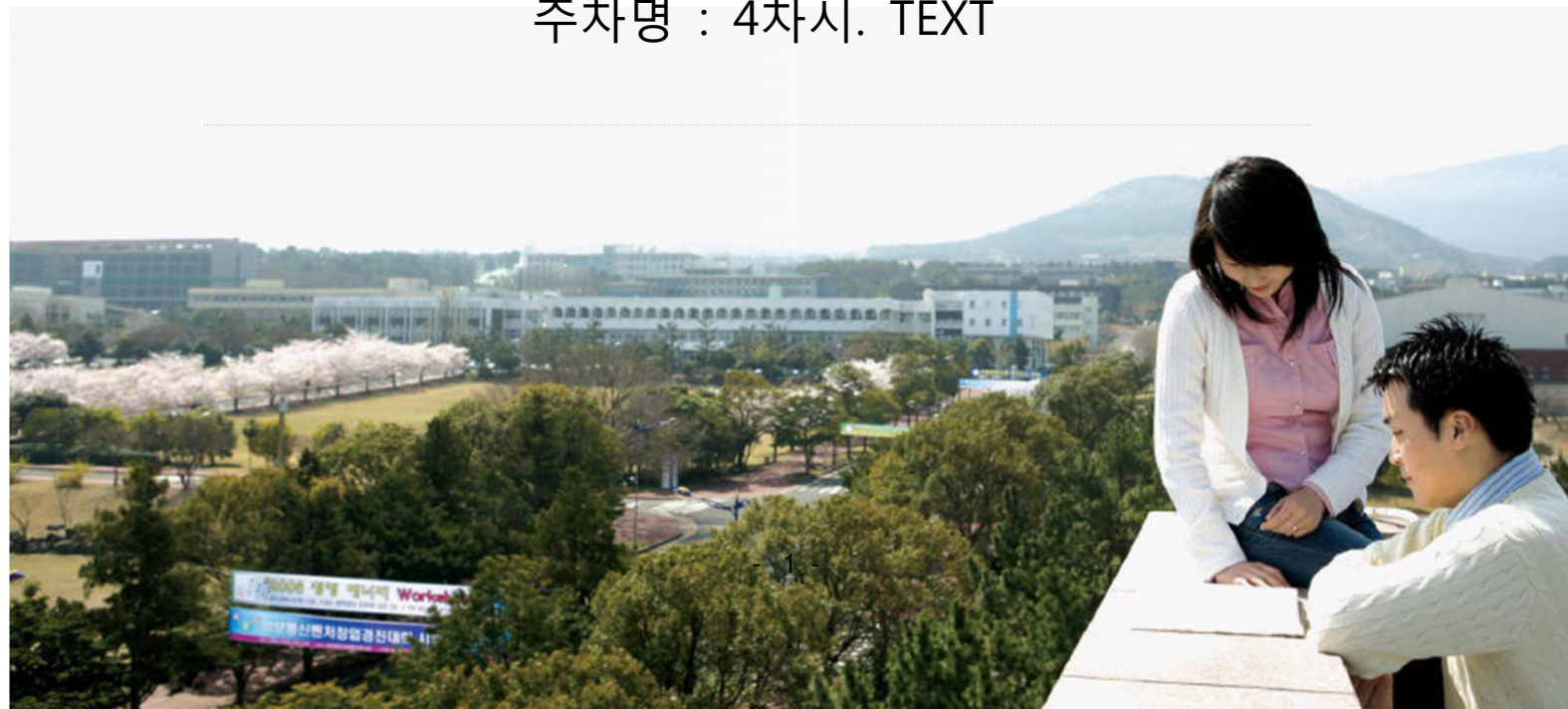


[교안]

디지털콘텐츠의 이해

주차명 : 4차시. TEXT



1. 텍스트의 기초

1) TEXT

- 언어 표현의 첫 요소이며, 의사전달의 수단으로서 문자를 사용하기 시작하였다.
- 컴퓨터의 발달로 키보드상에서 입력하고 사용이 가능하게 되었으며, ISO 표준으로 정의되어지고 있다.
- 이절에서는 가장 기본적인 텍스트에 대하여 살펴보도록 하겠습니다.
- 멀티미디어 데이터 중 비교적 저용량의 데이터 용량으로 만들수 있는 콘텐츠
- 데이터의 입력은 키보드와 스캐너 등을 이용
- 다른 멀티미디어 데이터들 보다 적은 기억장치의 용량
- 특이점 : 사용하는 언어에 따라 다른 특성을 소유
- 이용자가 작성하고자 하는 언어를 지원하는 도구를 사용

