수로 변환할 수 있다.

▶ R(Red, 빨강), G(Green, 녹색), B(Blue, 파랑) 3가지 색을 섞어 한 가지 색상으로 표현한다.



R(빨강), G(녹색), B(파랑)는 각각 1바이트씩 할당되어 다음과 같이 적색을 표현한다.



생활 속의 정보는 인코딩 기법에 의해 디지털 정보로 변환된다. 문자 인코딩에서 사용하는 문자 코드와 체계에 따라 같은 문자가 다른 디지털 값으로 표현된다. 인코딩된 디지털 값을 원래의 정보로 변환하기 위해서는 디코딩하는 과정이 필요하다.

1. 아스키코드(ASCII Code: American Standard Code for Information Interchange)

아스키코드는 영문 대·소문자, 숫자, 특수 기호 등 $2^7(=128)$ 개의 서로 다른 문자를 표현하기 위해 7비트를 사용한다. 7비트 아스키코드는 8번째 비트 영역(128~255)에 패리티 비트나 새로운 유럽 문자 등을 포함하여 단일 문자를 1바이트 형식으로 인코딩한다.

2진수	000	001	010	011	100	101	110	111	← 상위 비트
0000	NUL	DLE	Space	0	@	P	'	p	
0001	SOH	DC1	!	1	A	Q	a	q	
0010	STX	DC2	"	2	B	R	b	r	
0011	ETX	DC3	#	3	C	S	c	s	
0100	EOT	DC4	\$	4	D	T	d	t	
0101	ENQ	NAK	%	5	E	U	e	u	
0110	ACK	SYN	&	6	F	V	f	v	
0111	BEL	ETB	'	7	G	W	g	w	
1000	BS	CAN	(8	H	X	h	x	
1001	HT	EM)	9	I	Y	i	y	
1010	LF	SUB	*	:	J	Z	j	z	
1011	VT	ESC	+	;	K	[k	{	
1100	FF	FC	,	<	L	¥	l		
1101	CR	GS	-	=	M]	m	}	
1110	SO	RS	.	>	N	^	n	~	
1111	SI	US	/	?	O	_	o	DEL	

A = 100 0001₍₂₎
상위 비트 하위 비트

↑ 하위 비트

▲ 7비트 아스키코드 표

2. 유니코드(Unicode)

유니코드는 전 세계의 모든 문자를 동일한 자릿수로 표현하기 위한 표준 코드로, 최초의 유니코드 인코딩은 고정 2바이트를 사용하였다. 이후 2바이트를 넘어서는 문자에 대해서는 유니코드를 4바이트로 늘리지 않고, 문자에 따라 다른 길 이로 인코딩하는 기법들이 등장하고 있다.

문자		아스키코드	유니코드	
			이진수로 표현	십육진수로 표현
숫자	1	0110001	0000000000110001	0031
영문	A	1000001	000000001000001	0041
한글	가	표현할 수 없다.	1010110000000000	AC00
아랍어	ا	표현할 수 없다.	0000011000100110	0626
한자	上	표현할 수 없다.	0100111000001010	4E0A
히브리어	ד	표현할 수 없다.	0000010111100100	05E4



선택 활동 1



정답은?

컴퓨터 사고력

문제 해결 능력


비트 이해하기

다음 정보를 이진(0과 1) 값으로 표현하기 위해 필요한 비트 수를 알아보자.

- 1 주사위 눈 1에서 6을 표현하는 데 필요한 비트 수를 적어 보자.



- 2 윗가락의 평평한 면을 1, 둥근 면을 0이라고 가정할 때 4개의 윗가락을 던져 일어날 수 있는 모든 경우의 수를 생각해 보고, 이를 표현하기 위한 비트 수를 적어 보자.



평가하기

- 1 디지털 자료의 단위를 설명할 수 있는가?
- 2 컴퓨터로 정보를 표현하는 원리를 설명할 수 있는가?

잘	중	하
상	중	하

모듬 선택 활동 2

이 활동의 핵심 역량은?

