

“담대한” 테이북

제 4회 SKKU AI x BOOKATHON



목차



1. 데이터 수집



2. 데이터 전처리



3. 모델 학습



4. 수필 생성

데이터 수집

〈수집한 데이터〉

1. 글틴
2. 대표 에세이 문학회
3. 브런치
4. 동아신춘문에
5. 모두의 말뭉치 - 비출판물 데이터

〈데이터 수집 시 고려사항〉

1. 학습을 위해 최대한 많이 수집하기
2. 어떤 종류의 데이터를 수집할지
3. 저작권이 위배되지 않는지

≡

문학광장

🔍

글틴 > 쓰면서 뒀글 > 수필

수필

📁

지난해

우리는 모두 바람 아래 사는 사람

2022-04-30 [1]

또 왔다. 새학기가. 새로운 교실과 새로운 아침이. 들뜬 분위기와 영 켜이지 않는 공기가. 이미 여러번의 새학기를 경험한 나는 익숙하게 적응에 몸을 던졌다. 처음에나 정신 없이 금방 익숙해질 거라고. 어느새 십대의 끝자락에 서게 되었지만 뭐 크게 다르겠냐고. 그런 열은 믿음으로 하루하루 견뎌보고자 했다. ...

Representation

ESSAY

대표 에세이 문학회

격인 할마디!

사는 것이 내 일이요, 내 기술이다. -괴테

문학회소개

회원소개

동인지소개

작품보기

열린마당

MAIN

CONTACT

총 게시판 : 41 (Total 41 Articles)

(1 / 3)

번호	[동인지소개]	최근/전체	회 장	개설일
41	[제 39 호] < 존재와 시간 > (2022)	0 / 0	고재종	23-01-01
40	[제 38 호] < 모든 이의 아침 > (2021)	0 / 53	허문정	21-10-23
39	[제 37 호] <<生, 푸른 불빛>> (2020)	0 / 43	허해순	20-12-11
38	[제 36 호] 나는 ㅇㅇ이다(2019)	0 / 45	김경순	19-12-05
37	[제 35 호] 나에게로 온 날들(2018)	0 / 43	이은영	19-01-02
36	[제 34 호] 씬- 여행, 씬표와 느낌표 사이(2017)	0 / 44	김현희	17-12-01
35	[제 33 호] 골목길의 고백(2016)	0 / 45	김윤희	16-11-10
34	[제 32 호] 짧지만 깊은 이야기(2015)	0 / 39	박경희	15-12-22
33	[제 31 호] 바람결에 수긋수긋(2014)	0 / 40	김선화	14-11-27
32	[제 30 호] 대표에세이 選集 30주년기념(2013)	0 / 53	정태현	13-11-24
31	[제 29 호] 마흔다섯 개의 느낌표(2012)	0 / 45	장경환	12-11-27
30	[제 28 호] 결혼 닷 또는 닷(2011)	0 / 37	김지현	11-11-28
29	[제 27 호] 교과서에 실고 싶은 나의 수필(2010)	0 / 47	조현세	10-11-05
28	[제 26 호] 그림자, 연인(2009)	0 / 33	박미경	09-12-04

공모요강

역대수상자

올해 당선작

지난 당선작

2022 동아신춘문에 당선작

≡ 중편소설

플렉시테리언 (이안리)

≡ 단편소설

무겁고 높은 (김기태)

≡ 시

경유지에서 (채윤희)

≡ 시 조

꽃을 더듬어 읽다 (김성애)

≡ 희곡

뉴 트롤러 딜레마 (구지수)

≡ 동 화

아무 일도 아닌 것 같지만 (김란)

≡ brunch

감성 에세이

🔍

에세이

일상

감성에세이

일상에세이

수필

감성

감성글

산문

사진에세이

포토에세이

감성일기

산문집

카페 사장님을 올렸습니다.(feat. 만원의 행복)

마음을 나누고 행복을 선물 받았습니다-♡ | 지난해 출간을 두 달여 앞둔 어느 겨울 주말이었 습니다. 남편은 회의 자료를 준비해야 했고, 전 마음을 앞두고 글을 써야만 했습니다. 그...

댓글1 · 1시간전 · by 초원의 빛

내가 그렇게도 좋으니?