2. 텍스트의 표현

1) 문자코드

- 문자 데이터에는 숫자, 영문자, 특수 문자, 한글, 한자 등이 있으며, 각각을 구별할 수 있도록 여러 비트를 모아 조합하여 나타낸다.
- 문자 데이터의 표현 방식 - 언어에 따라 알파벳 사용권에서는 대개 8비트 코드를 사용하고, 한자를 사용하는 동양 권에서는 16비트 코드를 사용, 같은 8비트를 사용하는 경우에도 사용하는 코드에 따라 다소 차이점이 있음
- 2진화 10진 코드(BCD; Binary Coded Decimal)
- 아스키코드(ASCII; American Standard Coded for Information Interchange)
- 확장 2진화 10진 코드(EBCDIC; Extended Binary Coded Decimal Interchange Code)
- Unicode

2) BCD Code

- 10진수의 숫자를 나타내기 위하여 4비트로 조합
- 2의 4승의 문자 표현 - 영문자, 특수문자를 표현하기 위해 6비트로 구성
- 6비트 BCD 코드 - 2의 6승 = 64개의 표현, EBCDIC - BCD Zone 부분에서 2비트를 확장한 코드

10진수	BCD(Binary Coded Decimal)
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001

1자리 10진수 코드

3) ASCII Code

- ASCII 코드는 7개의 비트로 구성된 코드
- 3개의 존과 4개의 숫자로 구성된다.
- ASCII 코드의 특징은 앞부분 3개 부분의 코드는 영문자, 숫자, 특수 문자 등을 구분할 수 있도록 하였다.
- 정보의 호환성과 컴퓨터 통신을 효율적으로 이루기 위해 ASCII 코드에 패리티 비트를 추가하여 8자리의 2진 코드를 많이 사용하고 있음
- 패리티 비트는 패리티 비트를 포함하여 모든 비트열에 있는 1의 합이 항상 짝수 또는 홀수 개를 유지할 수 있도록 비트열에 추가한 검사 비트이다.
- ASCII 코드는 PC 기종에서 널리 사용되며, 데이터 통신용으로 많이 사용된다.

