인공지능기초응용 II

5 주차 과제

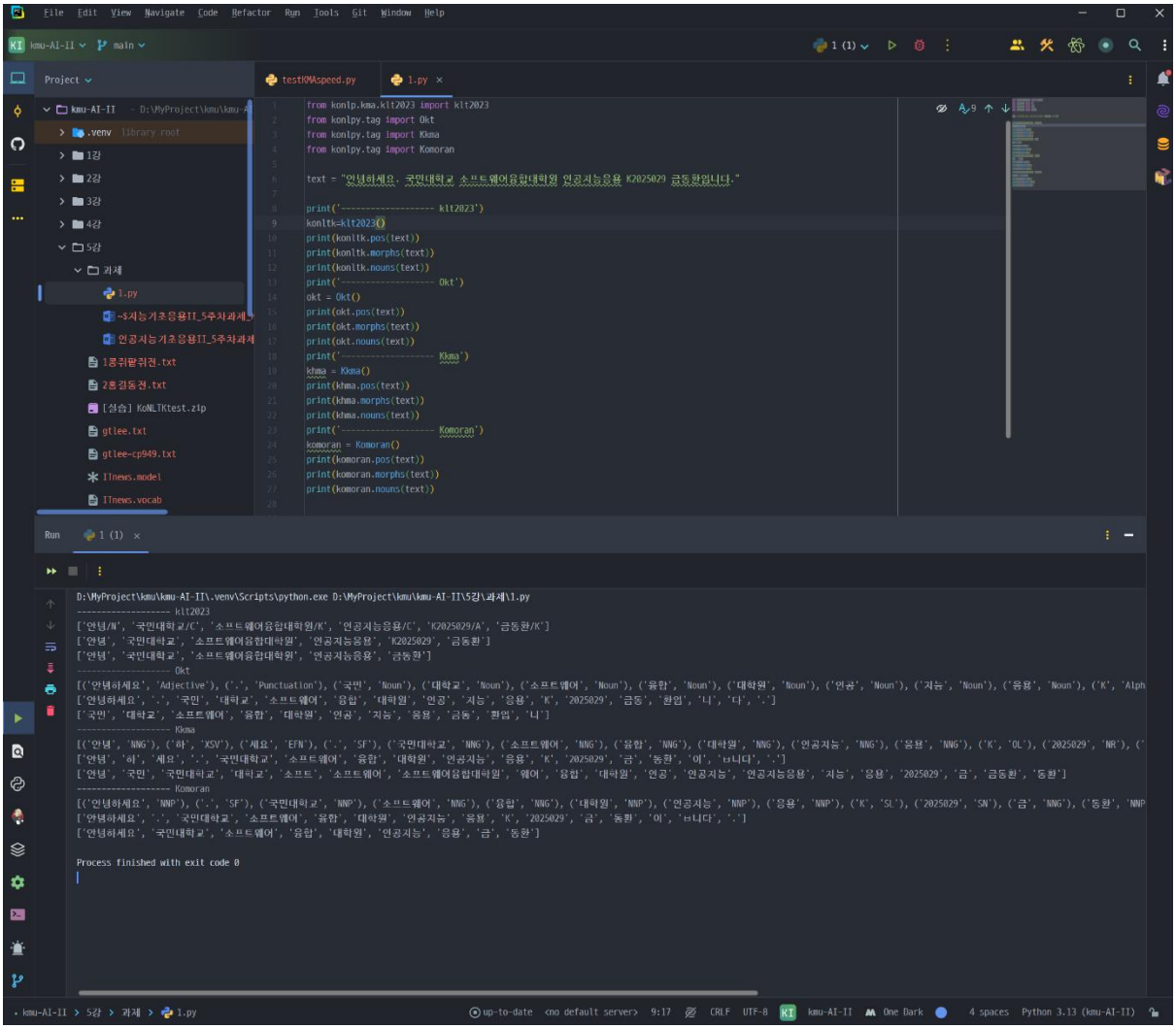
인공지능응용

K2025029 금동환

목차

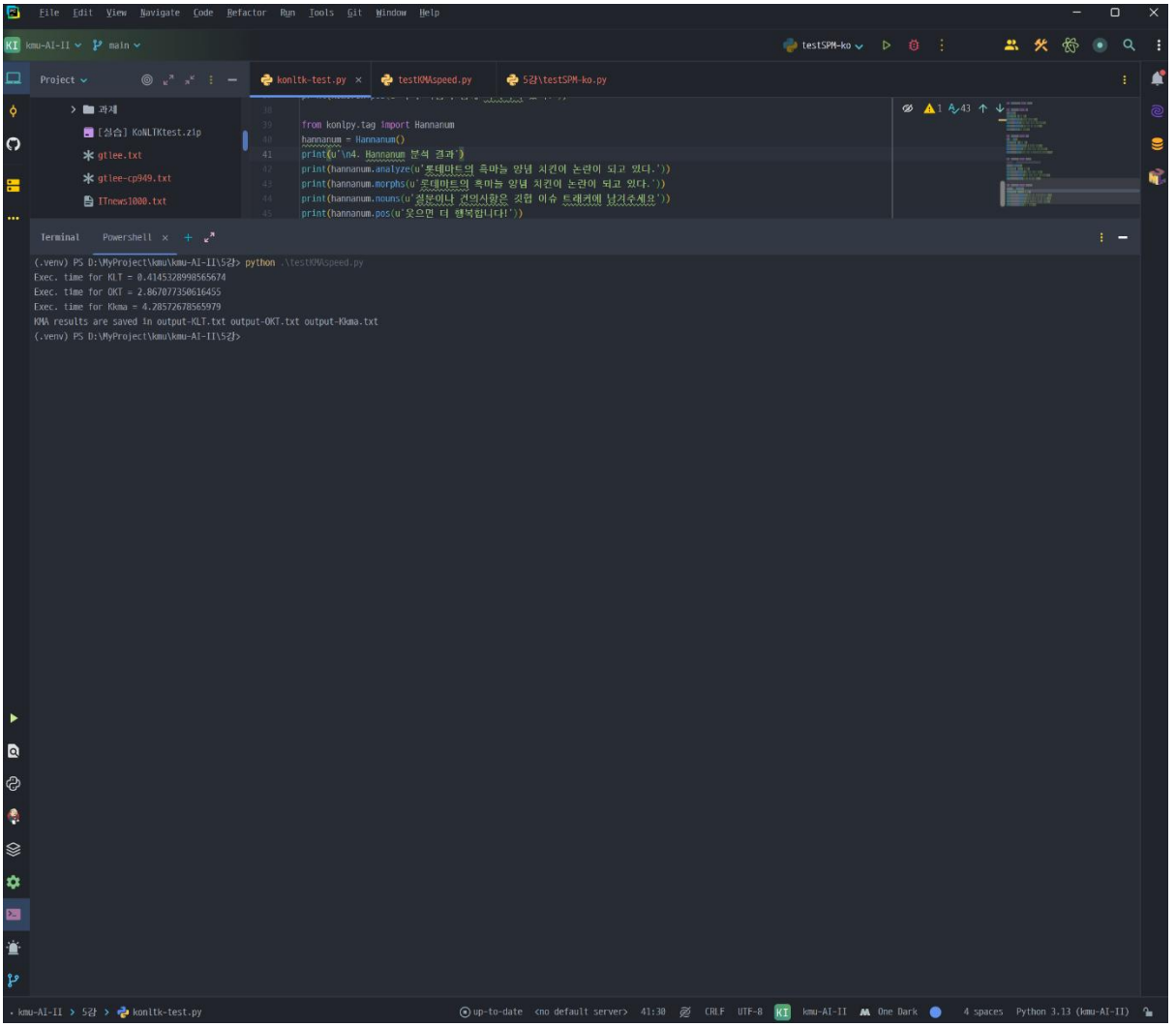
1. 파이썬 형태소 분석기(KoNLTK와 KoNLPy) 실습
2. "[실습] KoNLTKtest.zip"의 파이썬 코드 3개
3. [과제1]에서 준비한 텍스트 파일에 대해 파이썬/윈도용 형태소 분석기 실습

