**0. 실험 수행 순서**

1) 80개의 data-set으로 Fine-tuning 학습된 pretrained T5 model 사용 추상적요약 수행

(사실은 학습하지않고 wiki-how 로 학습된 오픈소스 이용하여 추상적 요약 수행)

2) bleu metric으로 성능 분석

3 keyword 추출 성능 분석 (YAKE, BERT, TextRank <-> 사람)

결과 한눈에 확인 (전체 99개 text)

<https://github.com/English-Vending-Machine/jupyter_EVM/blob/koh/ML/abstract/main.ipynb>

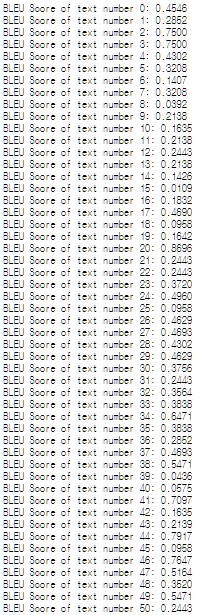
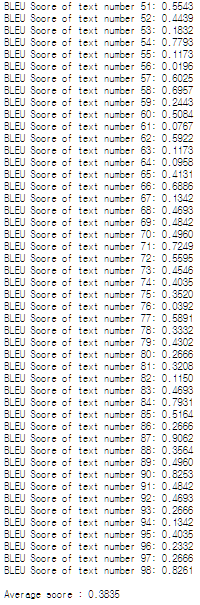
**1. 추상적 요약 결과**





**2. bleu metric score**

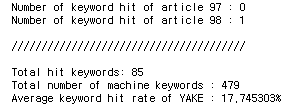
※ 팀원 각자 다른 지문을 하나씩 요약했으므로 각 지문에 대한 reference(사람이 요약한 결과)가 1개뿐이고, T5 요약(1문장)에 비해 reference(3,4 문장)가 너무 길어 결과 score가 너무 낮게 나왔다. 따라서 reference 가중치를 (1:0:0:0)으로 두었다. (각각 1~4번째 reference의 가중치를 나타냄)

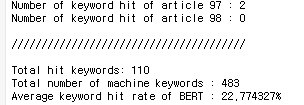
**평균 score : 0.3835 : 두 text가 완전히 일치할경우 점수는 1.0이다. 점수가 절반인 0.5도안되므로, 사람이 요약 했을 때 보다 기계가 추상적 요약 했을 때 차이가 매우 크게 난다. 중심 요약문의 내용이 매우 다르므로 두 요약문을 읽고 이해한 방향이 다를 가능성이 크다.**

3. keyword 추출 결과분석

- YAKE



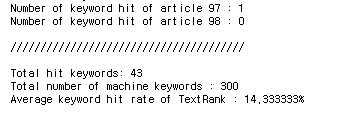
- BERT



- TextRank

TextRank는 마침표로 문장 구분을 하지못하여 마침표 + 공백으로 일일이 모든 요약문을수정해주었다

- 그리고 지문의 길이가 짧으면 설정한 keyword 개수 만큼 생성하지 못하여 전체 keyword 개수가 적다



-> 추출적 요약에 비해서(?) 너무나 낮은 결과가 나왔다

🡪 당연한 결과지만, 추상적 요약은 빈칸 생성에 사용 불가!