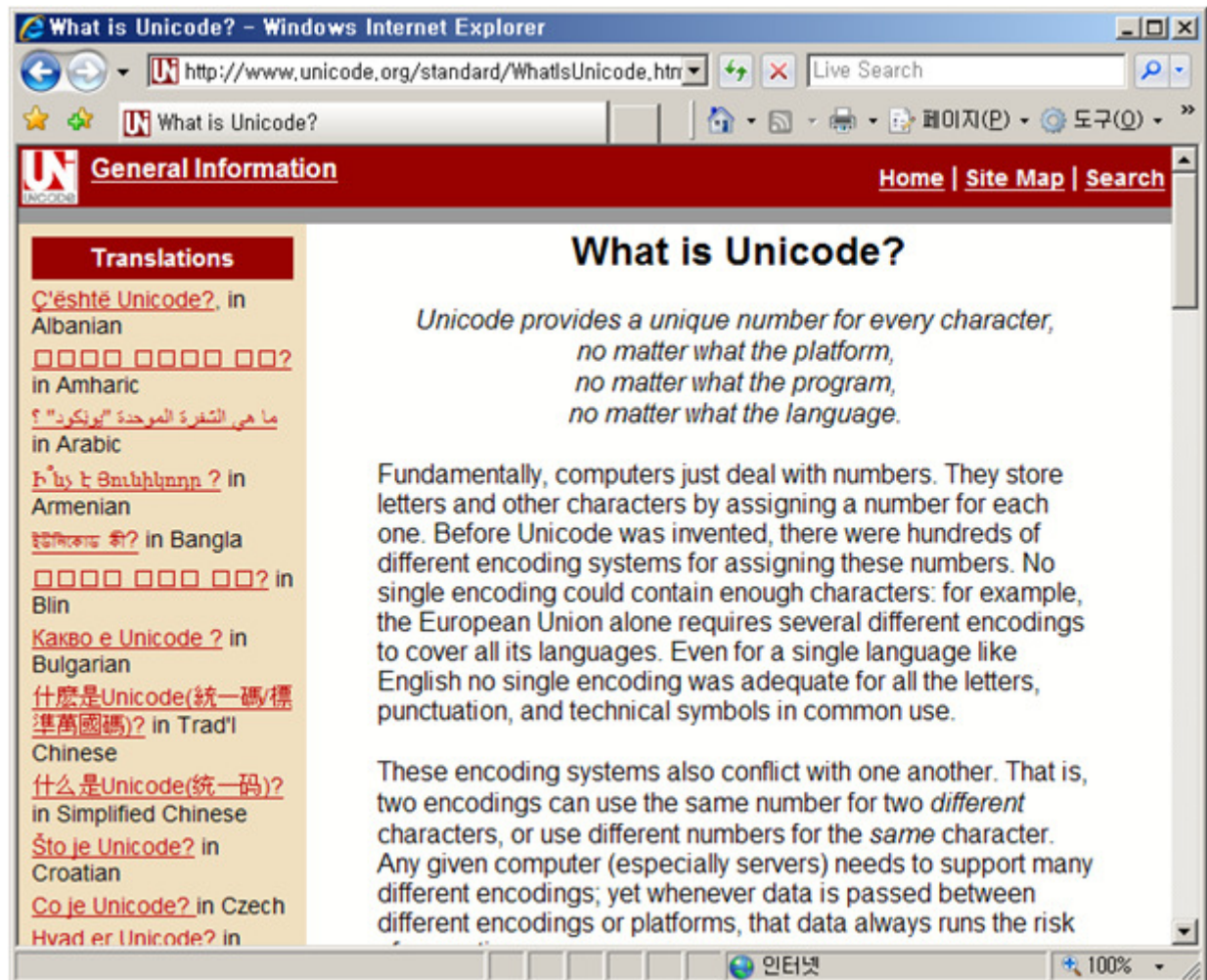
4) Unicode

- 각 나라별 코드 체계가 다르기 때문에 발생하는 문제 - 호환성 문제해결을 위해 유니코드 컨소시엄에서 관리 하는 산업 표준
- 현재 언어의 인코딩 체계를 포함 - 매우 큰 문자 집합 (단일 코드문자 집합), 하나의 문서에서 한국어, 영어, 일어 등을 작성 가능하게 함

- UTF-8 과 UTF-16으로 구별하여 표현

5) Unicode 콘소시엄

- <http://www.unicode.org/standard/WhatIsUnicode.html>



6) UTF-8

- 파일 사이즈를 줄이기 위해 표현 - ASCII 문자 코드는 1바이트, 다른 문자 코드 2바이트나 그 이상, 한글은 3바이트로 인코딩, ASCII 인코딩으로 저장된 파일을 ASCII 또는 UTF-8로 저장했을 경우 동일하게 8비트로 저장됨
- 상호 호환성을 유지함 - 이러한 호환성으로 인해 현재 기본 인코딩으로 많이 사용되어 지고 있음, 한글은 3바이트이므로 호환성에 문제 있음

7) UTF-16

- 2바이트를 이용하여 모든 문자를 표시
- 65356개의 문자 표시 가능 - 38885자는 주요 국가 언어 설정, 6400자는 사용자 정의 영역
- 영어로 작성된 문서는 UTF-8 보다 용량이 늘어남
- 영어 + 기타 문서로 작성된 문서는 UTF-16이 용량이 줄어듦

8) 한글코드

- KSC5601 - 한국표준 문자집합표준안, 2바이트를 사용하여 완성형 한글을 표현하는 방법, ASCII 문자코드에 대한 내용은 없음, KSC5636(ASCII 문자코드를 포함한 내용, ASCII 의 역슬래시를 ₩로 표현)
- EUC-KR - 벨 연구소에서 정의한 한글코드, 영어는 KSC5636체계로 표현하며, 한글은 KSC5601로 표현, ASCII 는 1Byte로 표현, 한글코드는 2Byte로 표현, 한국 표준협회에서 KSC5861로 등록됨

