

✓ 워드 클라우드(word cloud) = 태그 클라우드(tag cloud)

- 텍스트에서 나타나는 단어들을 빈도에 따라 시각적으로 표현한 그래픽
- 워드 클라우드의 특징

- 사용자가 입력한 데이터에서 얻어진 태그들을 분석
- 중요도나 인기도 등을 고려하여 시각적으로 늘어 놓음
- 언급 빈도가 높은 글자가 크게 표기
- 빈도에 따라 글자 컬러도 다르게 나타남
- 문서의 주요 키워드나 중요도를 한눈에 보고 이해할 수 있음

- 워드 클라우드의 활용

- 문서나 수집한 비정형 데이터의 특징을 도출하기 위해 사용
- 서술형 데이터를 분석할 때 효과적임
- 클라이언트의 문제 이해, 비즈니스 활동 가속화, 직원 감정 분석, 기술 데이터 > 단순화 등에 사용할 수 있음

- 워드 클라우드의 단점

- 단어 간의 관계를 표현할 수 없음
- 정보가 편향될 수 있음

```

1 #https://blog.naver.com/j7youngh/222902787053
2 !pip install wikipedia-api          # wikipedia api 설치
3 import wikipediaapi                  # wikipedia api 호출
4
5 #wiki = wikipediaapi.Wikipedia('en') # 한국어 사이트 접속(한글은 en => ko)
6 wiki = wikipediaapi.Wikipedia(user_agent='your-user-agent', language='en')
7 page = wiki.page("Harry Potter")   # 검색어에 해당되는 페이지 전체 가져오기
8 Harry_text = page.text                # 크롤링한 텍스트 저장
9
10 print(Harry_text)

```



J. K. Rowling's personal website
 Harry Potter movies—Official website (Warner Bros.)
 Harry Potter at Bloomsbury.com (International publisher)
 Harry Potter at Scholastic.com (US publisher)
 Harry Potter at Raincoast.com (Canadian publisher)
 Harry Potter collected news and commentary at The Guardian
 Harry Potter collected news and commentary at The New York Times
 The Wizarding World of Harry Potter at Orlando resort, Florida





https://colab.research.google.com/drive/1kvDStME3MI-2jJC9uv_-MtCYvMFRQzXX#scrollTo=gCaxJrLehWEf

→ ndarray (900, 900) [show data](#)



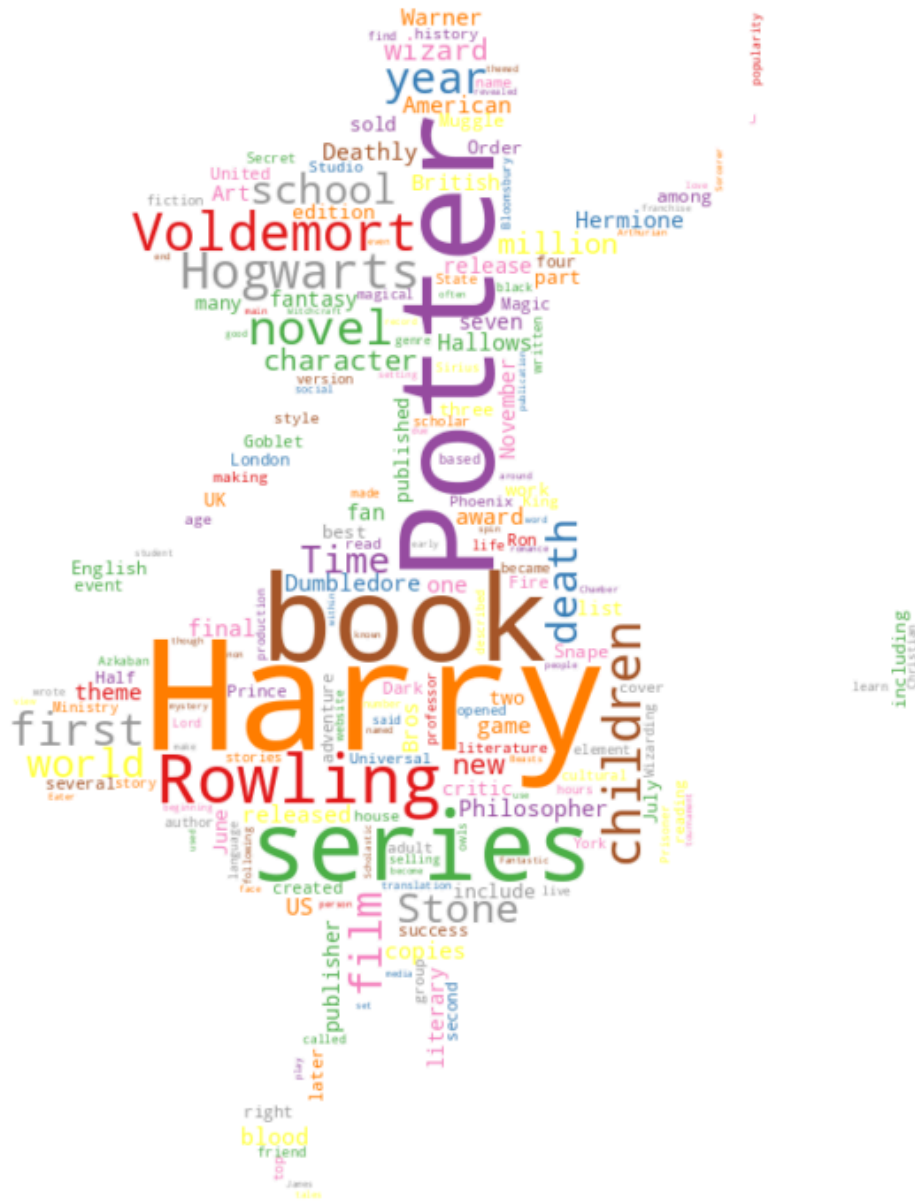


```
1 from wordcloud import WordCloud, STOPWORDS # WordCloud와 STOPWORD 호출
2 import matplotlib.pyplot as plt           # matplotlib.pyplot를 plt로 단축
3 import numpy as np                         # 이미지를 array 데이터로 변환
4 from PIL import Image                     # 이미지 분석 및 처리
5
6 mask = np.array(Image.open("/content/drive/MyDrive/데이터과학과인공지능/데이터과학과인공지능-1-빅데이터-인공지능-컴퓨팅사고-엔트리/alice_mask.png")) # 이미지를 불러옴
7
8 wordcloud = WordCloud(width = 3000, height = 2000, random_state=1, W
9                       background_color='white', colormap='Set1', W
10                      collocations=False, stopwords = STOPWORDS, W
11                      mask=mask)
12
13 wordcloud.generate(Harry_text)             # 워드 클라우드 생성
14
15 plt.figure(figsize=(10,10))               # 이미지 크기 지정
16 plt.imshow(wordcloud, interpolation="bilinear") # 픽셀을 부드럽게 색으로 채움
17 plt.axis("off")
```

25. 3. 5. 오후 1:31

워드클라우드(WordCloud) - 핵심 단어 시각화 하기 - Colab

→ $(-0.5, 899.5, 899.5, -0.5)$



워드클라우드, BiqZami로 해보기