**Techniques for Working with Big Data**

**(빅데이터를 다루기 위한 기술)**

raw data -> data-pre-proeccsing -> processing -> information

Data Collection 데이터 수집

1>. Data Pre-processing

- Class Labeling

-> number, text, digital images, digital video data, digital audio data

- Data cleansing = data scrubbing

> 단순히 text의 오타를 정정하는 걸 넘어서 이미지 보정, 오디오 음질보정, 동영상 품질보정 등등 범위가 훨씬 넓음

- Dealing with Missing values

-> 다루기 힘듬. 빅데이터에서 missing value는 큰일임.

**2> Case Specific**

- Text data mining

-> 텍스트로부터 가치있고, 구조화되지 않은 데이터를 끌어내는 작업

- Data Masking

-> 보안이 필요한 데이터를 식별해 암호화하거나 블라인드 처리하는 기술

(Confidentiality - preserving data mining techniques)