9) Font

- 글자의 모양을 표현하는 것
- 우리말로 글꼴이라 함 - 형태와 크기를 가리키는 용어로 사용되었으나 현재는 글자의 형태만을 지칭하는 서체와 같은 의미로 사용
- 폰트의 크기 - 글자의 높이를 기준으로 사용하는데 72포인트가 1인치에 해당
- 폰트의 속성 - 굵게, 가늘게, 글자색, 음영, 장평, 자간 등을 지정 가능, 영문의 경우에는 고정간격(Fixed) 폰트와 비례간격(Proportional) 폰트를 사용하며 대부분의 폰트는 비례 간격이며 고정 간격의 대표적인 폰트로 System font, Courier font 등을 들 수 있음, 굴림, 굴림체, 바탕, 바탕체(굴림,바탕은 비례간격 폰트, 굴림체, 바탕체는 고정간격)

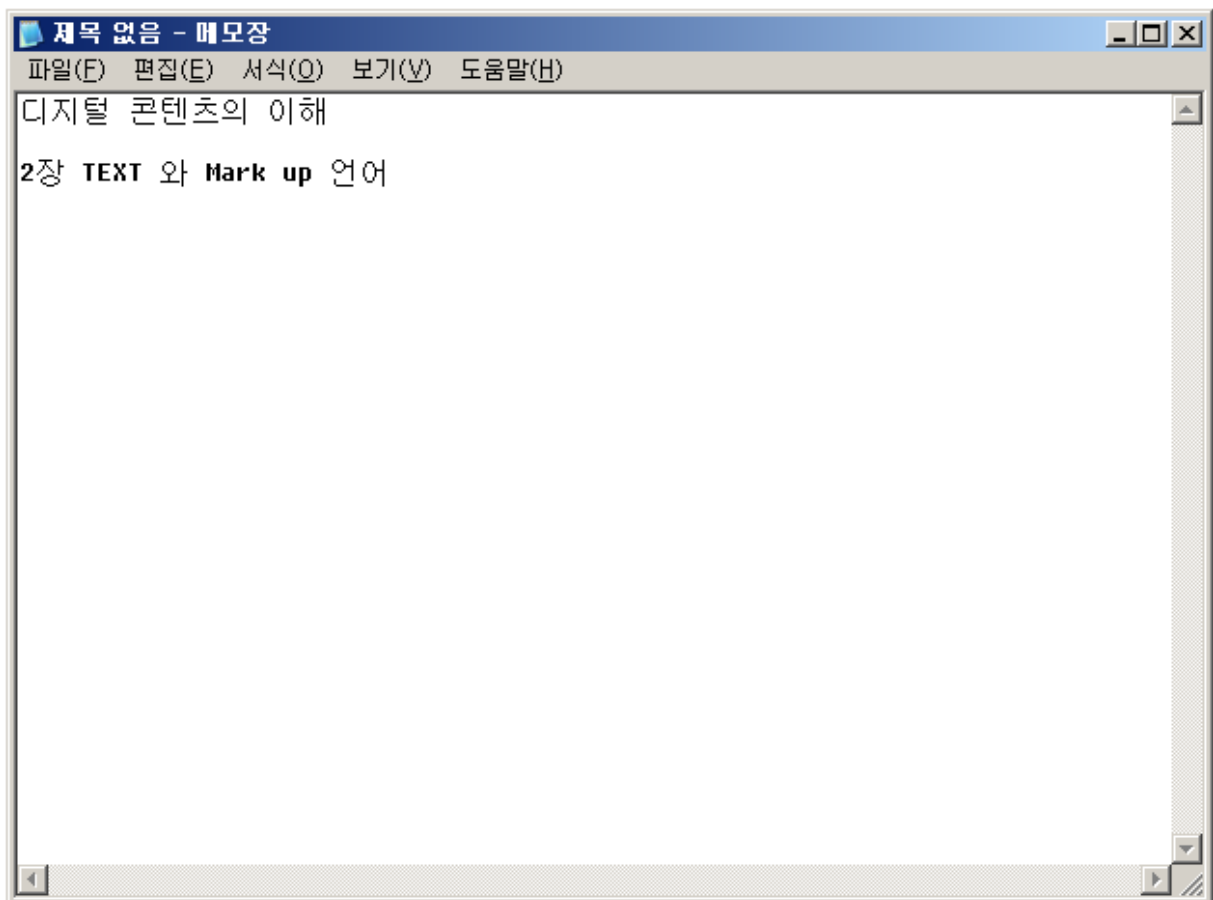
10) 폰트의 구분

- 비트맵 폰트 - 여러 개의 점으로 하나의 글자 폰트를 기록, 글자의 크기에 따라 용량이 늘어 날 수 있음, 지정된 크기가 아닌 경우 앨리어싱현상 이 나타날 수 있음
- 벡터 폰트 - 트루타입 폰트라고 하며, 글자 모양을 나타내기 위해 윤곽선을 수학적인 함수로 표현함, 시작하는 좌표의 위치($X1, Y1$)와 끝나는 좌표의 위치($X2, Y2$)를 