

# 텍스트 마이닝과 데이터 마이닝

# Part 05. 토픽 모델링과 워드 클라우드

정 정 민

## Chapter 14. 워드 클라우드

---

1. 워드 클라우드란?
2. 워드 클라우드 실습

워드 클라우드란?

## 워드 클라우드

- 백문이 불여일견
- 옆의 시각화 이미지를 ‘워드 클라우드’라고 함



# 워드 클라우드(Word Cloud)란?

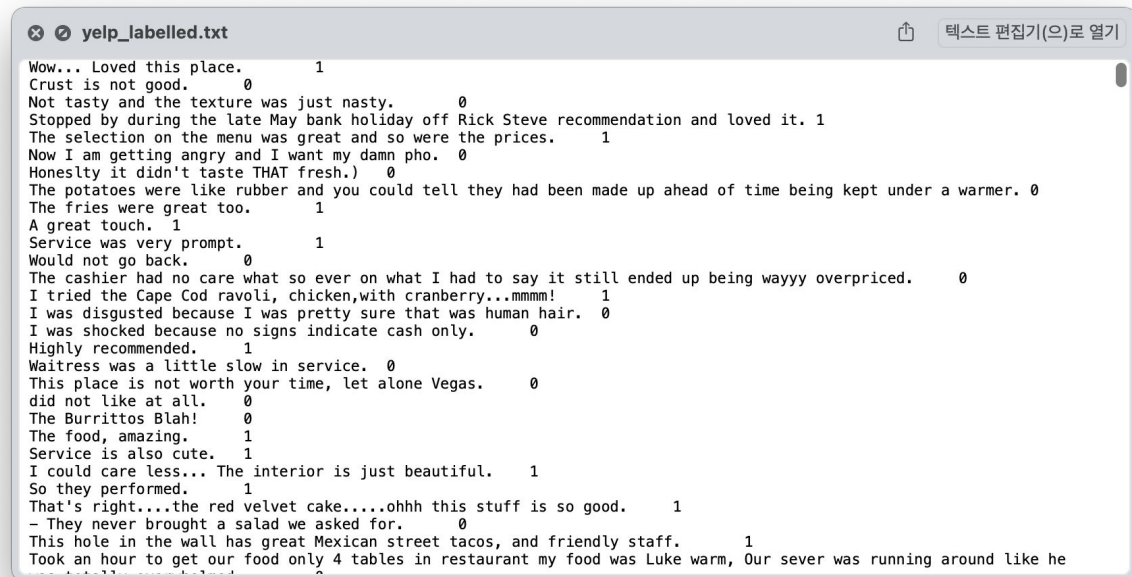
---

- 워드 클라우드란
  - 텍스트 데이터를 시각적으로 표현한 것
  - 문서나 데이터에서 가장 빈번하게 등장하는 단어들을 시각적으로 돋보이게 하는 방법
- 자주 등장하는 단어는 더 크고 두드러지게 표시됨
- 대량의 텍스트 데이터에서 중요한 키워드의 빠르고 직관적 이해를 위함
- 데이터 해석을 위한 직관적인 접근 가능
- 아래와 같은 활용 분야
  - 소셜 미디어 : 트렌드 & 핫 토픽
  - 시장 조사 분석 : 소비자 리뷰 & 피드백
  - 교육 : 학습 자료 내용 요약 및 시각화

# 워드 클라우드 실습

# 전처리 및 사용 데이터

- 보고자 하는 **목적에 맞춰 데이터를 전처리** 해야 함
  - 뉴스, 소셜 미디어, 학술 논문 등은 ‘명사’를 기반으로
  - 리뷰, 고객 피드백과 같은 감정의 통찰을 얻고자 한다면 ‘형용사’ 혹은 ‘동사’ 활용
  - 전체 맥락을 파악하기 위해 전체 데이터 활용
- 일반적으로 **특수 문자와 숫자는 제거**
- 이번 실습에서도 Yelp 데이터 활용 ([링크](#))
  - 리뷰 데이터이므로 전체 데이터 활용
  - 정규화(소문자화), 특수 문자 제거,
  - 단어 기반 tokenize, Stop word 제거, Stemming



