

# Assignment 6

## 프로젝트 최종 보고서



과목명	영어분석을 위한 기계학습
담당교수	최승택
제출일	2025년 12월 11일
전공	한국어교육과
학번	202302592
이름	임채운

## 1. 프로젝트 개요

### 1.1. 프로젝트명

: 교재 학습활동 추출 도구

### 1.2. 문제 상황

사범대 전공 특성상 교과서나 교재를 보아야 하는 경우가 매우 많다. 교과서의 종류가 과목별, 수준별, 학년별로 전부 다르고 출판사도 다양하여 교재 분석을 하거나 수업 시연 준비를 위해 교과서를 검토하는 데에 시간이 많이 소요된다. 이 수고를 덜기 위해 교과서 안에 어떤 글이 들어있는지, 어떤 질문이 들어있는지 등을 추출하여 보여주는 도구가 필요하다고 생각했다.

## 2. 진행 과정

### 2.1. 주제 설정 및 문제 정의

앞선 문제 상황을 해결하기 위해, 교과서 속 학습활동 질문 추출 도구를 개발하는 것을 이 프로젝트의 주제로 설정하였다. 처음에는 '성취 기준'이나 '작품'을 기준으로 학습활동을 추출하는 것을 목표로 하여 데이터 수집과 분석까지 진행하였으나, 실제 모델을 학습하는 과정이 너무 복잡하고 이 계획대로 프로젝트를 진행하기에는 개인 역량이 부족하다고 판단하여 기준을 설정하지 않고 단순히 지시문만 추출할 수 있는 도구를 개발하는 것으로 계획을 일부 수정하였다.

### 2.2. 데이터 수집 및 분석

#### - 최초 계획에 따른 데이터 수집 및 분석

데이터 수집과 분석 과정에서는 최초 계획에 따라 Assignment 4에서 '성취 기준'을 기준으로 설정하여 교과서 속 지시문을 수동으로 수집하였다. 수집한 데이터에 대한 개요는 다음과 같다.

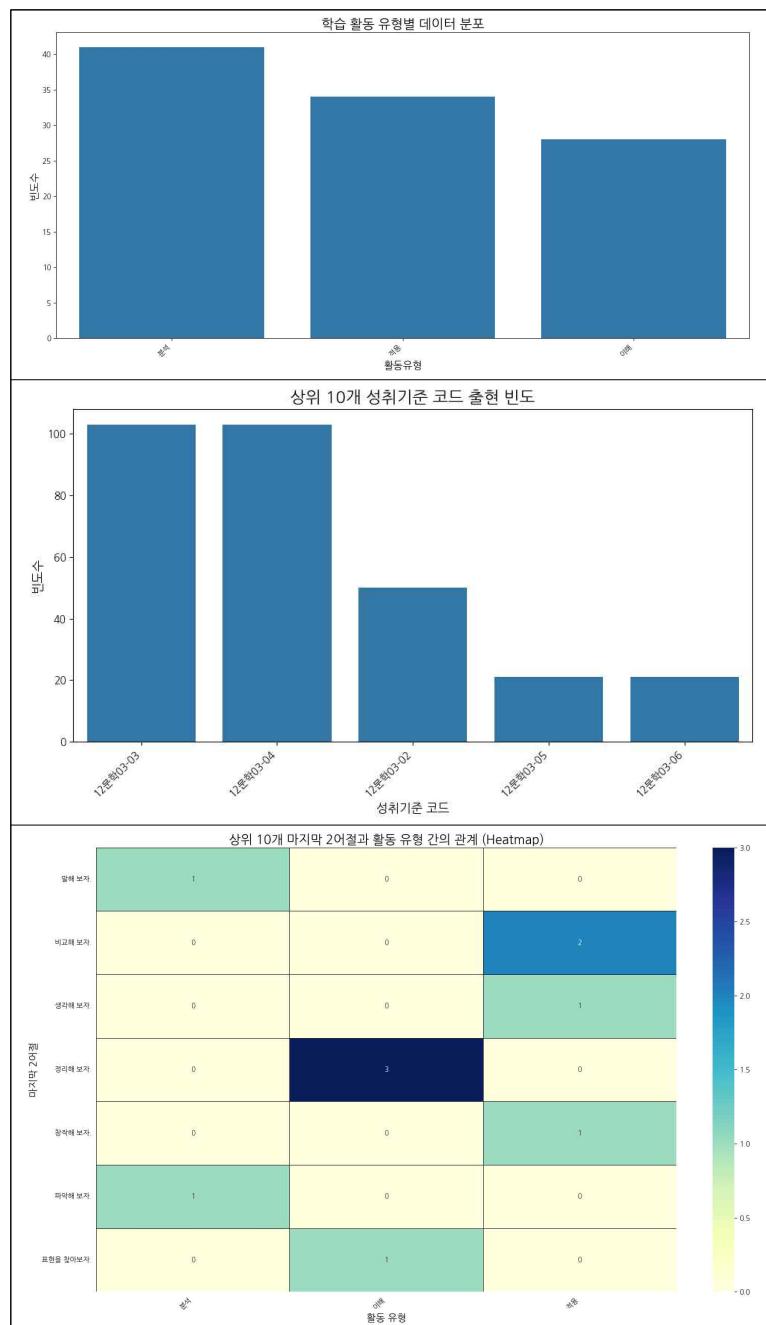
수집 데이터	2015 국어과 교육과정 성취기준, 국어 문학 교과서 (천재교과서, 동아출판, 비상교육, 천재교육)
개수	교육과정 성취기준 16개, 교과서 학습활동 지시문 103개

교육과정 성취기준은 '성취기준 코드 / 영역 / 학년 / 내용'으로 나누어 정리하였고, 교과서 학습활동 지시문은 '출판사 / 단원명 / 페이지 / 코드 / 글감제목 / 글쓴이 / 학습활동 지시문 / 활동 유형 / 질문 형식'으로 나누어 정리하였다. 수집한 자료들은 모두 csv 파일 형태로 저장하였다.