지정하는 경우가 되므로 일그러짐이 없이 깨끗한 표현, 문자의 크기를 변화시킬 경우 수식의 크기를 나타내는 함수의 매개변수를 바꿈으로 해서 모양을 찌그러뜨리지 않고 보여줄 수 있으므로 수식적인 정보를 가지고 모든 크기에 대해 사용이 가능

3. 텍스트의 파일 형식

1) TEXT(*.TXT)

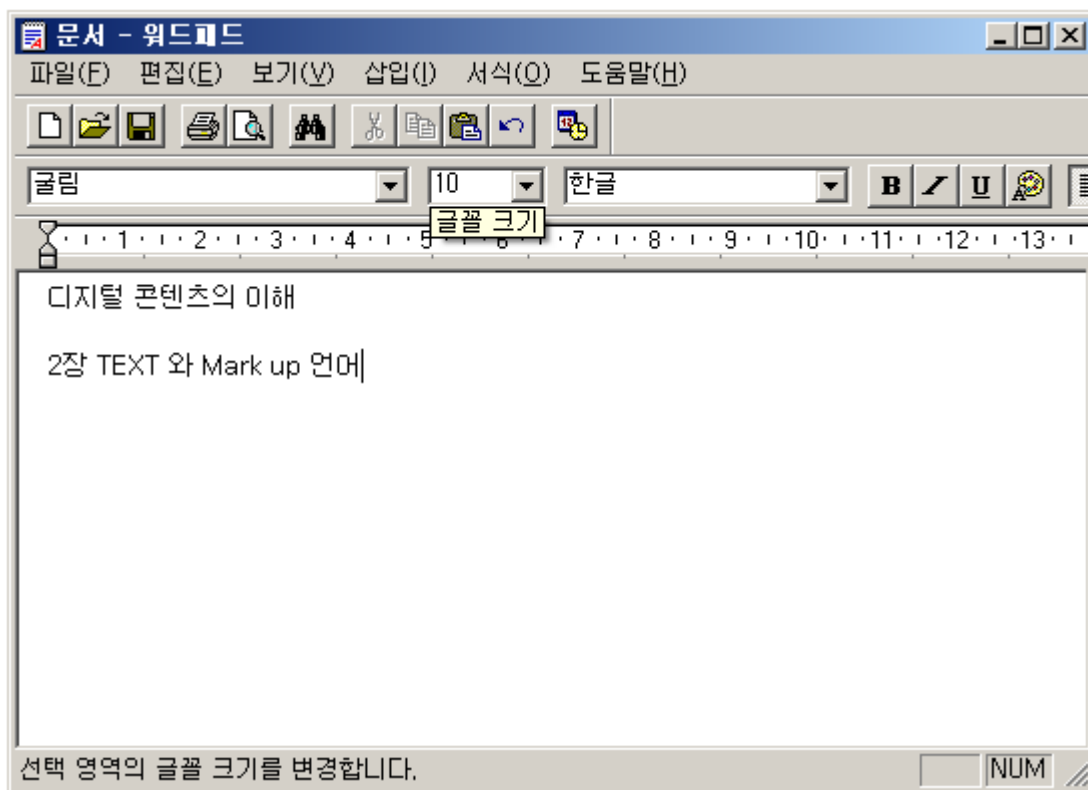
- 텍스트 정보는 영문의 경우 ASCII 코드, 한글의 경우에는 KS5601 코드를 사용하여 저장할 수 있으며, 파일은 *.txt 라는 확장자를 사용하여 저장
- TXT 포맷의 파일은 단순한 문자 코드의 집합으로 구성되어 있고 특정 문자를 표현하기 위한 폰트 정보 문장의 정렬 등 워드프로세서에서 필요한 정보를 갖고 있지 않기 때문에 편집되기 이전의 원시 자료로서의 가치만 있을 뿐 편집된 결과를 활용하기에는 부적합한 형식
- TXT 포맷으로 운영체제에서 기본적으로 제공하는 텍스트 편집기를 사용하여 만들어진 파일 형식
- ASCII 코드 또는 KS 한글 코드만으로 구성되어 있어 모든 텍스트 편집기나 워드프로세서에서 읽을 수 있어 단순한 텍스트 정보를 교환하기 위한 파일 포맷이므로 그래프나 수식 등을 제공하지 않음