-나를 조건 없이 사랑해 주는 사람에게 오늘도 배운다. | 나를 조건 없이 사랑해 주는 사람이 존재할 줄이야. 그 사람은 정말 나만 본다. 내가 왜 좋냐고 하면 그냥 다 좋다고 한다. 내...

댓글0 · 2시간전 · by 재하

러닝 80분 뒀몸 36개 다리 160회

D15 체지방률 33.2 골격근량 21.2/인바디 73점 BMI 24.3

댓글0 · 2시간전 · by Ophelia

추천작가

지수연

글 97 · 구독자 730

부크림

글 934 · 구독자 1,104

Anarchist

글 123 · 구독자 20

예경

글 5 · 구독자 12

여마

글 6 · 구독자 10

세인

글 12 · 구독자 19

글틴
브런치
대표 에세이 문학회
동아신춘문에

=> 웹 크롤링을 통한
데이터 수집

+) 모두의 말뭉치
=> 홈페이지 승인 신청

데이터 전처리

중복되는 공백과 문자 처리, 개행문자 제거, 그 밖에 의미 없는 text 제거

```
In [2]: #전처리 함수
def PreProcess(text):
    text = re.sub(pattern='Posted on [0-9]{4} [0-9]{2} [0-9]{2}', repl='', string=text)
    text = re.sub(pattern='Posted on [0-9]{8}', repl='', string=text)
    text = re.sub(pattern='[0-9]{4}년 [0-9]{,2}월 [0-9]{,2}일 [0-9]{,2}시 [0-9]{,2}분 [0-9]{,2}초', repl='', string=text)
    text = re.sub(pattern='[0-9]{4}. [0-9]{,2}. [0-9]{,2}', repl='', string=text)
    _filter = re.compile('[\s]')
    text = _filter.sub(' ', text)

    _filter = re.compile('[\r\n]+')
    text = _filter.sub('', text)

    _filter = re.compile('[\n]')
    text = _filter.sub(' ', text)

    _filter = re.compile('[^가-힣 0-9 a-z A-Z ., #, ? !]+')
    text = _filter.sub('', text)

    text = re.sub("[.]", ". ", text).strip()

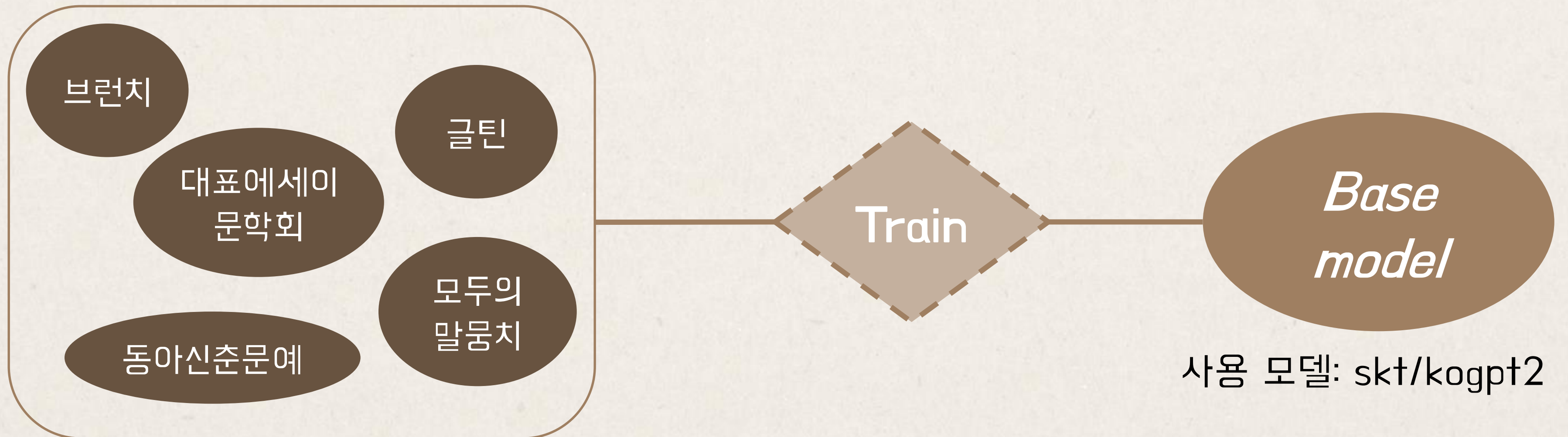
    # 연속된 공백 치환
    text = re.sub(r"#+", " ", text).strip()

    return text
```

모델 학습

목표: base model 생성하기, Pre-training

전처리한 데이터



담대함?