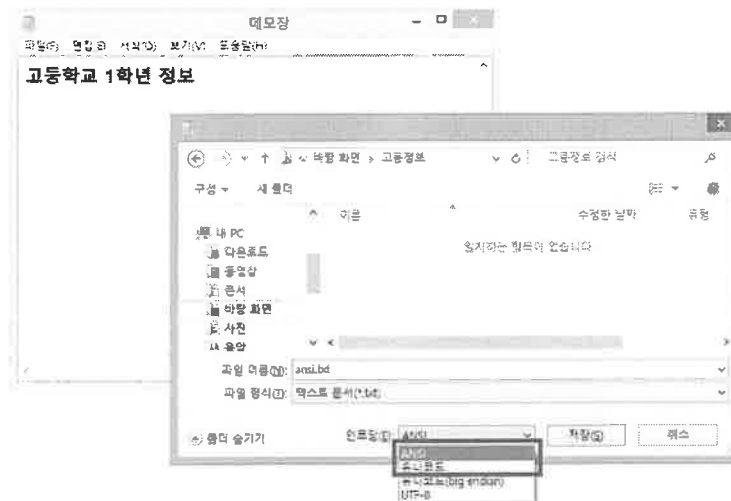
정보 문화 소양

컴퓨팅 사고력

창의적 문제 해결력

문자 인코딩하기

문서 작성 프로그램으로 텍스트 파일을 만들고, 만들어진 디지털 자료의 용량을 확인해 보자.



1 ANSI, 유니코드 인코딩으로 각각 저장해 보고, 파일의 용량을 비교해 보자.

파일명	용량
ANSI.txt	
유니코드.txt	

2 ANSI, 유니코드 인코딩에 대해 조사해 보자.

인코딩	설명
ANSI	
유니코드	

평가하기

- 문자 인코딩 과정을 이해할 수 있는가?
- 문자 인코딩 기법에 따라 같은 문자열이 다르게 표현되는 원리를 설명할 수 있는가?

상	중	하
상	중	하

2 효율적인 디지털 표현

정체 기술

- 문자열, 그림, 소리 등의 정보를 효율적으로 관리하기 위한 여러 가지 디지털 표현 방법을 이해할 수 있다.
- 자료와 정보의 **활용** 목적에 따라 효율적인 디지털 표현 방법을 선택하고, 비교·분석할 수 있다.

▲ 학습 요소 효율적인 디지털 표현, 문자열 정보, 그림 정보, 소리 정보



1 효율적인 디지털 표현 방법

생활 속의 정보를 디지털 정보로 표현할 때에는 각 정보의 특징을 살려 최대한 비슷하게 표현하는 것이 중요하다. 하지만 한정된 물리적 자원을 사용하는 컴퓨터에서 디지털 정보를 효율적으로 표현하기는 쉽지 않다.

따라서 표현하고자 하는 정보의 특성과 주어진 상황, 목적 등을 제대로 파악하여 디지털화하는 과정이 필요하다. 또한 효율적으로 자원을 활용하기 위해 같은 정보를 더 적은 수의 비트를 사용하여 표현하는 것이 매우 중요하다.

효율적인 자원 활용

디지털 정보의 크기가 작을수록 기억 장치의 저장 공간과 장치 간 데이터 전송 시간이 줄어들어 효율적으로 자원을 활용할 수 있게 도와준다.



▲ 디지털 표현의 효율성

② 문자열 정보 표현

문자열은 문자들을 하나로 묶은 것으로, 각 문자는 문자 집합의 이진 코드로 표현된다. 이때, 문자열에 존재하는 반복되는 문자나 부분 문자열을 찾아 더 적은 수의 비트로 반복을 표현하면 문자열을 효율적으로 표현할 수 있다.

이와 같은 과정을 데이터 압축이라고 하며, 가장 기본적인 문자 압축 알고리즘인 런-길이(Run-Length), 렘펠-지브(Lempel-Ziv) 압축 기법에 대해 알아보자.

1. 런-길이 압축

런-길이 압축은 동일한 문자가 연속해서 반복되는 경우, 그것을 문자와 반복된 문자 개수의 쌍으로 표현하는 알고리즘이다.

원래 문자열 BBBBWWBBBB

압축 문자열 B4W2B3

런-길이 압축 효율

런-길이 압축 알고리즘은 반복되는 문자가 길거나 더 자주 나타날수록 압축 효율이 높아지며, 흑백으로만 표현된 그림이나 팩스 정보, 단색의 만화 등을 압축할 때 많이 사용된다.

2. 렘펠-지브 압축

렘펠-지브 압축은 가까이 있는 동일한 패턴의 상대적인 위치와 패턴의 길이로 문자열을 치환하여 표현하기 때문에 떨어져 있는 동일한 패턴에 대해서도 압축할 수 있다.