1. 파이썬 형태소 분석기(KoNLTK와 KoNLPy) 실습

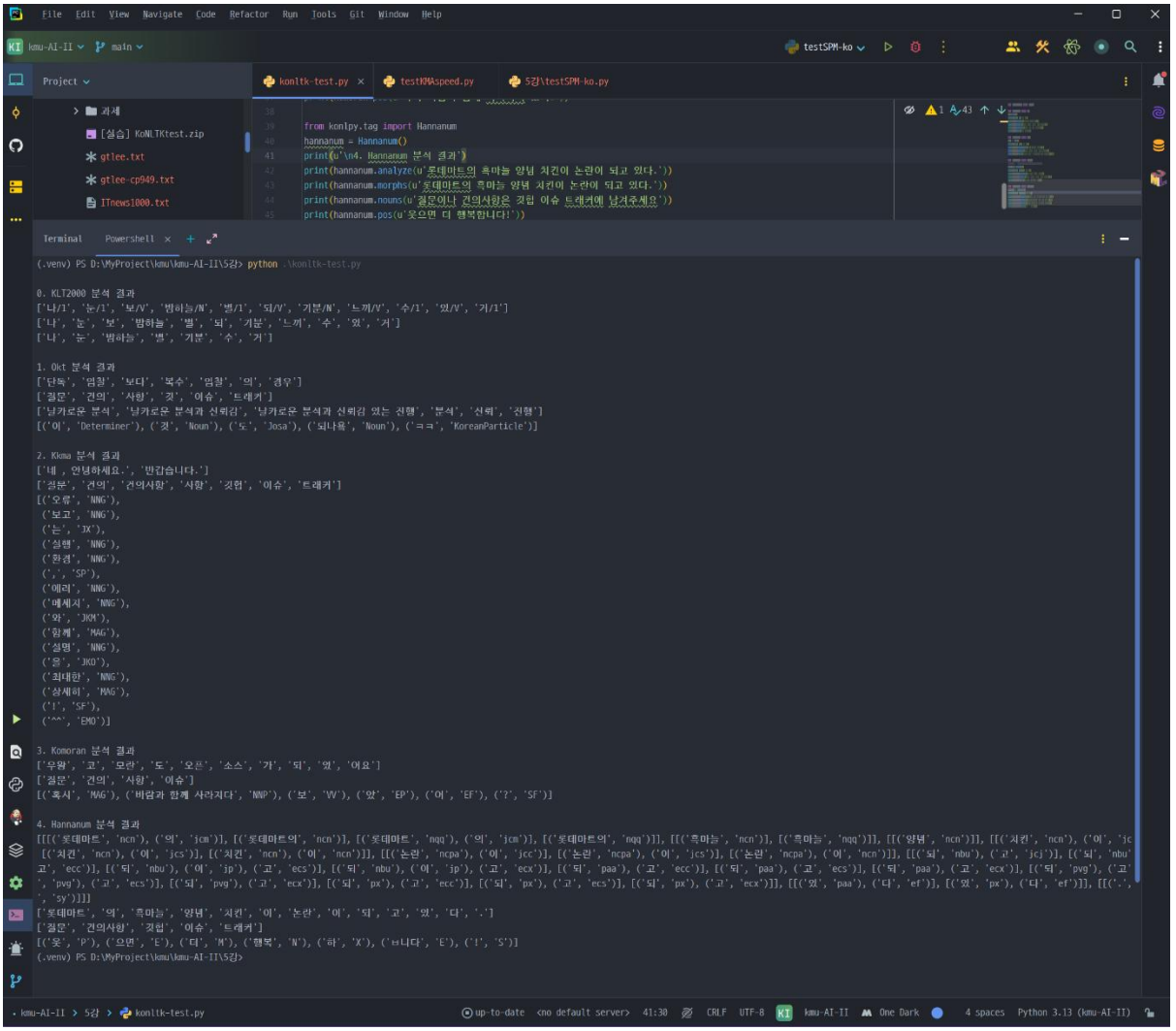