데이터 분석 과정에서 결측값은 존재하지 않았고, 활동 유형별 분포 확인 결과 '분석', '적용', '이해' 순으로 많았던 것을 확인하였다. 이 유형은 학습활동 순서대로 이해, 분석, 적용 영역으로 임의로 부여한 것으로, 라벨링하는 주체에 따라 그 범위가 조금씩 달라질 수 있다. 성취 기준 코드의 출현 빈도 확인 결과 출현 빈도로는 '12문학03-03'과 '12문학03-04'가 가장 많았고, 이 두 항목은 수집된 모든 데이터에 적용된 코드임을 확인하였다. 문학 교과서에

한정되고, 특히 '문학사' 소단원에 한정하여 데이터를 수집하고 정리하였기 때문에 이 결과가 나타난 것으로 보인다. 이에 관해서는 향후 불균형한 데이터를 보완할 필요가 있다. 마지막으로는 학습활동 질문과 활동 유형과의 관계를 찾아보고자 하였다. 이에 따라, '학습활동 질문'의 마지막 2어절을 추출하여 활동 유형 '이해, 분석, 적용'과의 관계를 파악해 보았다. (예시: '말해 보자'로 끝나는 질문은 '이해' 유형에 해당할 것이다) 그러나 분석 결과 두 항목 간의 관계는 나타나지 않았고, 이 질문을 통해 활동 유형을 파악하기는 어려울 것으로 예상되었다.

다음은 데이터 분석과 관련된 그래프다.



수집한 데이터가 문학 교과서의 문학사 부분에만 해당하여 데이터의 불균형이 나타난 것으로 보이므로, 이를 보완하기 위해 국어의 다른 세부 과목 데이터나 다른 소단원의 데이터를 더 수집하고자 하였다.

#### - 수정된 계획에 따른 데이터 수집

계획이 수정됨에 따라 불필요한 성취 기준 데이터는 폐기하고, 기존에 수집한 학습활동 지시문만 채택하여 다시 정리하였다. 이 과정에서 데이터를 조금 더 추가하여 지시문 데이터를 총 115개를 수집하였다. 지시문이 아닌 일반 문장의 경우, 매우 정제된 패턴을 보이는 교과서의 문장에서 벗어나서 더 다양한 형태의 문장을 수집하기 위해 한국어 위키피디아에서 193개의 문장을 수집하였다. 지시문은 '1'을 할당하였고, 일반 문장은 '0'을 할당하여 정리한 자료는 csv 파일 형태로 저장하였다. 다음은 라벨링한 데이터의 예시다.

sentence	label
단군의 탄생 과정을 중심으로 '변화'라는 관점에서 살펴보고 빈칸을 채워 보자.	1
이 대회의 예선 방식은 조별 리그와 플레이오프로 되어 있다.	0

계획 수정에 따라 시간 관계상 이 데이터에 대해서는 분석 절차를 생략하였다.

### 2.3. ML 모델 학습 및 평가

한국어 데이터 처리를 위해 KoBERT 기반 정보추출 모델(skt/kobert-base-v1)을 활용하여 Assignment 4에서 수집한 데이터를 바탕으로 모델을 학습하였다. 다음은 학습 과정에서 분할한 데이터의 개수다.

분류	개수(개)
Train	196
Validation	50
Test	62

이를 바탕으로 학습을 시작하였고, 최초 epoch을 3으로 하여 실행하였으나, evaluation과 inference 성능이 너무 좋지 않아 epoch의 수를 10으로 늘렸다. epoch을 3에서 10으로 점차 늘린 결과, epoch의 수를 늘릴수록 validation accuracy가 항상 높아지는 것은 아님을 발견하였다. 그러나 전반적으로 우상향하는 추세를 보이고 있고, 더불어 training loss가 함께 줄어들고 있는 모습을 볼 수 있었다. epoch 10일 때의 주요 성능지표를 살펴보면, 지시문 추출 기준으로 Accuracy가 0.8, Precision 1.0, Recall 0.65, F1-score 0.79였다.

epoch을 늘릴수록 성능은 개선되었지만, Recall 수치가 너무 낮았다. 이 도구의 활용 목적이 '지시문 추출 후 검토'이므로 지시문이 아닌 문장을 추출하더라도 지시문인 문장을 빠짐없이 추출하도록 하고자 Recall 수치를 높이는 방식으로 조정하였다. 이에 따라, learning rate 를 2e-5에서 3e-5로 조정하여 모델 학습을 진행하였고, recall 수치가 0.77, F1-score가 0.87로 개선되었다. 다음 표는 모델 학습 및 평가에 관한 요약이다.