원래 문자열 ABCDABCE

압축 문자열 ABCD(4,3)E

상대적인 위치

패턴의 길이

현재 위치에서 4문자 앞에서부터 3문자

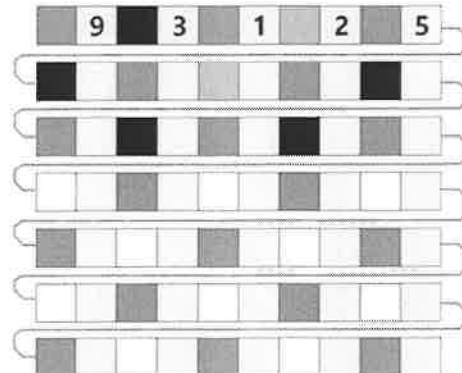
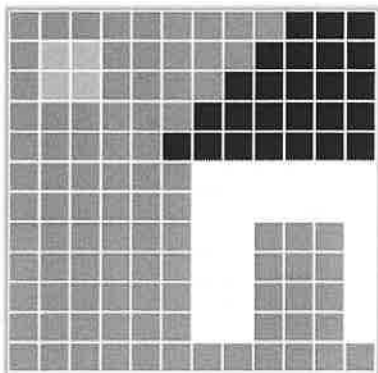
렘펠-지브 압축 효율

렘펠-지브 압축 알고리즘은 압축 효율이 높아서 대부분의 압축 형태에서 확장 및 변형하여 사용된다.

예제 생활 속 그림 압축하기

문제 다음 그림의 규칙을 찾아 오른쪽 빈칸에 숫자를 채워 런-길이 알고리즘으로 압축해 보자.

풀이



3 그림 정보 표현



* 화소

컴퓨터에서 그림을 구성하는 최소 단위로 그림이나 사진을 구성하는 낱알의 점들을 화소 또는 픽셀(pixel)이라고 한다.



화소(픽셀)

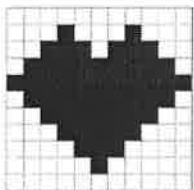

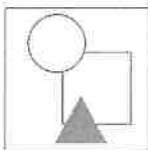

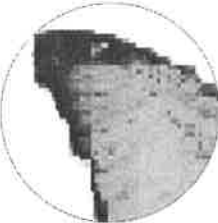
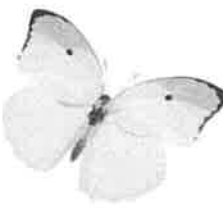

컴퓨터에서 그림 정보를 표현하는 방식은 비트맵 방식과 벡터 방식으로 나눌 수 있다.

비트맵 방식은 *화소들의 집합으로 그림을 표현하고, **벡터 방식**은 점·선·면 등의 수학적 수식으로 그림을 표현한다.

효율적으로 그림 정보를 표현하기 위해서는 그림 정보의 복잡한 정도, 색채 변화에 민감한 정도 등의 특징에 따라 그림 정보를 표현하는 방식을 적절히 선택해야 한다.