2. "[실습] KoNLTKtest.zip"의 파이썬 코드 3개

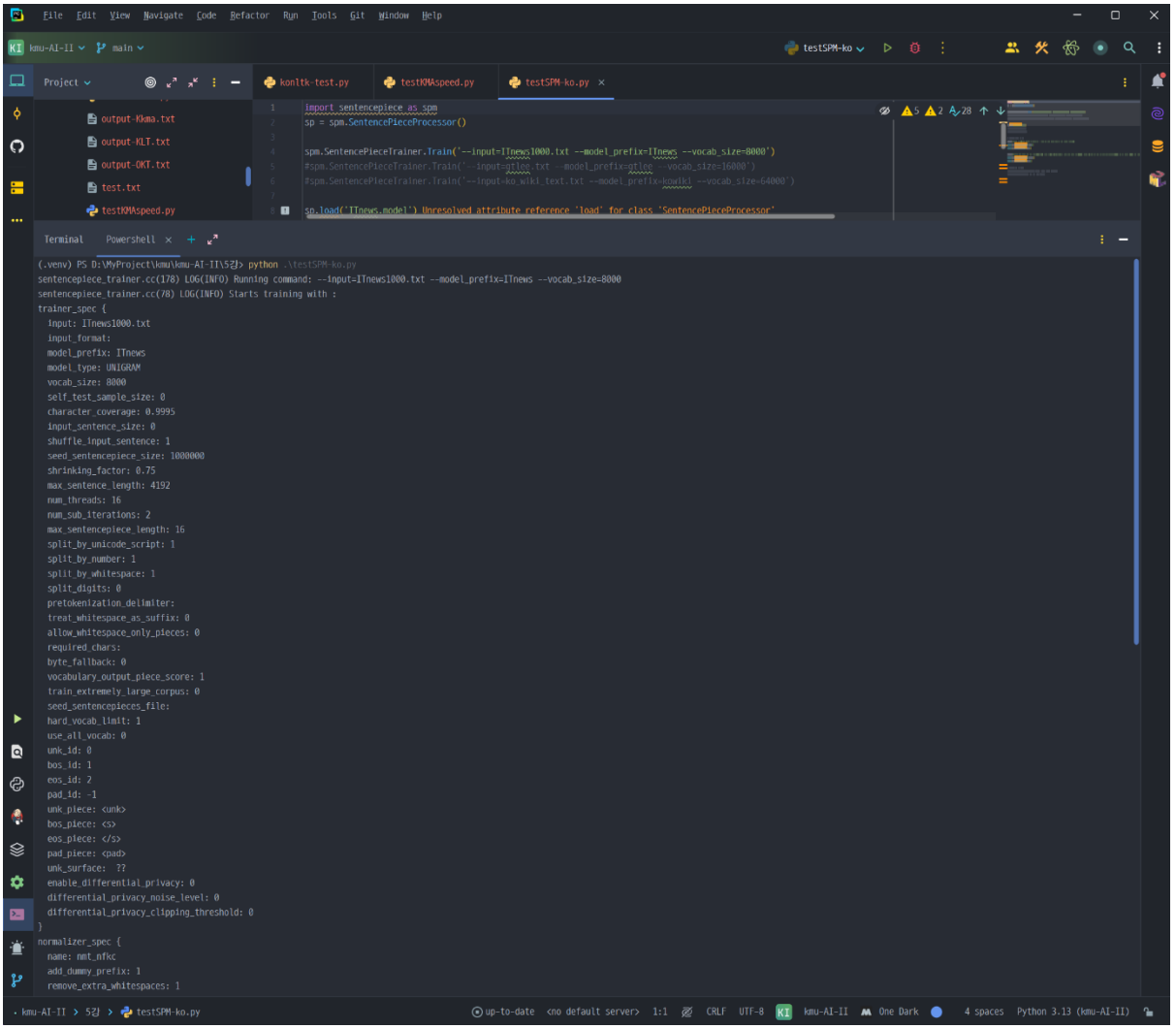