--- 3단계: Test set 최종 평가 수행 --- 평가 지표 (Precision, Recall, F1-score, Support) :				
	precision	recall	f1-score	support
0	0.8000	1.0000	0.8889	36
1	1.0000	0.6538	0.7907	26
accuracy		0.8548	0.8548	62
macro avg	0.9000	0.8269	0.8398	62
weighted avg	0.8839	0.8548	0.8477	62

=====

<learning rate 조정 전>

--- 3단계: Test set 최종 평가 수행 --- 평가 지표 (Precision, Recall, F1-score, Support) :				
	precision	recall	f1-score	support
0	0.8571	1.0000	0.9231	36
1	1.0000	0.7692	0.8696	26
accuracy		0.9032	0.9032	62
macro avg	0.9286	0.8846	0.8963	62
weighted avg	0.9171	0.9032	0.9006	62

=====

<learning rate 조정 후 최종>

### 3. 모델을 서비스로 만든 구조

#### 3.1. 서비스화

학습을 완료하고 실제 사용이 가능하도록 CLI 기반 도구로 만들었다. txb\_inference.py 파일이 핵심 로직이며, 모델 가중치는 instruction\_classifier\_model 폴더에 저장하였고, 모델 서비스에 필요한 라이브러리 목록을 requirements.txt에 저장하였다.

python 기반 서비스이기 때문에 이 서비스를 사용하기 위해서는 python이 설치된 환경이 필요하다. 또한, txt 파일을 입력으로 받기 때문에, txt 형태가 아닌 다른 형태의 파일이라면 미리 변환이 필요하다.

#### 3.2. 시스템 구조

- 1) CMD나 Anaconda prompt에서 사용자로부터 txt 파일의 경로를 입력받는다.

실행형식:

```
python      txb_inference.py      input_data\{파일명}.txt      -model-dir
instruction_classifier_model
```

- 2) 입력받은 파일에서 모델이 사전 학습된 ‘지시문’만 추출한다.
- 3) 추출된 지시문들이 콘솔 창을 통해 사용자에게 제시된다.

### 4. 실제 사용 결과

#### 4.1. 5회 이상 사용 기록 (스크린샷)

- 1) <국어논리및논술> 국어 교과서 제작에 필요한 교재 분석

<국어논리및논술> 과목에서 논리 교과서 제작을 하였다. 이를 위해 중학교 국어 교과서 속 논리 논술 단원을 분석하였고, 그 과정에서 이 도구를 활용하여 교과서 5종의 학습활동 속 지시문을 추출하여 교과서 제작에 참고하였다. OCR을 활용하여 미리 교과서에서 수동 추출한 텍스트를 txt 파일로 저장하여 도구에 제공하였고, 그 파일을 바탕으로 도구가 추출한 문장들을 출력하였다. 전반적으로 지시문을 거의 빠짐없이 추출하는 편이었으나, 불필요한 단어나 지시문이 아닌 문장들을 함께 추출하는 모습을 보였다. 다음은 실제 도구를 사용한 스크린샷 기록이다.

(kobert\_final\_env) C:\final>python txb\_inference.py input\_data\2015\_miraen.txt --model\_dir instruction\_classifier\_model  
모델 로드 중: instruction\_classifier\_model  
모델 로드 성공.

[ 최종 추출 결과 ]  
총 문장 수: 186  
추출된 시지문 수: 24

1. 2 자신이 노을이라면 어떤 방법으로 이슬이를 살득할 수 있을  
지 생각해 보자.  
2. (1) 이 글을 다음과 같이 정리할 때, 빈칸에 들어갈 알맞은 말  
을 써 보자.  
3. 애나히면,  
(3) 위 말동을 바탕으로 하여 이 글의 구조를 파악해 보자.  
4. 이해 탐구  
내 인생 벼슬  
2 나음을 떠올리며 쓰임 논증 방법을 파악해 보자.  
5. (1) 이 글은 이가 주장하는 바가 무엇인지 말해 보자.  
6. (2) 다음 설명을 참고하여 이 글의 논증 방법과 구조를 파악해  
보자.  
7. 논증 방법을 중심으로 글의 구조를 알아보고, 논증의 타당성  
을 판단해 보자.  
8. (1) 이 글의 글쓴이가 주장하는 바가 무엇인지 말해 보자.  
9. (2) 다음 설명을 참고하여 이 글의 논증 방법과 구조를 파악해  
보자.  
10. 논증 방법을 중심으로 글의 구조를 알아보고, 논증의 타당성  
을 판단해 보자.  
11. 또한, 논증에서 근거를 '전제'라고도 하고, 주장을 '결론'이

10. 논증 방법을 중심으로 글의 구조를 알아보고, 논증의 타당성  
을 판단해 보자.  
11. 또한, 논증에서 근거를 '전제'라고도 하고, 주장을 '결론'이  
라고도 한다는 점을 다시 한번 떠올려 보세요.  
12. ○ 이 글이 논증의 타당성을 판단할 때  
(3) 이 글에서 주장할 평가는 방식이 타당한지 판단해 보자.  
13. (1) 논증방법 파악하며 읽기 183  
4 단원

문제 해결  
적용  
○ 이글기  
논증방법을 사용하여 주장 할 때는 전제 가운데 일부가 생략되기도 해요.  
14. 살득의 힘  
1 다음 광고에 쓰인 논증 방법을 파악해 보자.  
2 정민숙에서 모둠원과 함께 논증 방법을 사용하여 이야기를 나누어  
보자.  
16. 연가 활동으로  
좋은 영화를 보자.  
17. 고심 우리 가 생활하는 중요한 공간이다.  
18. 살득의 힘

종은 영화를 보자.  
17. 교실은 우리가 생활하는 중요한 공간이다.  
18. 살득의 힘  
배운 내용을 떠올리며 빈칸에 들어갈 알맞은 말을 넣어 보자.  
19. (1)  
온/는 일반적 원리나 법칙에서 구체적 사실을 이끌어 내는 논증 방  
법으로, 잠깐 대전제를 사용하는 삼단 논법이 대표적이다.  
20. (3) 둘 이상의 대상이 비슷한 속성을 가진다는 것을 근거로  
다른 속성도 유사 할 것이라고 추론하는 것을 (이)라고 한다.  
21. 2 다음을 읽고 맞는 말은 아래, 틀린 말은 X에 표시해 보자.  
22. (1) 논증이란 근거를 들어 주장이 참임을 증명하는 것으로,  
근거가 참이면 주장은 언제나 참이다.  
23. (Q, X)  
(3) 논증 방법을 파악하며 글을 읽으면 글의 타당성을 판단하는 대  
도움이 된다.  
24. 내가 확실하게 믿고 있던 것이 나만의 편안 편견은 아닌지 의  
심 해 본다.