=> 주체성을 찾는 과정, “무언가를 깨트리는”

=> 성장, 모더니즘 발상

=> 인상주의 화가 모네 발상

=> 주인공이 모네의 작품과 타인을 만나며

자신의 삶을 담대하게 맞서 나가는 스토리 및 에피소드 구상

에피소드 별 키워드 지정



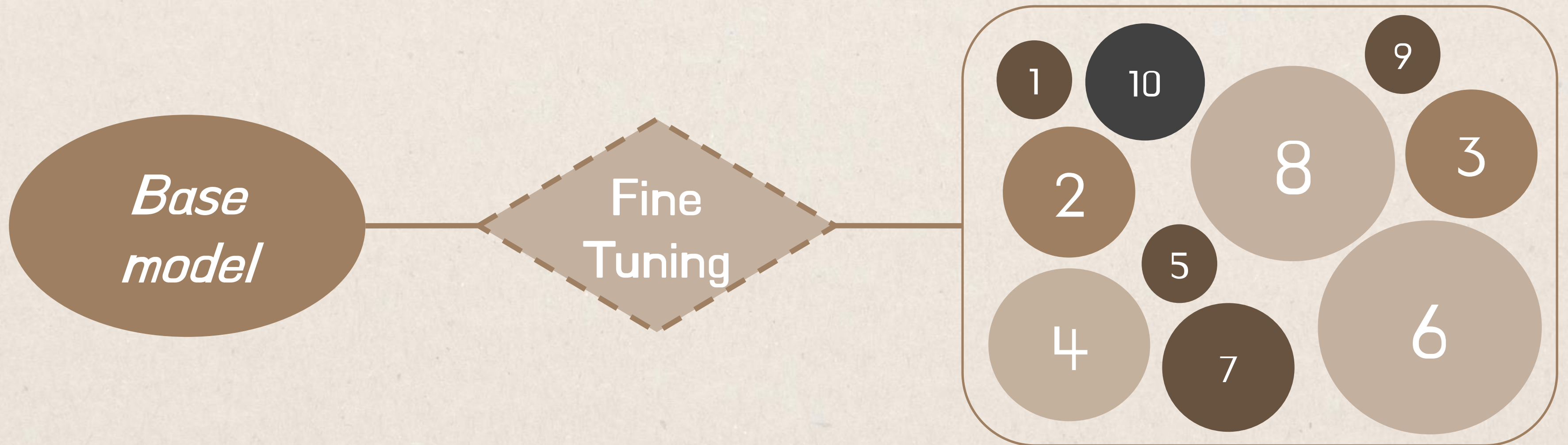
-> 키워드에 따른 데이터 추가 수집 이후 fine-tuning

Fine-tuning Data 수집

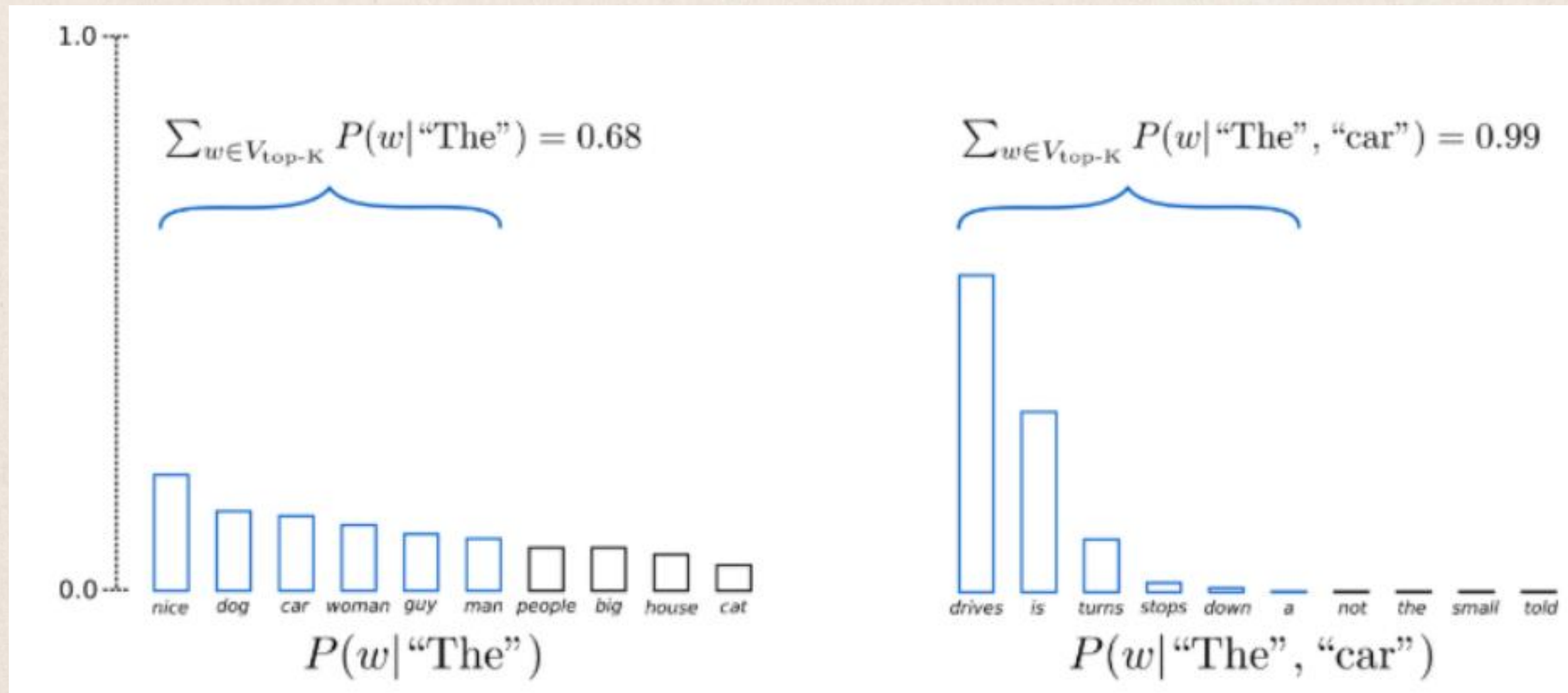
```
# 2. 화면에 나타나는 카테고리 박스를 css를 기준으로 찾아 카테고리의 session 들을 category_boxes 변수에 저장함
search_box = driver.find_element(By.XPATH, "/html/body/div[3]/header/div[1]/div/div[2]/div/button")
search_box.click()
search_box = driver.find_element(By.XPATH, "/html/body/div[2]/div/div[2]/div/div[1]/div/div/div/span/input[1]")
search_keyword = '추억'
search_box.send_keys(search_keyword)
search_box.send_keys(Keys.RETURN)
driver.switch_to.window(driver.window_handles[-1])
```

- 각 에피소드의 키워드를 브런치에 검색하여 해당 키워드를 가지고 있는 수필들을 웹 크롤링을 통해 수집
- 해당 데이터는 전처리 후 Base model에 학습시킴

Fine-tuning을 통해 각 에피소드 별로 모두 10개의 모델 생성



Top P Sampling & Top K Sampling



생성한 수필 소개

제목: 과거에서 찾은 ‘나’

줄거리

지극히 평범한 사람이고 평범한 삶을 사는 ‘나’는 어렸을 때 갔던 전시회에 다시 가보며 과거를 회상한다. 과거에는 잠시나마 평범을 탈출하여 ‘나’를 찾았던 시기가 존재했다. 나를 평범에서 벗어나 담대한 삶을 보여준 ‘그’와 무척이나 가까워졌고 같이 그림에 대한 열정을 불태웠지만 모종의 이유로 그와 멀어지며 나는 다시 평범으로 돌아가게 된다. 어느덧 시간이 흐른 지금 나는 과거를 기억하며 다시 담대한 삶을 살아가고자 한다.

감사합니다.