원도우 메모장

2) 서식이 있는 텍스트 문서(*.rtf)

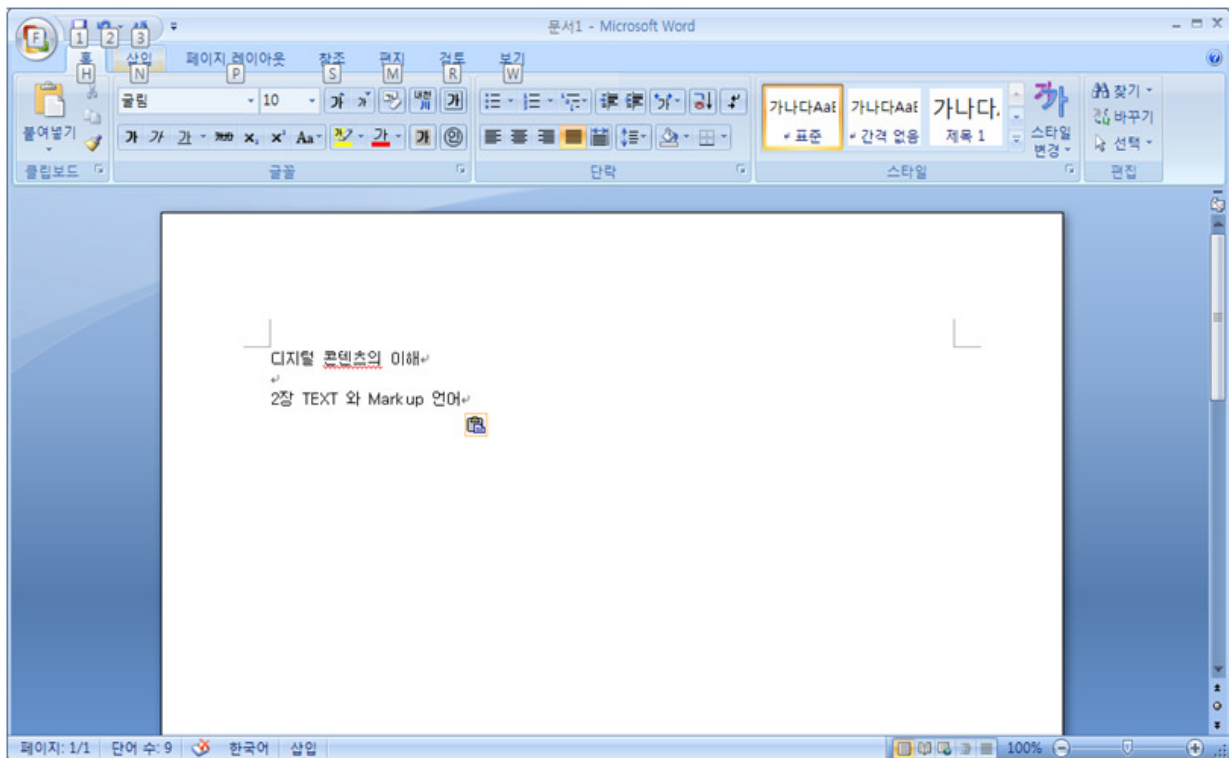
- RTF는 서로 다른 오퍼레이팅 시스템과 서로 다른 워드프로세서 사이에서도 텍스트 파일을 교환 할 수 있도록 위해 사용되는 문서형식으로 문서의 구조에 관련된 정보 역시 화면에서 읽을 수 있는 텍스트로 표시
- 표나 방정식과 같은 복잡한 기능을 사용하지 않는 경우 워드프로세서 사이에 문서를 교환할 수 있는 방법
- 파일 크기가 워드 프로세서 파일에 비해 수 배 이상 커지며 처리에 많은 시간이 걸리기 때문에 서로 다른 워드 프로세서 사이의 교환에만 사용하는 것이 바람직함
- MS 윈도우 워드패드에서 사용되어 지는 텍스트 문서