1. 정윤이가 발견한 문제 상황을 써 보자.  
2. 개념 쪽  
국정윤이와 친구들이 쓸 글의 주장과 이유, 예상 독자를 정리해 보자.  
3. 그럼 우리가 쓸 글의 주장과 주장을 뒷받침하는 이유를 정리해  
보자.  
4. 정윤이와 친구들이 수집한 자료를 살펴보며, 자료를 수집할 때  
유의사항 점을 생각해 보자.  
5. 험금(출판) 물질이 어떤 성분  
을 포함하고 있는지 분석  
6. 카페인은 청소년  
의 성장에 필요한 키클리과 철의 흡  
수를 방해하고 몸밖으로 배출시  
켜 깔 숨, 철의 결  
3 핵 문제 가 생  
길 수 있다.  
7. .... 카페인은 신경을 예민하게 만들기  
도 합니 다. - 머리 카펜터, 『카페인 권하는 사회』에서  
자

19. (1) 주장하는 글의 내용을 조직하기 위한 개요를 작성해 보자.  
20. (2) 평가 기준에 따라 개요를 점검해 보자.  
21. 개요  
서론  
본론  
(2) 다음 평가 기준을 바탕으로 하여 수집한 자료의 타당성을 점검하  
고, 자료 활용  
방안을 정리해 보자.  
22. (3) 예제로 정리한 자료 활용 방안을 바탕으로 하여 타당한 근거  
를 마련해 보자.  
23. 6 적성한 글을 평가 기준에 따라 점검하고, 이를 바탕으로 하  
여 글을 써 보자.  
24. 7 예상 독자를 고려하여 완성한 글을 친구들과 공유해 보자.  
25. (3)  
다음 단어의 뜻을 보고, 보기에서 알맞은 단어를 찾아 빈칸에 써  
보자.  
26. C  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
698  
699  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
798  
799  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
898  
899  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
989  
990  
991  
992  
993  
994  
995  
996  
997  
997  
998  
999  
999  
1000  
1001  
1002  
1003  
1004  
1005  
1006  
1007  
1008  
1009  
1009  
1010  
1011  
1012  
1013  
1014  
1015  
1016  
1017  
1018  
1019  
1019  
1020  
1021  
1022  
1023  
1024  
1025  
1026  
1027  
1028  
1029  
1029  
1030  
1031  
1032  
1033  
1034  
1035  
1036  
1037  
1038  
1039  
1039  
1040  
1041  
1042  
1043  
1044  
1045  
1046  
1047  
1048  
1049  
1049  
1050  
1051  
1052  
1053  
1054  
1055  
1056  
1057  
1058  
1059  
1059  
1060  
1061  
1062  
1063  
1064  
1065  
1066  
1067  
1068  
1069  
1069  
1070  
1071  
1072  
1073  
1074  
1075  
1076  
1077  
1078  
1079  
1079  
1080  
1081  
1082  
1083  
1084  
1085  
1086  
1087  
1088  
1089  
1089  
1090  
1091  
1092  
1093  
1094  
1095  
1096  
1097  
1097  
1098  
1099  
1099  
1100  
1101  
1102  
1103  
1104  
1105  
1106  
1107  
1108  
1109  
1109  
1110  
1111  
1112  
1113  
1114  
1115  
1116  
1117  
1118  
1119  
1119  
1120  
1121  
1122  
1123  
1124  
1125  
1126  
1127  
1128  
1129  
1129  
1130  
1131  
1132  
1133  
1134  
1135  
1136  
1137  
1138  
1139  
1139  
1140  
1141  
1142  
1143  
1144  
1145  
1146  
1147  
1148  
1149  
1149  
1150  
1151  
1152  
1153  
1154  
1155  
1156  
1157  
1158  
1159  
1159  
1160  
1161  
1162  
1163  
1164  
1165  
1166  
1167  
1168  
1169  
1169  
1170  
1171  
1172  
1173  
1174  
1175  
1176  
1177  
1178  
1179  
1179  
1180  
1181  
1182  
1183  
1184  
1185  
1186  
1187  
1188  
1189  
1189  
1190  
1191  
1192  
1193  
1194  
1195  
1196  
1197  
1197  
1198  
1199  
1199  
1200  
1201  
1202  
1203  
1204  
1205  
1206  
1207  
1208  
1209  
1209  
1210  
1211  
1212  
1213  
1214  
1215  
1216  
1217  
1218  
1219  
1219  
1220  
1221  
1222  
1223  
1224  
1225  
1226  
1227  
1228  
1229  
1229  
1230  
1231  
1232  
1233  
1234  
1235  
1236  
1237  
1238  
1239  
1239  
1240  
1241  
1242  
1243  
1244  
1245  
1246  
1247  
1248  
1249  
1249  
1250  
1251  
1252  
1253  
1254  
1255  
1256  
1257  
1258  
1259  
1259  
1260  
1261  
1262  
1263  
1264  
1265  
1266  
1267  
1268  
1269  
1269  
1270  
1271  
1272  
1273  
1274  
1275  
1276  
1277  
1278  
1279  
1279  
1280  
1281  
1282  
1283  
1284  
1285  
1286  
1287  
1288  
1289  
1289  
1290  
1291  
1292  
1293  
1294  
1295  
1296  
1297  
1297  
1298  
1299  
1299  
1300  
1301  
1302  
1303  
1304  
1305  
1306  
1307  
1308  
1309  
1309  
1310  
1311  
1312  
1313  
1314  
1315  
1316  
1317  
1318  
1319  
1319  
1320  
1321  
1322  
1323  
1324  
1325  
1326  
1327  
1328  
1329  
1329  
1330  
1331  
1332  
1333  
1334  
1335  
1336  
1337  
1338  
1339  
1339  
1340  
1341  
1342  
1343  
1344  
1345  
1346  
1347  
1348  
1349  
1349  
1350  
1351  
1352  
1353  
1354  
1355  
1356  
1357  
1358  
1359  
1359  
1360  
1361  
1362  
1363  
1364  
1365  
1366  
1367  
1368  
1369  
1369  
1370  
1371  
1372  
1373  
1374  
1375  
1376  
1377  
1378  
1379  
1379  
1380  
1381  
1382  
1383  
1384  
1385  
1386  
1387  
1388  
1389  
1389  
1390  
1391  
1392  
1393  
1394  
1395  
1396  
1397  
1397  
1398  
1399  
1399  
1400  
1401  
1402  
1403  
1404  
1405  
1406  
1407  
1408  
1409  
1409  
1410  
1411  
1412  
1413  
1414  
1415  
1416  
1417  
1418  
1419  
1419  
1420  
1421  
1422  
1423  
1424  
1425  
1426  
1427  
1428  
1429  
1429  
1430  
1431  
1432  
1433  
1434  
1435  
1436  
1437  
1438  
1439  
1439  
1440  
1441  
1442  
1443  
1444  
1445  
1446  
1447  
1448  
1449  
1449  
1450  
1451  
1452  
1453  
1454  
1455  
1456  
1457  
1458  
1459  
1459  
1460  
1461  
1462  
1463  
1464  
1465  
1466  
1467  
1468  
1469  
1469  
1470  
1471  
1472  
1473  
1474  
1475  
1476  
1477  
1478  
1479  
1479  
1480  
1481  
1482  
1483  
1484  
1485  
1486  
1487  
1488  
1489  
1489  
1490  
1491  
1492  
1493  
1494  
1495  
1496  
1497  
1497  
1498  
1499  
1499  
1500  
1501  
1502  
1503  
1504  
1505  
1506  
1507  
1508  
1509  
1509  
1510  
1511  
1512  
1513  
1514  
1515  
1516  
1517  
1518  
1519  
1519  
1520  
1521  
1522  
1523  
1524  
1525  
1526  
1527  
1528  
1529  
1529  
1530  
1531  
1532  
1533  
1534  
1535  
1536  
1537  
1538  
1539  
1539  
1540  
1541  
1542  
1543  
1544  
1545  
1546  
1547  
1548  
1549  
1549  
1550  
1551  
1552  
1553  
1554  
1555  
1556  
1557  
1558  
1559  
1559  
1560  
1561  
1562  
1563  
1564  
1565  
1566  
1567  
1568  
1569  
1569  
1570  
1571  
1572  
1573  
1574  
1575  
1576  
1577  
1578  
1579  
1579  
1580  
1581  
1582  
1583  
1584  
1585  
1586  
1587  
1588  
1589  
1589  
1590  
1591  
1592  
1593  
1594  
1595  
1596  
1597  
1597  
1598  
1599  
1599  
1600  
1601  
1602  
1603  
1604  
1605  
1606  
1607  
1608  
1609  
1609  
1610  
1611  
1612  
1613  
1614  
1615  
1616  
1617  
1618  
1619  
1619  
1620  
1621  
1622  
1623  
1624  
1625  
1626  
1627  
1628  
1629  
1629  
1630  
1631  
1632  
1633  
1634  
1

