**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>

92~93 : 준영

94~95 : 은후

96~97 : 지민

97~99 : 선우

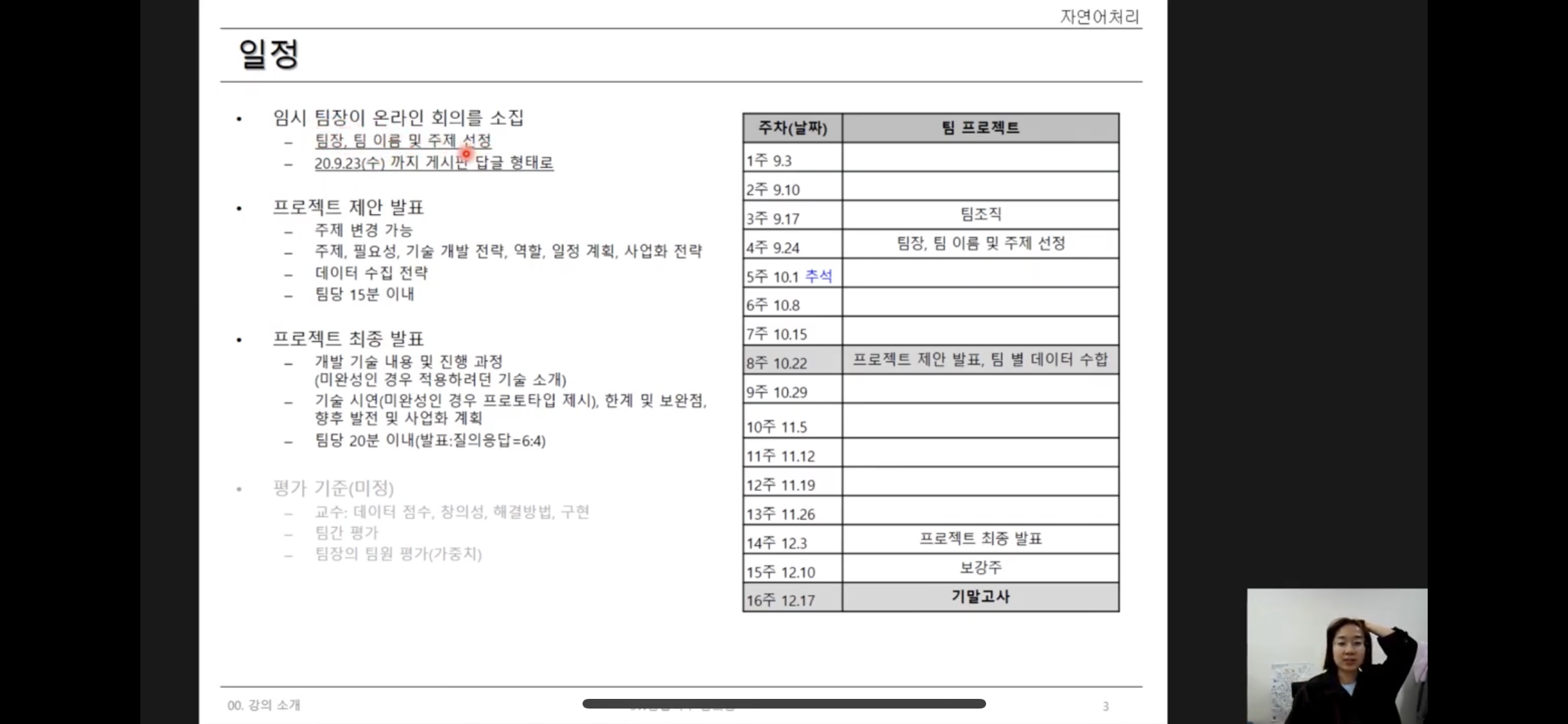
* 방법2: 1992~1999, 총 8년 이므로 2년씩 나누기

1. **데이터 수집 방법 논의**

* 캐글
* 웹 크롤링

1. **데이터 수집 기간 및 중간 발표 준비 기간**

* 10.11 까지 데이터 수집
* 10.11 중간발표 준비 기간

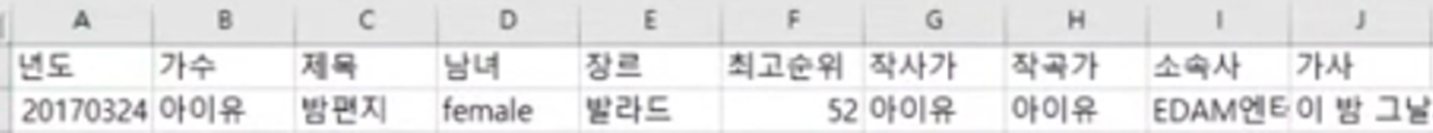


