

인코딩과 디코딩 (Encoding & Decoding)

컴퓨터는 문자를 인식할 수 없기 때문에 숫자로 변환되어 저장됩니다. 변환해주기 위해서는 기준이 있어야하는데 이것을 **문자 코드**라고 하며 대표적으로 **ASCII코드** 또는 **유니코드**가 있습니다.

이렇게 문자 코드를 기준으로 문자를 코드로 변환하는 것을 **문자 인코딩(encoding)** 이라고하고 코드를 문자로 변환하는 것을 **문자 디코딩(decoding)** 이라고 합니다.

인코딩/디코딩은 정보의 형태나 형식을 변환하는 처리에 대해 표준화하고 보안, 처리 속도 향상, 저장 공간 절약 등의 목적으로 사용합니다.

글 내용이 많지만 문자 인코딩이 궁금하다면 꼭 읽기 바란다.

우선 ^{원화}인코딩을 위키 백과에 나온 설명으로 보면 '정보의 형태나 형식을 변환하는 처리나 처리 방식이다.' 라고 정의 되어 있다.
(출처 : <https://ko.wikipedia.org/wiki/%EC%9D%B8%EC%BD%94%EB%94%A9>)
즉, 어떤 정보를 미리 약속한 규칙으로 가공한다고 보면 된다.

인코딩은 문자, 동영상, 사진 등 다양하게 분야에서 많이 사용 되고 있다.
많은 인코딩 중 문자 인코딩에 대해 말하려고 한다.

먼저, 우리가 쉽게 접할 수 있는 문제로 보겠다.

문자 인코딩으로 인한 피해 사례

- 인터넷에 좋은 자료가 있어서 첨부 파일을 다운로드 받았는데 글이 다 깨져 있는 경우
- 어떤 사이트에 들어갔는데 사이트의 내용 글이 깨져 있는 경우
- 공개된 소스 코드 파일을 다운로드 받았는데 주석 내용이 깨져 있는 경우
- 내가 전달한 파일이 다른 사람에게는 그 파일 내용이 제대로 보이지 않을 때

위와 같이 문자가 깨져서 알아볼 수 없는 상황은 모두가 아마도 한번씩은 겪어 봤을 것이다.

원인이 뭘까?

글의 제목 때문에 짐작은 하고 있으리라 본다.
문제는 문자 인코딩이 다르기 때문이다.

예를 들어 설명 하겠다. (실제 규칙이 아닙니다.)

나는 "니 내 누군지 아니?"라는 내용을 저장해서 파일로 만들었다. 그 파일은 vigli 라는 인코딩 방법으로 저장된 파일이다.
vigli의 규칙은 아래와 같다.

문자	코드표
니	0
내	1
누	2
군	3
지	4
아	5
니	6
?	7
(공백)	8

표1. vigli 인코딩의 코드표



눈으로는 '니 내 누군지 아니?' 라고 보이지만 컴퓨터에 저장된 파일은 vigli 인코딩으로 '08182348567' 이렇게 저장 되어있다.
그런데 상대방은 내가 보낸 파일을 vigli의 규칙과 동일한 규칙을 가진 인코딩이 아닌 다른 인코딩 방식으로 보기 때문에
08182348567'을 파일에서 읽었을때 사람의 눈으로 표현한 문자를 '뽕뽕뽕뽕뽕뽕' 와 같이 보이게한다. 그래서 내가 원하는
내용과 다르게 보이는 것이다.