윈도우 워드패드

3) MS Word(*.doc, *.docx)

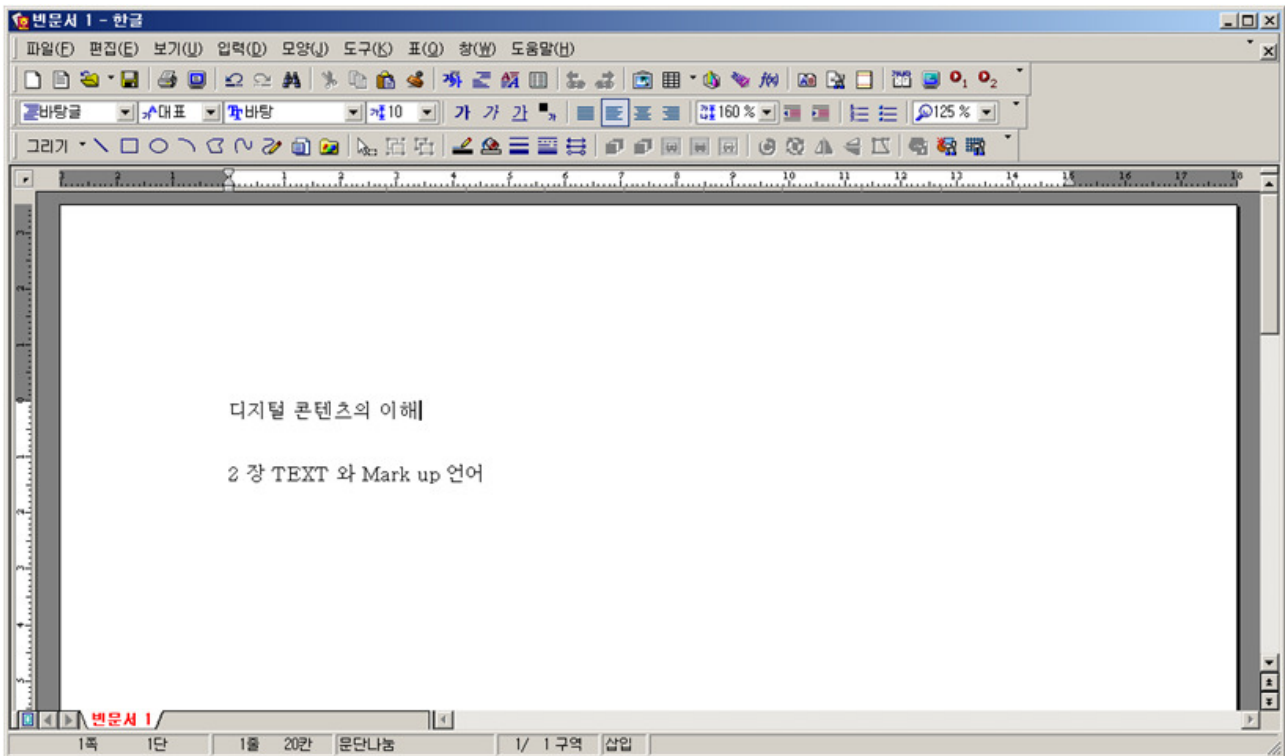
- DOC(Document)는 마이크로소프트사(MS)의 워드를 사용하여 작성된 문서 포맷
- 그래픽 효과, 그림 및 특수 기호, 서식 정보, 탭, 줄 간격 등 워드에서 필요한 모든 정보를 포함하고 있음
- 최근 2007 버전부터는 docx 라는 파일 확장명을 가짐
- 같은 포맷으로 Digital Equipment Corporation (DEC) 에서 Decwrite라는 워드에디터가 있으며 같은 *.doc 이라는 확장자를 사용하나 서로간의 호환성은 없음



MS 워드

4) 한글워드프로세서(*.hwp)

- HWP는 한글 워드 프로세서에서 사용하는 문서 포맷
- 다른 워드 프로세서의 경우 완성형 코드를 사용하는데 비해 조합형 코드를 사용하고 있는 특징이 있음
- 그래픽 정보, 수식 정보, 특수문자 등 워드프로세서에 필요한 모든 정보를 포함



한글 워드 프로세서