비트맵과 벡터 방식의 비교

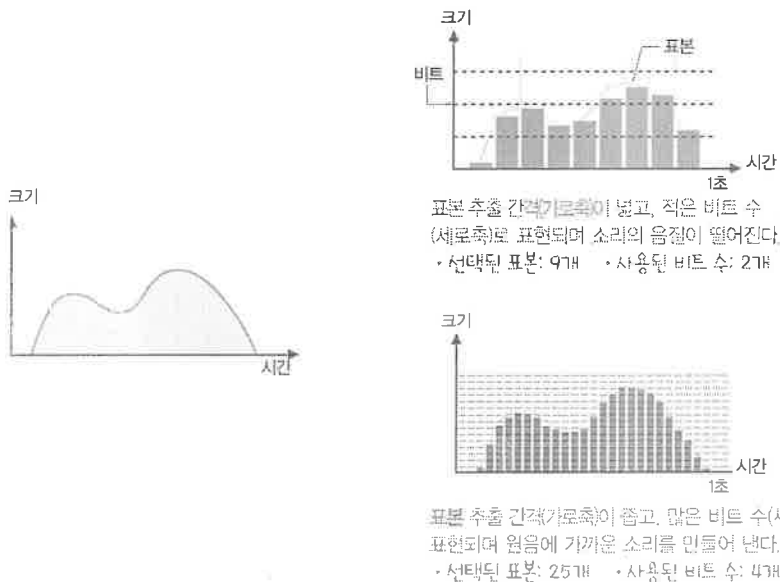
구분	비트맵(bitmap)	벡터(vector)
특징	화소들이 모여 하나의 그림을 표현	점의 좌표값, 선 길이, 선 굵기, 색상 등에 대한 수학적 정보로 그림을 표현
그림 표현	 비트맵 방식 그림  비트맵 그림 표현	 벡터 방식 그림 <p>원의 중심 x, y 좌표 Circle 20, 20, 12 → 반지름 크기 Triangle 40, 35, 22, 78, 32, 78 green Rectangle 25, 25, 65, 65</p>
적용 분야	다양한 색상이 쓰인 고해상도 사진, 수채화나 회화 작품 등	아이콘, 캐릭터, 로고 디자인 등
그림 비교	 * 계단 현상 그림을 확대하면 화소의 크기가 커져 테두리가 깨지는 현상이 나타난다. 	 그림을 확대해도 깨지는 현상이 나타나지 않는다. 
장점	섬세하고 자연스러운 색의 변화를 표현할 수 있다.	<ul style="list-style-type: none"> 그림 확대나 축소에 영향을 받지 않는다. 수학적 수식만 기억하기 때문에 그림이 단순할수록 파일의 크기가 작다.
단점	<ul style="list-style-type: none"> 해상도(그림 전체 화소 수)가 높을수록 저장 공간이 많이 필요하다. 그림을 확대할 때 계단 현상이 나타난다. 	<ul style="list-style-type: none"> 정교한 색상 표현이 어려워 비트맵에 비해 사실성이 떨어질 수 있다. 그림이 복잡할수록 파일의 크기가 커지고, 직접 처리 시간이 많이 소요된다.
확장자	.gif, .jpg, .bmp, .png	.ai, .eps, .wmf

4 소리 정보 표현

컴퓨터에서 소리 정보를 표현하기 위해서는 입력되는 아날로그 신호를 디지털 형태의 이진 코드로 변환해야 한다.

일반적으로 아날로그 신호에서 일정한 시간 단위로 나누어 표본을 추출하고, 추출한 신호의 크기를 이진 코드로 변환하여 디지털 신호를 만든다. 이때 표본을 추출하는 간격이 넓을수록 음질이 떨어지고, 추출 간격이 좁을수록 원음에 가까운 소리를 만들 수 있다. 또한 추출한 신호를 이진 코드로 변환할 때, 신호 저장에 사용된 비트 수가 많을수록 정밀하게 소리를 표현할 수 있다.

선택 사항



표본화와 표본 채취율

아날로그 형식의 자연 음을 디지털 신호로 추출하는 과정을 표본화라 하며, 1초 단위의 표본화 횟수를 표본 채취율(샘플링 레이트)이라 한다.

소리 확장자

파일 형식	확장자
WAV	.wav
MP3	.mp3
MID	.mid

동영상 확장자

파일 형식	확장자
MPEG	.mpeg, .mpg
AVI	.avi
MP4	.mp4

소리 정보의 디지털 표현



지식 충전소

동영상 정보

동영상 정보는 문자, 소리, 그림 등의 정보를 조합하여 사용하는 멀티미디어 정보로 파일 용량이 큰 특징을 갖는다. 프레임이라고 부르는 정지 영상을 연속적으로 빠르게 보여 주면 우리는 눈의 잔상 효과에 의해 프레임들이 움직이는 것처럼 인식한다. 1초에 15~30 프레임을 연속적으로 보여 주면 자연스러운 움직임을 느끼게 된다.

일반적으로 동영상 파일은 파일 용량이 커서 압축하여 저장하며, 음성과 영상의 압축 기법에 따른 다양한 *코덱들이 존재한다. 같은 파일 형식이라도 압축할 때 사용한 동일한 코덱을 사용해야 동영상 파일을 재생할 수 있다.

* 코덱: 음성이나 영상 자료를 컴퓨터가 처리할 수 있게 디지털로 바꾸어 주고, 그 데이터를 컴퓨터 사용자가 알 수 있게 모니터에 본래대로 재생시켜 주기도 하는 소프트웨어이다.