(kobert\_final\_env) C:\final>python txb\_inference.py input\_data\2022\_visanparkn.txt --model\_dir instruction\_classifier\_model  
모델 로드 중: instruction\_classifier\_model  
모델 로드 성공.

[ 최종 추출 결과 ]  
총 문장 수: 213  
추출된 시시문 수: 21

1. 노란 별 아리와 하얀 새는  
'승전'이라는 말을 서로 다른  
2. (2) 이기거나에서 여학생이 남학생의 말을 각각 어떠한 의미로 추  
론하였는지 그 근거와 함께 써 보자.  
3. 그리고 화자의 의도를 적절하게 추론한 것을 골라 보자.  
4. (1) "지금 몇 시니?"라는 담화가 이루어지는 상황을 파악해 보자.  
5. (2) 나에게 여학생이 남학생의 말을 각각 어떠한 의미로 추  
론하였는지 그 근거와 함께 써 보자.  
6. 그리고 화자의 의도를 적절하게 추론한 것을 골라 보자.  
7. (1) "지금 몇 시니?"라는 담화가 이루어지는 상황을 파악해 보자.  
8. 2. '중화' 민서의  
대화, 음원  
괜찮아, 좀 하, 너 나한테 하고 싶은 말이 있는 표정인걸?  
9. (1) 상황을 고려하여 '고'에 담긴 화자의 의도를 추론해 보자.  
10. (2) 안전모 의무화 규정과 관련한 기자 의 말에서 추론할 수  
있는 대안은 무엇 일까?  
11. (3) 통계 자료와 전문가의 인터뷰를 활용하여 기자가 말하고  
자 한 바는 무엇 일까?

