

0110과제 (a.minkuk.seo)

1. RAM 용어 조사

RAM(Random Access Memory)(<->Sequential Access Memory)

SAM

이전 순차적인 접근식 메모리의 경우 특정 장소에 저장된 데이터에 접근하기 위해서는 특정 순서를 거쳐야 했다(ex-카세트 테이프 다음곡 점프)

RAM

반면 ram 은 용어 그대로 무작위 접근 방식, 특정 장소에 저장된 데이터에 접근하기 위한 어떤 정해진 순서, 차례가 없다는 것을 의미한다. 램은 cpu가 처리할 데이터,명령어가 임시로 저장되는 곳이다. 명칭의 random 의 의미는 데이터에 접근(access)시간의 차이를 의미한다. 하드디스크가 데이터의 저장위치(주소)에 따라 접근시간의 차이가 발생하는 것에 반해 램은 데이터 처리시간이 그 저장위치와 무관하다는 점에서 random이란 용어가 사용된다

(이러한 구분도 현재 random access 아키텍처를 사용하는 타 메모리가 증가함에 따라

휘발성을 의미하는 RAM으로서의 의미가 더 강해지고 있다)

한글 표현 문자 인코딩 방식 조사

-컴퓨터 내 문자 표현?

기존 컴퓨터는 수치 연산을 위한 용어 그대로의 전자계산기의 역할만을 수행하였기에 문자 표현의 요구는 없었다. 하지만 시대의 흐름에 따라 컴퓨터의 역할이 증대되었고 이는 컴퓨터간의 문자 데이터 교환 시 표준 이슈를 발생시켰다. 이러한 연유로 ASCII(American Standard Code for Information Interchange)와 같은 표준 문자 인코딩이 나오게 되었다

문자를 표현하기 위해서는 우선 문자 집합을 정의하고 순서를 지정해야 하고(CCS, coded character set)이러한 문자 집합을 컴퓨터에 저장할 목적으로 비트 단위의 형태로 표현한 것을 **인코딩 방식 (CES, character encoding scheme) 이라고 한다**

2. 한글 표현 문자 인코딩 방식?

알파벳을 기초로하는 언어는 256개의 코드를 이용하여 문자 집합 표현이 가능하다. 하지만
한국어의 경우는 그 필요 개수가 더 많아 확장ASCII 코드로 모두 처리가 불가능 했다
한글표현 방법은 크게 조합형 과 완성형으로 나눌 수 있다.

-조합형 : 한글 제자원리 따라 초성,중성,종성에 각 코드를 할당

-완성형 : '가' '한' '나' 같이 완성된 한 글자마다 코드를 할당하는 방식 (한글 표준안)

| 구 분 | 자 판 입 력 | 한 글 구 현 |
|-----|-----------|-----------------------------|
| 조합형 | ㄱ + ㅏ + ㅇ | ㄱ + ㅏ + ㅇ = 강 |
| 완성형 | ㄱ + ㅏ + ㅇ | 가 각 간 갠 갈 감 갑 갓 강 갯 갯... |

-대표적 한글 인코딩 방식 (UTF-8 / EUC-KR, CP949(MS949))

1. UTF-8

조합형 방식의 문자 집합/ 유니코드 인코딩 방식 / 가장 범용적

초성 ,중성 , 종성 각1byte(일반적으로 한글 3byte인식) / 공백,영문은 1byte 인식

유니코드 통해 다양한 언어로 작성되는, 다양한 국가의 사용자가 참여하는 웹과 같은 분야에 적합한 방식

UTF-8은 유니코드를 위한 가변 길이 문자 인코딩 방식 중 하나로, 켄 톰프슨과 롭 파이크가 만들었다.
본래는 FSS-UTF(File System Safe UCS/Unicode Transformation Format)라는 이름으로 제안되었
다. UTF-8 인코딩은 유니코드 한 문자를 나타내기 위해 1바이트에서 4바이트까지를 사용한다. 예를
들어서, U+0000부터 U+007F 범위에 있는 ASCII 문자들은 UTF-8에서 1바이트만으로 표시된다. 4
바이트로 표현되는 문자는 모두 기본 다국어 평면(BMP) 바깥의 유니코드 문자이며, 거의 사용되지
않는다. UTF-16과 UTF-8 중 어느 인코딩이 더 적은 바이트를 사용하는지는 문자열에서 사용된 코드
포인트에 따라 달라지며, 실제로 DEFLATE와 같은 일반적인 압축 알고리즘을 사용할 경우 이 차이는
무시할 수 있을 정도이다. 이러한 압축 알고리즘을 사용하기 힘들고 크기가 중요할 경우 유니코드 표
준 압축 방식을 대신 사용할 수 있다.

-출처: [위키](#)

2. EUC-KR, CP949(MS949)

EUC-KR은 KS X 1001와 KS X 1003을 사용하는 8비트 문자 인코딩으로, EUC의 일종이며 **대표적인 한글 완성형 인코딩**이기 때문에 보통 완성형이라고 불린다.

- 출처 : [위키](#)

코드 페이지 949(CP949)는 마이크로소프트사가 도입한 코드 페이지이다. 본래는 KS C 5601의 완성형 한글을 표현한 코드 페이지였으나, 윈도 95부터는 **확장 완성형** 혹은 **통합형 한글 코드(Unified Hangul Code)**이라는 명칭으로 확장되어 **현대의 모든 한글을 수용**하게 되었다. 마이크로소프트에서는 이 인코딩을 기반 문자 집합 이름인 "ks_c_5601-1987"로 사용하고 있다. 다만 이 코드 페이지는 IANA에 등록되어 있지 않으므로 인터넷 상에서 정보를 주고받는 데 대한 표준은 아니다.

- 출처 : [위키](#)

완성형 인코딩 방식 / 한글을 2byte로 인식 / 한글 사용환경 에서만 한글이 제대로 출력되는 단점 / 한글과 영어만 사용하는 환경에 적합



완성형 중에서 EUC-KR의 경우에는 종종 빠져있는 문자가 있지만, **확장 완성형인 CP949의 경우에는 유니코드(UTF-8)에서 표현할 수 있는 모든 문자를 표현** 가능합니다. 따라서 웹을 사용하지 않는 윈도우 환경이라면 크게 불편함이 없을 수 있습니다.

참조 <http://studyforus.tistory.com/167>