2020/09/29 회의록

1. 데이터 파트 나누기(각 50개 총 200개)

* 방법1: 가수별로 나누기

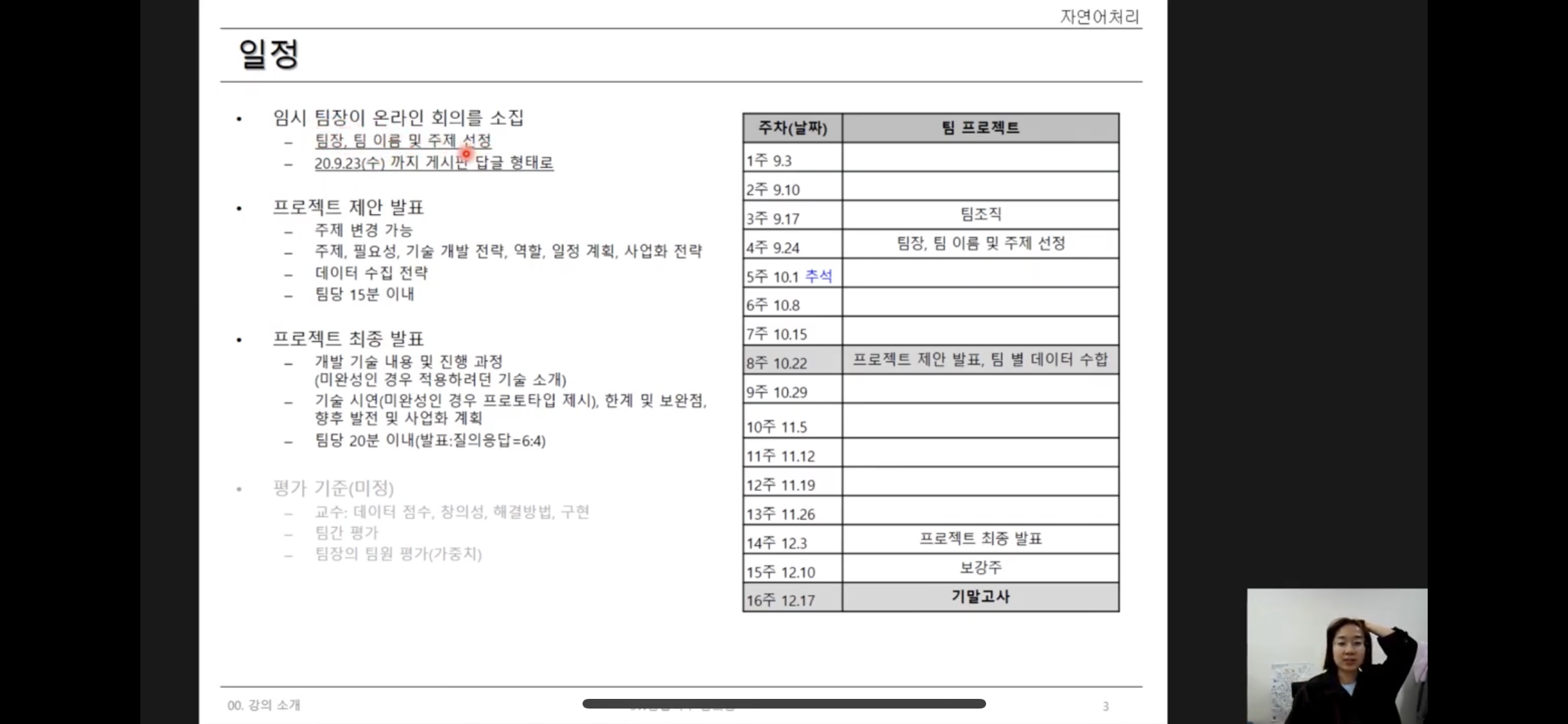
참고-위키피디아 시대별 아이돌 정리 표 <https://ko.wikipedia.org/wiki/%EB%8C%80%ED%95%9C%EB%AF%BC%EA%B5%AD%EC%9D%98\_%EC%95%84%EC%9D%B4%EB%8F%8C\_%EA%B7%B8%EB%A3%B9\_%EB%AA%A9%EB%A1%9D>

* 방법2: 1992~1999, 총 8년 이므로 2년씩 나누기

1. 데이터 수집 방법 논의

* 캐글
* 웹 크롤링

1. 데이터 수집 기간 및 중간 발표 준비 기간

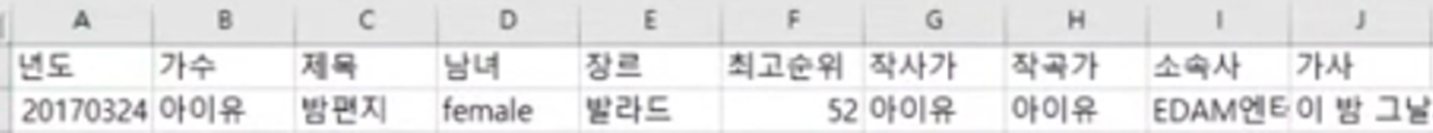


1. 깃허브 레포지터리

https://github.com/jiminAn/Kpop\_NLP\_Project

1. 데이터 수집 시 주의사항

* Cvs 파일이 아닌 tvs파일로 저장할 것(서브라임 텍스트 이용)
* 칼럼은 아래와 같이 수집



* 웹 크롤링 시 노이즈 제거 및 수정이 필요할때는 아래 방법이용 ctrl+h