(kobert\_final\_env) C:\final>python txb\_inference.py input\_data\2022\_visanparky.txt --model\_dir instruction\_classifier\_model  
모델 로드 중: instruction\_classifier\_model  
모델 로드 성공.

[ 최종 추출 결과 ]  
총 문장 수: 202  
추출된 시시문 수: 23

1. 이 친구들이 하는 말 중 누구의 말이 가장 설득력 있게 들리는  
지 이유와 함께 말해  
보자.  
2. 내 생각은  
준희  
자신에게 해당하는 것을 고르고, 자신의 학습 출발점을 확인해 보자.  
3. 이 단원의 학습 흐름을 살펴보고 스스로 학습을 계획해 보자.  
4. '신비'의 하루  
제시원 문제 상황을 바탕으로 '신비'네 모둠의 글쓰기 계획을 파악  
해 보자.  
5. 운동선  
이왕 운동하는 거 잘 써서 학교 신문에도 투고해 보자.  
6. 각 자료를  
의 내용을 파악하고, 근거로 쓸 내용을 마련해 보자.  
7. 그리고 우리가 원활히 매번 언제든 우리를 전 세계 와 연결해 준  
다.  
8. (1) 가사를 주장하는 글의 근거로 활용할 수 있을지 평가해 보자.  
9. (2) 앞의 (1)에서 평가한 자료를 바탕으로 주장의 이유를 정리  
해 보자.  
10. 앞의 3에서 작성한 개요를 바탕으로 주장하는 글을 쓰고, 글  
을 점검하며 고쳐 써 보자.  
11. 문장 등으로 하는 답변.  
12. 고쳐쓰기  
타당하다  
쓴 글을 을 찾아 고쳐 씀.  
13. 점검  
학습 내용을 스스로 점검해 보고, 더 학습하고 싶다면 발전 활동  
을 해 보자.

(kobert\_final\_env) C:\final>python txb\_inference.py input\_data\2015\_mirra.sin.txt --model\_dir instruction\_classifier\_model  
모델 로드 중: instruction\_classifier\_model  
모델 로드 성공.

[ 최종 추출 결과 ]  
총 문장 수: 252  
추출된 시시문 수: 32

1. 하늘이와 바다의 말에 친구들이 다르게 반응한 까닭을 말해 보자.  
2. 2 다른 사람을 설득할 때 무엇이 필요한지 생각해 보자.  
3. (2) 주장하는 글쓰기 189  
4. 단원  
이해  
친구 니  
1. 문제 상황 발견하기  
동물 사랑 동아리인 민호와 친구들은 동물원에 관한 뉴스 영상을 보며 동물원에 살고 있는 동물의 삶을 살피고자 했지만 영상에 대한 자신의 생각을 말해 보자.  
2. 우리 동아리의 문제에 맞는 동물원에 관한 자신의 생각을 말해 보자.  
3. 세상을 향한 목소리를 이기야 해 보자.  
4. 세상을 향한 목소리를 듣는 동물원 문제에 관한 동아리 친구들의 주장은 말해 보자.  
5. 둘째 친구들이 찾은 자료가 주장은 뒷받침할 근거로 적절  
한지 평가해 보자.

(2) 주장하는 글쓰기 201  
"단원  
문제 해결  
1. 주장하는 글 쓰기  
2. 청소년과 관련한 사회 문제 가운데 하나를 골라 문제 상황을 찾아보자.  
3. 3 주장은 뒷받침할 근거를 마련해 보자.  
4. (2) (1)을 바탕으로 하여 근거를 정리해 보자.  
5. 주제  
서론  
본론  
결론  
4월  
(2) 주장하는 글쓰기 203  
문제 해결  
작용 :  
5 탄당한 근거를 들어 주장하는 글을 써보자.  
1 다음을 위고 맞는 말은 ○에, 둘린 말은 X에 표시해 보자. 넣어 보  
2 배운 내용을 떠올리며 빙간에 들어갈 알맞은 말을 넣어 보자.  
28. 2 배운 내용을 떠올리며 빙간에 들어갈 알맞은 말을 넣어 보자.

대화 ' 음원  
괜찮아, 좀 하, 너 나한테 하고 싶은 말이 있는 표정인걸?  
9. (1) 상황을 고려하여 '고'에 담긴 화자의 의도를 추론해 보자.  
10. (2) 안전모 의무화 규정과 관련한 기자 의 말에서 추론할 수  
있는 대안은 무엇 일까?  
11. (3) 통계 자료와 전문가의 인터뷰를 활용하여 기자가 말하고  
자 한 바는 무엇 일까?  
12. 당시 화자의 상황을 고려하여 '고'의 의미를 추론해 보자.  
13. 화자의 관점  
•빠져 살았습니다."  
조종학교 6학년, 권투연습장의 회원 모집 전단을 보다.  
14. 화자의 관점  
(3) 화자가 두 사진을 비교하여 제시한 의도를 파악해 보자.  
15. (4) 화자가 설득하고 싶은 대상과 궁극적으로 전하려는 바를  
말해 보자.  
16. 다음 광고에 담긴 의도를 추론해 보자.  
17. QR 퀴즈  
(2) 이 광고의 의도를 추론해 보자.  
18. (1) 다음 광고 속 "5조만!"이라는 말의 의미를 추론해 보자.  
19. 다음 광고에 담긴 의도를 추론해 보자.  
20. QR 퀴즈  
(2) 이 광고의 의도를 추론해 보자.  
21. (1) 다음 광고 속 "5조만!"이라는 말의 의미를 추론해 보자.

(kobert\_final\_env) C:\final>python txb\_inference.py input\_data\2015\_mirra.y.txt --model\_dir instruction\_classifier\_model  
모델 로드 중: instruction\_classifier\_model  
모델 로드 성공.