1. **깃허브 레포지터리**

<https://github.com/jiminAn/Kpop_NLP_Project>

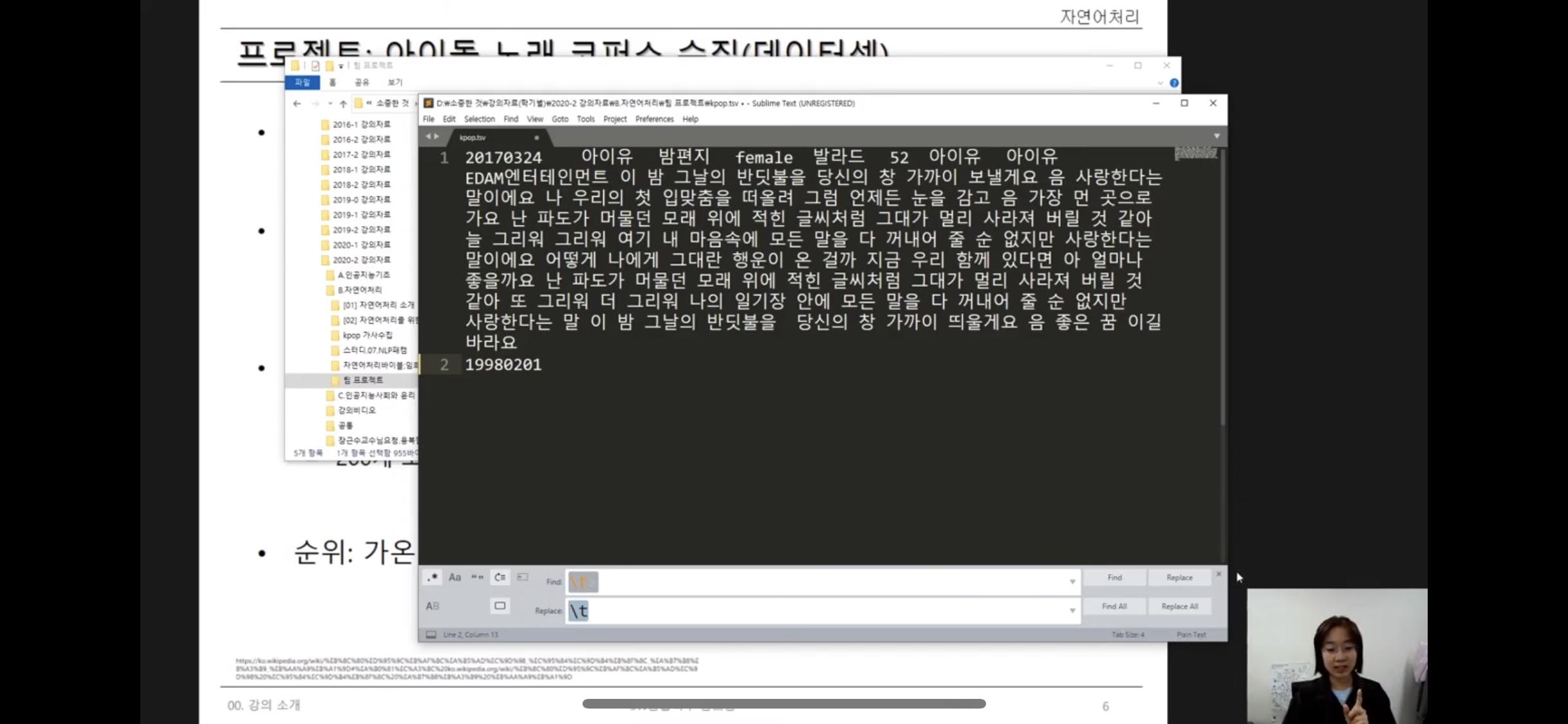
1. **데이터 수집 시 주의사항**

* Csv 파일이 아닌 tsv파일로 저장할 것(서브라임 텍스트 이용)
* 칼럼은 아래와 같이 수집





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



1. **10/4일까지 준영님이 올려주신 크롤링방법 사용해서 각자 맡은 부분 데이터 수집해보기**
2. **분류 방법 고려해보기 : 주제 구체화**

* 장르, 가사