

## 4주차 1차시 텍스트의 표현과 코드구성의 이해

### 【학습목표】

1. 텍스트의 개념을 설명할 수 있고, 유형별 특징을 설명할 수 있다.
2. 코드의 각각의 종류를 파악하고 구분할 수 있다.

### 학습내용1 : 텍스트의 표현

#### 1. 텍스트의 개념

- 텍스트의 사전적 의미 : 문장보다 더 큰 문법 단위 즉 문장이 모여서 이루어진 한 덩어리의 글
- 멀티미디어에서의 텍스트란? 멀티미디어에서의 텍스트란 다수 문장으로 구성된 글이라고 할 수 있다.
- 텍스트 장점 : 조작이 가장 수월하고 키보드로 쉽게 입력할 수 있기 때문에 멀티미디어 데이터 중에 가장 많이 사용되고 있다.
- 멀티미디어 데이터 중에 우리가 가장 많이 사용하는 텍스트는 주로 숫자와 문자로 구성되며 멀티미디어 구성 요소 중에서 가장 기본적인 미디어이다.
- 다수의 미디어중 가장 전통적인 미디어임
- 미디어는 비디오나 사운드를 이용하지 않고도 정보를 표현할 수 있지만 텍스트를 제외하고는 정확한 정보를 전달하기가 어렵다.
- 다수의 미디어 중 텍스트는 가장 정확한 표현을 가지고 있음.
- 다른 미디어 유형과의 차이점 : 텍스트는 다른 미디어 유형처럼 시각적인 흥미를 제공하지는 않지만 정교한 정보를 전달할 경우에 매우 많이 사용된다.
- 사물을 설명할 때 가장 세밀한 표현을 할 수 있는 미디어가 텍스트임
- 텍스트 데이터 :
  - 텍스트 데이터는 가장 수월하게 조작할 수 있는 데이터로서 주로 키보드의 입력으로 생성되지만 스캐너 같은 장비로 입력된 글을 문자 인식 기술을 사용하여 텍스트 데이터를 생성할 수도 있다.
  - 텍스트 데이터는 한 권의 책 분량을 1MB정도로 저장할 수 있다. 즉, 이미지 한 장 정도의 분량의 상대적으로 적은 용량으로 많은 내용을 전달할 수 있는 미디어로서 텍스트 데이터는 가장효율적인 미디어라 할 수 있다.

## 2. 텍스트의 유형

- 멀티미디어에서 텍스트를 나타내는 방법에는 두 가지 방법이 있다. 첫 번째는 정적인 텍스트(Static Text) 이고, 두 번째는 동적인 텍스트(Dynamic Text) 이다.

### ① 정적인 텍스트

- 정적인 텍스트는 각 화면에 정적인 문자로 표현되며, 화면에서 가장 보기 좋은 형태로 디자인되어 보이게 된다.
- 정적인 텍스트는 디자인이 중요시되기 때문에 텍스트 정보는 항상 하나의 화면 형태로 한정된다는 단점이 있다.

### ② 동적인 텍스트

- 동적인 텍스트는 비선형적이고 비순차적인 정보 체계로 사용자가 원하는 정보를 탐색할 수 있다는 장점이 있다.

## 학습내용2 : 코드의 종류

### 1. 아스키 (ASCII) 코드

아스키코드란 1963년 미국표준협회 (ASA) 에 의해 결정되어 미국의 표준부호가 되었으며 PC와 같은 소형 컴퓨터를 중심으로 보급되어 현재 국제적으로 널리 사용되고 있다.

- 아스키는 128개의 가능한 문자조합을 제공하는 7비트 부호로, 처음 32개의 부호는 인쇄와 전송 제어용으로 사용된다. 보통 기억장치는 8비트 (256조합) 이고, 아스키는 단지 128개의 문자만 사용하기 때문에 나머지 비트는 패리티 비트나 특정 문자로 사용된다.