[ 대화 ]  
대화 ' 음원  
9. (1) 상황을 고려하여 '고'에 담긴 화자의 의도를 추론해 보자.  
10. (2) 안전모 의무화 규정과 관련한 기자 의 말에서 추론할 수  
있는 대안은 무엇 일까?  
11. (3) 통계 자료와 전문가의 인터뷰를 활용하여 기자가 말하고  
자 한 바는 무엇 일까?  
12. 당시 화자의 상황을 고려하여 '고'의 의미를 추론해 보자.  
13. 화자의 관점  
•빠져 살았습니다."  
조종학교 6학년, 권투연습장의 회원 모집 전단을 보다.  
14. 화자의 관점  
(3) 화자가 두 사진을 비교하여 제시한 의도를 파악해 보자.  
15. (4) 화자가 설득하고 싶은 대상과 궁극적으로 전하려는 바를  
말해 보자.  
16. 다음 광고에 담긴 의도를 추론해 보자.  
17. QR 퀴즈  
(2) 이 광고의 의도를 추론해 보자.  
18. (1) 다음 광고 속 "5조만!"이라는 말의 의미를 추론해 보자.  
19. 다음 광고에 담긴 의도를 추론해 보자.  
20. QR 퀴즈  
(2) 이 광고의 의도를 추론해 보자.  
21. (1) 다음 광고 속 "5조만!"이라는 말의 의미를 추론해 보자.

(kobert\_final\_env) C:\final>python txb\_inference.py input\_data\2015\_mirra.y.txt --model\_dir instruction\_classifier\_model  
모델 로드 중: instruction\_classifier\_model  
모델 로드 성공.

[ 대화 ]  
대화 ' 음원  
8. (1) 가사를 주장하는 글의 근거로 활용할 수 있을지 평가해 보자.  
9. (2) 앞의 (1)에서 평가한 자료를 바탕으로 주장의 이유를 정리해 보고 글을 쓰기 위해 더 필요한 자료가 있다면 찾으보자.  
10. '신비'의 질문에 답하며 개요를 점검하고 수정해 보자.  
11. '신비'가 모둠 구성원과 함께 글을 점검하고 고쳐 쓰는 과정을 살펴보자.  
12. 주장하는 글쓰기 183  
고쳐 쓰기 전 모둠 구성원이 점검한 내용을 참고하여 글을 고쳐 쓰자.  
(1) '신비'네 모둠 구성원이 점검한 내용을 참고하여 글을 고쳐 쓰자.  
13. (2) 앞의 (1)에서 고쳐 쓴 글을 다음 기준에 따라 평가해 보자.  
14. •개요 쓰기 내용을 마련하여 '서론 본론 결론'의 형식으로 짜임세 있게 개요를 쓴다.  
15. •글쓰기 타당한 이유와 근거를 들고 적절한 표현을 사용하여 설득력 있는 글을 쓴다.  
16. •고쳐쓰기 쓴 글을 점검하고, 수정·보완해야 할 부분을 찾아 고쳐 쓴다.  
17. 1 우리 주변의 문제 상황 중 하나를 고르고, 글의 주제와 목적에 맞는 글을 실을 예제를 정해 보자.  
18. 2 앞의 3에서 정리한 주장과 이유, 근거를 바탕으로 개요를 작성해 보자.  
19. 2 자료를 수집하여 주장은 뒷받침할 이유와 근거를 마련해 보자.  
20. 앞의 3에서 작성한 개요를 바탕으로 주장하는 글을 쓰고, 글을 점검하며 고쳐 써 보자.  
21. 문장 등으로 하는 답변.

(kobert\_final\_env) C:\final>python txb\_inference.py input\_data\2015\_mirra.y.txt --model\_dir instruction\_classifier\_model  
모델 로드 중: instruction\_classifier\_model  
모델 로드 성공.

[ 대화 ]  
9. (2) 주장하는 글쓰기 195  
이해  
와  
답구  
21을 바탕으로 하여 주장은 뒷받침할 근거를 정리해 보자.  
10. 근거  
2  
3 동아리 친구들이 찾은 자료 외에 다른 자료를 더 찾아 근거를 마련해 보자.  
11. 1 다음은 평가 기준에 따라 개요를 점검해 보자.  
12. 2 개요를 살펴보며 이 글에서 사용할 논증방법을 생각해 보자.  
13. 글 전체 내용을 포함하고 주장이 분명하게 드러나는 제목으로 바꾸자.  
14. 즐고 제한된 공간에서 살아야 하며, 만들어 놓은 환경도 자연과 다르다.  
15. 그뿐만 아니라 즐은 우리에서 관광객의 시선에 시달리며 극심한 스트레스를 받는다.  
16. (2) 주장하는 글쓰기 199  
이해 탁구  
1. 문제 상황 발견하기  
동물 사랑 동아리인 민호와 친구들은 동물원에 관한 뉴스 영상을 보며 동물원에 살고 있는 동물의 삶을 살피고자 했지만 영상에 대한 자신의 생각을 말해 보자.  
2. 우리 동아리의 문제에 맞는 동아리 친구들의 주장은 말해 보자.  
3. 세상을 향한 목소리를 듣는 동물원 문제에 관한 동아리 친구들의 주장은 말해 보자.  
4. 둘째 친구들이 찾은 자료가 주장은 뒷받침할 근거로 적절한지 평가해 보자.