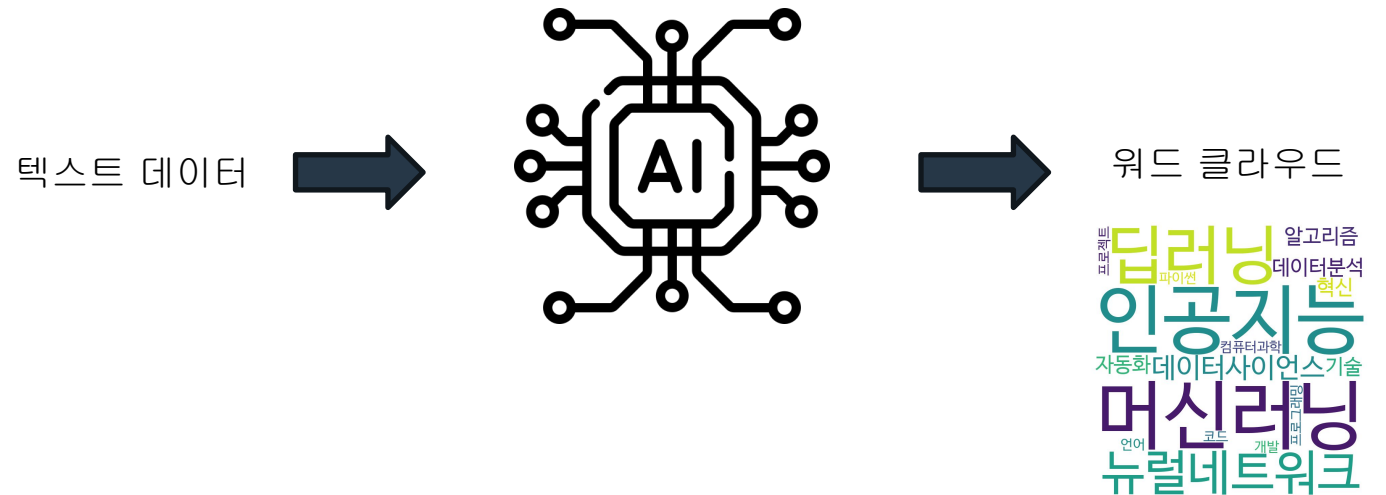
The screenshot shows a text editor window titled 'yelp\_labelled.txt'. The text inside consists of multiple lines of Yelp reviews, each followed by a sentiment label (0 for negative, 1 for positive). The reviews cover various topics like food quality, service, and pricing. The labels are placed at the end of each line, indicating the sentiment of the entire review.

```
Wow... Loved this place. 1
Crust is not good. 0
Not tasty and the texture was just nasty. 0
Stopped by during the late May bank holiday off Rick Steve recommendation and loved it. 1
The selection on the menu was great and so were the prices. 1
Now I am getting angry and I want my damn pho. 0
Honeslty it didn't taste THAT fresh.) 0
The potatoes were like rubber and you could tell they had been made up ahead of time being kept under a warmer. 0
The fries were great too. 1
A great touch. 1
Service was very prompt. 1
Would not go back. 0
The cashier had no care what so ever on what I had to say it still ended up being wayyy overpriced. 0
I tried the Cape Cod ravoli, chicken,with cranberry...mmmm! 1
I was disgusted because I was pretty sure that was human hair. 0
I was shocked because no signs indicate cash only. 0
Highly recommended. 1
Waitress was a little slow in service. 0
This place is not worth your time, let alone Vegas. 0
did not like at all. 0
The Burrittos Blah! 0
The food, amazing. 1
Service is also cute. 1
I could care less... The interior is just beautiful. 1
So they performed. 1
That's right....the red velvet cake....ohhh this stuff is so good. 1
- They never brought a salad we asked for. 0
This hole in the wall has great Mexican street tacos, and friendly staff. 1
Took an hour to get our food only 4 tables in restaurant my food was Luke warm, Our sever was running around like he
```



## 문제 정의

- 풀어야 하는 문제
  - 데이터 전체를 활용한 워드 클라우드 생성
- 입력과 출력
  - 입력 : 텍스트 전체 텍스트
  - 출력 : 워드 클라우드





**E.O.D**