선택 활동 2



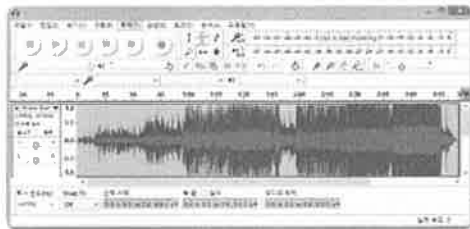
정보 문화
소양

컴퓨터
사고력

창의적 문제
해결력

소리 정보 디지털 변환하기

프리웨어로 제공하는 소리 편집 프로그램인 'Audacity'를 사용하여 디지털 음원을 만들어 보자.



Audacity 프로그램은 wav, mp3 등과 같은 소리 파일을 재생, 녹음, 편집할 수 있는 소프트웨어이다.



1 소리를 녹음하고 wav 파일과 mp3 파일로 저장하여 각각의 특징을 조사해 보자.

tip. 소리, 그림, 문자 파일의 확장자만 변경한다고 파일이 바뀌는 것은 아니다.

파일 실행 순서

<http://www.audacityteam.org> 사이트 접속 - Audacity 설치 - Audacity 실행 - [녹음 버튼] - [정지] - [File] 메뉴 - Export Audio - 파일 형식 선택

파일 형식	확장자	특징
WAV		
MP3		

2 다음 동영상 파일 형식의 확장자 및 특징을 조사해 보자.

동영상 파일 형식	확장자	특징
MPEG		
AVI		
MP4		
MOV		
WMV		

평가하기

- 1 소리 정보의 저장 방식을 설명할 수 있는가?
- 2 동영상 파일 형식과 확장자를 구분하고, 설명할 수 있는가?

☐ 상 ☐ 중 ☐ 하
☐ 상 ☐ 중 ☐ 하

중단원 마무리



핵심 내용 정리 노트

1. 디지털 표현

(핵심 개념) 비트, 디지털 정보, 문자 정보, 수치 정보

- 1) 컴퓨팅 기기는 아날로그 자료를 이진(0과 1) 값으로 변환하여 디지털 정보를 생성한다.
- 2) 0과 1, 각각을 자료 표현의 최소 단위인 비트라고 부른다.
- 3) 문자 인코딩에서 사용하는 문자 코드와 체계에 따라 동일한 문자 정보가 서로 다른 디지털 값으로 표현될 수 있다.
- 4) 컴퓨팅 기기에서 수치 정보는 인간이 사용하는 십진수를 0과 1의 이진수로 변환하여 처리된다.

2. 효율적인 디지털 표현

(핵심 개념) 문자열 정보, 그림 정보, 소리 정보

- 1) 정보의 특성이나 상황, 목적 등을 파악하여 효율적으로 디지털 정보를 활용할 수 있어야 한다.
- 2) 런-길이 기법과 램펠-지브 기법을 통해 문자열 정보를 압축할 수 있다.
- 3) 그림 정보는 화소들로 그림을 표현하는 비트맵 방식과 수학적 정보로 그림을 표현하는 벡터 방식으로 표현할 수 있다.
- 4) 추출되는 표본이 많을수록 소리를 표현하는 비트 수가 많을수록 원음에 가까운 정밀한 소리 표현이 가능하다.



이 단원의 성취 수준은?



1. 디지털 정보의 특성을 이해하고 컴퓨터의 자료 표현 원리를 설명할 수 있다.

- 상 디지털 정보의 특징을 잘 이해하고 컴퓨터의 자료 표현 원리를 설명할 수 있다.
- 중 디지털 정보의 특징과 컴퓨터의 자료 표현 원리를 이해할 수 있다.
- 하 디지털 정보의 특징을 이해할 수 있다.

→ 51쪽



2. 같은 정보가 다르게 표현되는 이유를 설명할 수 있다.

- 상 문자 정보와 수치 정보의 표현 방법과 어떻게 다르게 표현되는지 설명할 수 있다.
- 중 문자 정보와 수치 정보의 표현 방법을 설명할 수 있다.
- 하 문자 정보와 수치 정보의 개념을 설명할 수 있다.

→ 52쪽



3. 정보를 효율적으로 활용하기 위한 디지털 정보의 다양한 표현 방법을 설명할 수 있다.

- 상 정보를 활용 목적에 따라 문자열, 그림, 소리 등을 효율적으로 표현하는 방법을 알고, 보다 효율적인 방법을 선택하여 비교·분석할 수 있다.
- 중 정보를 효율적으로 표현하는 방법을 알고 효율적인 방법을 선택할 수 있다.
- 하 정보를 효율적으로 표현하는 방법을 알고 있다.

→ 56쪽