(2) 주장하는 글쓰기 203  
문제 해결  
작용 :  
5 탄당한 근거를 들어 주장하는 글을 써보자.  
1 다음을 위고 맞는 말은 ○에, 둘린 말은 X에 표시해 보자.  
28. 2 배운 내용을 떠올리며 빙간에 들어갈 알맞은 말을 넣어 보자.  
2) Y (2) (1) O(E) O(?) X(1) O  
이해 탁구  
1 다음은 평가 기준에 어울리는 속담을 찾아 바르게 연결해 보자.  
31. (3)  
"분명히 처음에는 크고 화려한 궁전을 만들려고 했는데 말이야....  
가전을 했더니 그려 려다가 강아지를 그린다.  
32. 2) 키를 들을 때마다 아나도 못 접는다.

## 2) 한국어 문화 수업 자료 만들기를 위한 교재 분석

위의 과정에서 학습 데이터에 등장했거나 그와 비슷한 표현을 찾아냈던 것을 볼 수 있었는데, 눈여겨볼 것은 ‘-하자’ 형태의 지시문이 아닌 ‘-무엇일까?’로 끝나는 지시문을 찾아낸 것이다. 혹시 ‘무엇’과 같은 다른 질문 패턴을 학습하였는지 궁금증이 생겨 다른 교재에도 적용해 보았다. 학습 데이터의 거의 모든 지시문이 ‘-자’ 형태였기 때문에 추출할 수 있는 문장은 없을 것이라 예상했지만, 확인된 결과에 따라 모델을 어떻게 개선하면 좋을지 파악할 수 있을 것 같아 시도해 보았다. <한국문화의 이해> 과목에서 한국어 문화 자료를 만들어야 했고, 이 때 한국어 교재의 문화 파트를 분석하면서 도구를 사용해 보았다.

The image shows two side-by-side terminal windows. Both windows have identical command-line inputs at the top:

```
(kobert_final_env) C:\final>python txb_inference.py input_data\sejong1.txt --model_dir instruction_classifier_model
```

The left window's output is:

```
모델 로드 중: instruction_classifier_model
✓ 모델 로드 성공.

[ 최종 추출 결과 ]
총 문장 수: 113
추출된 지시문 수: 0
-----
추출된 지시문이 없습니다.
```

The right window's output is:

```
모델 로드 중: instruction_classifier_model
✓ 모델 로드 성공.

[ 최종 추출 결과 ]
총 문장 수: 35
추출된 지시문 수: 0
-----
추출된 지시문이 없습니다.
```

예상대로 추출된 지시문은 없었음을 확인하였다. 한국어 교재 속 지시문을 수집하여 학습 데이터로 활용한다면 한국어 교재에도 활용할 수 있을 것으로 보인다.

## 3) 그 외 활용 가능성

이 도구를 주로 이번 학기에 수강한 강의에서 사용하기는 했으나, 앞으로도 매 학기 전공 수업에서도 계속해서 활용할 수 있을 것으로 예상되며, 성능을 더 개선한다면 실제 교육 현장에서도 유용하게 사용할 수 있을 것으로 예상된다.

## 5. 결론

### 5.1. 배운 점

- 프로젝트에 활용할 모델을 선정하는 과정에서 다양한 모델의 역할과 목적을 알게 되었다. 이전까지는 BERT라는 모델에 대해 기본적인 내용만 알고 있었고, 실제로 사용해 본 적이 없었는데, 이번 프로젝트를 진행하면서 BERT뿐만 아니라 KoBERT라는 모델이 존재한다는 것을 알게 되었고, 실제로 활용해 보면서 모델의 기능을 알게 되었다. 또한, 어떤 모델을 활용할지 찾아보고 선정하는 과정에서 그 밖의 다양한 모델에 대해서도 많이 알게 된 시간이었다.
- 간단한 도구를 만든다고 해도 수집해야 할 데이터의 규모나 형식, 전처리 등 고려해야 할 것이 매우 많고 복잡하다고 느꼈다. 특히, 전처리만 잘 되어도 모델의 성능에 크게 영향을 준다는 것을 깨달았고, 그만큼 그 중요성을 크게 체감할 수 있었다.

### 5.2. 개선 방향

- 1) pdf 파일을 txt파일로 수동으로 변환하는 과정을 자동화한다.

프로젝트를 진행하면서 가장 생각지도 못했던 난관은 바로 모델에 입력하는 자료의 형태였다. 교과서 파일이 일반적인 pdf로 구성되어 있다고 생각하고 프로젝트를 계획하였지만, 모델

학습 과정에서 교과서 파일은 사진을 pdf로 변환한 형태로 제공되거나, e-book 형태로 배포되고 있었다는 것을 깨닫게 되었다. 이 때문에 OCR을 활용하여 직접 수동으로 텍스트를 추출하여 txt 파일로 저장해서 모델에 입력할 수밖에 없었는데, 이 과정 때문에 오히려 시간이 더 많이 소요되기도 했다. 이 도구가 계속해서 사용되려면 이 번거로운 과정에 대한 개선이 꼭 필요하다. OCR 과정을 모델에 넣거나, 그와 비슷한 수준의 자동화 과정이 필요하다.

2) 데이터를 더 정교하게 전처리한다.

실행 결과를 보면, 띠어쓰기의 문제로 출력된 결과가 정돈되지 않은 형태를 보이고 있고, 단원의 제목과 같은 불필요한 텍스트가 함께 출력되고 있다. 띠어쓰기나 문장부호를 처리하고 불필요한 표현은 사전에 없애서 정제된 데이터를 만든다.

3) 학습 데이터의 양과 종류를 확대한다.

학습 데이터의 양을 늘려 국어 교과서에서 한국어 교재나 다른 교재로 확장한다면 더 광범위한 활용이 가능할 것으로 예상된다.