### 2. BCD (Binary Coded Decimal) 코드

- BCD 코드는 이진화 십진 표기법이라고도 하며 십진법에서의 각 십진 숫자를 이진 숫자로 나타내는 방식이다. BCD 코드에서 무게가 8-4-2-1 일 때 십진수 9530의 예는 다음과 같이 한 자리씩 계산하여 변환할 수 있다.
- 9는 8과 1 의 합으로 나타낼 수 있기 때문에 1001
- 5는 4와 1 의 합으로 나타낼 수 있기 때문에 0101
- 3은 2와 1 의 합으로 나타낼 수 있기 때문에 0011
- 0은 그대로 모든 비트를 0으로 하여 0000
- 따라서 십진수 9530의 BCD 코드는 1001 0101 0011 0000으로 변환된다.

컴퓨터 내부에서는 보통 이진 표기법이 사용되지만 BCD가 사용할 때도 많은데, 그 이유는BCD를 사용하는 컴퓨터에서는 십진수가 그대로 계산될 수 있다는 장점이 있기 때문이다.

### 3. EBCDIC(Extended Binary Coded Decimal for Interchange Code)

EBCDIC는 확장 이진화 십진 코드라고도 하며 8비트를 써서 2567B 의 문자를 표시할 수 있게 한 코드이다. 주로 IBM 메인 프레임 컴퓨터에서 사용하는데 이진화 십진 코드에 4비트를 더한 것으로 숫자 영문 대문자 소문자 특수 문자 한글의 자음과 모음 등을 모두 표현할 수 있다. 이 코드는 초창기에는 주도적인 내부 부호로 사용되었지만 근래 에는 아스키코드가 압도적인 위치를 차지함에 따라 잘 사용되지 않는다.

#### 4. 한글코드

한글을 나타내려면 위에서 살펴본 코드와 다른 16비트로 이루어진 별도의 코드가 필요하다. 이와 같은 한글을 나타내기 위한 코드에는 완성형 한글코드와 조합형 한글 코드 두 가지 방식이 있다.

- 완성형 한글코드는 1987년 우리나라의 국가표준코드로 지정한 2바이트 코드 KSC 5601이다.
- 조합형 한글코드는 한글의 기본 구성 자소인 초성 19자 중성 21자와 종성 27자를 서로 결합하지 않은 상태로 컴퓨터 에 저장하여 화면이나 프린터 에 나타나도록 하는 방식의 조합형 한글 글자꼴의 처리에 사용하는 코드체계이다.
- 유니코드 : 국제표준으로 제정된 2바이트의 국제 문자부호 체계로서 데이터의 교환을 원활하게 하려고 문자 1 개에 부여되는 값을 16 비트로 통일하였다.

##### \* 특징

- 코드의 1 문자 당 영어는 7비트, 비영어는 8비트, 한글이나 일본어는 16비트의 값을 지니는데, 이들 모두 16비트로 통일함.
- 최대 수용할 수 있는 문자수는 65,536자인데, 이 중 38,885자는 주요 국가의 언어를 구현하는 용도로 이미 할당되어 있고 6,400자는 사용자 정의 영역으로 나머지 20,000 여 자는 새로 추가될 언어 영역으로 각각 비워두고 있다.
- 유니코드에서 코드 할당비율을 보면 한자가 39 . 89%로 가장 많고, 한글 17.04%, 아스키와 기호문자 10.39% 등의 순이다.

#### 【학습정리】

##### 1. 텍스트의 사전적 의미

- 문장보다 더 큰 문법 단위 즉 문장이 모여서 이루어진 한 덩어리의 글

##### 2. 멀티미디어에서의 텍스트란

- 다수 문장으로 구성된 글이라고 할 수 있음

##### 3. 유니코드란?

- 국제표준으로 제정된 2바이트의 국제 문자부호 체계로서 데이터의 교환을 원활하게 하려고 문자 1 개에 부여되는 값을 16 비트로 통일하였음