A. testKMAspeed.py

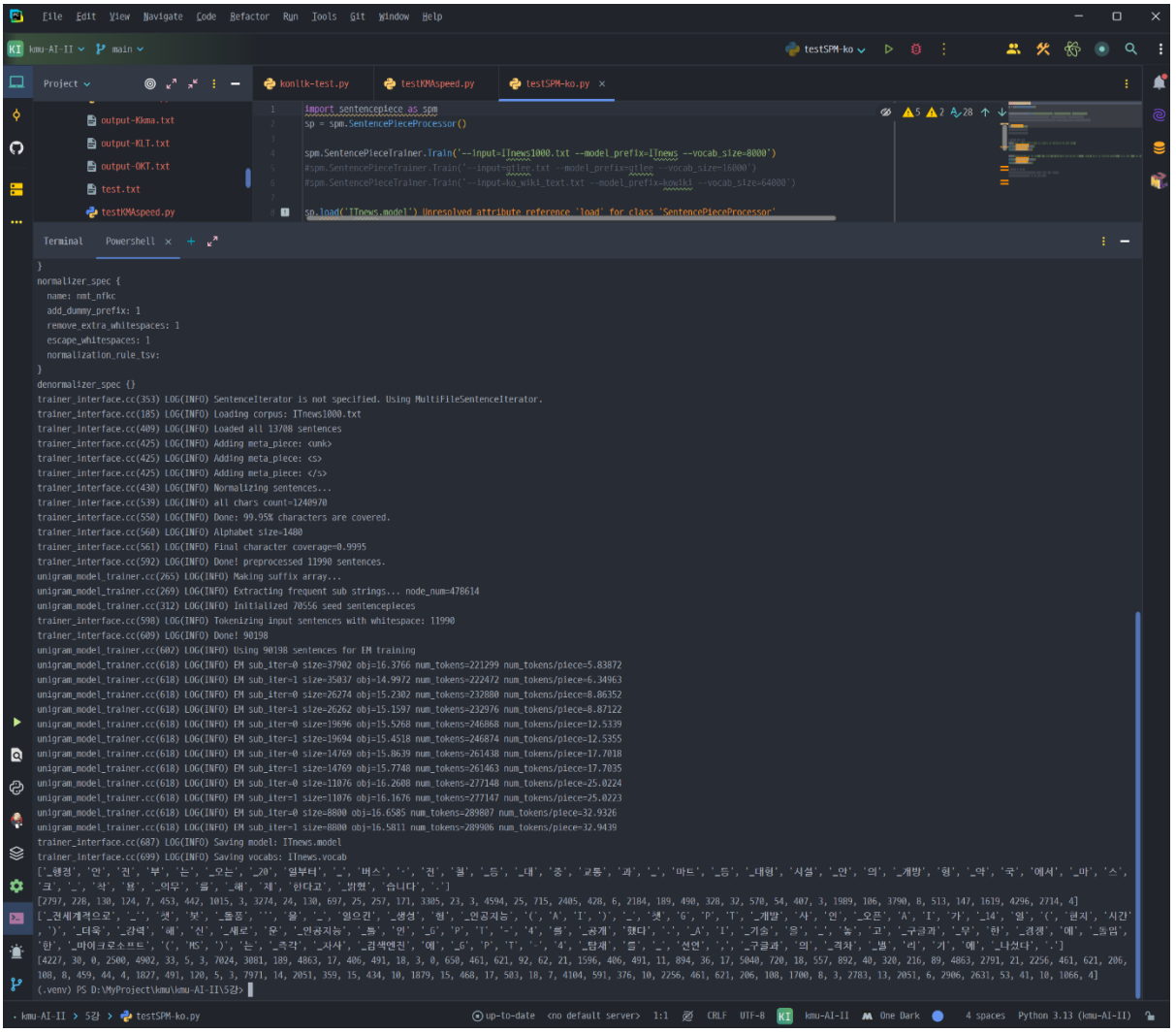


B. konltk-test.py



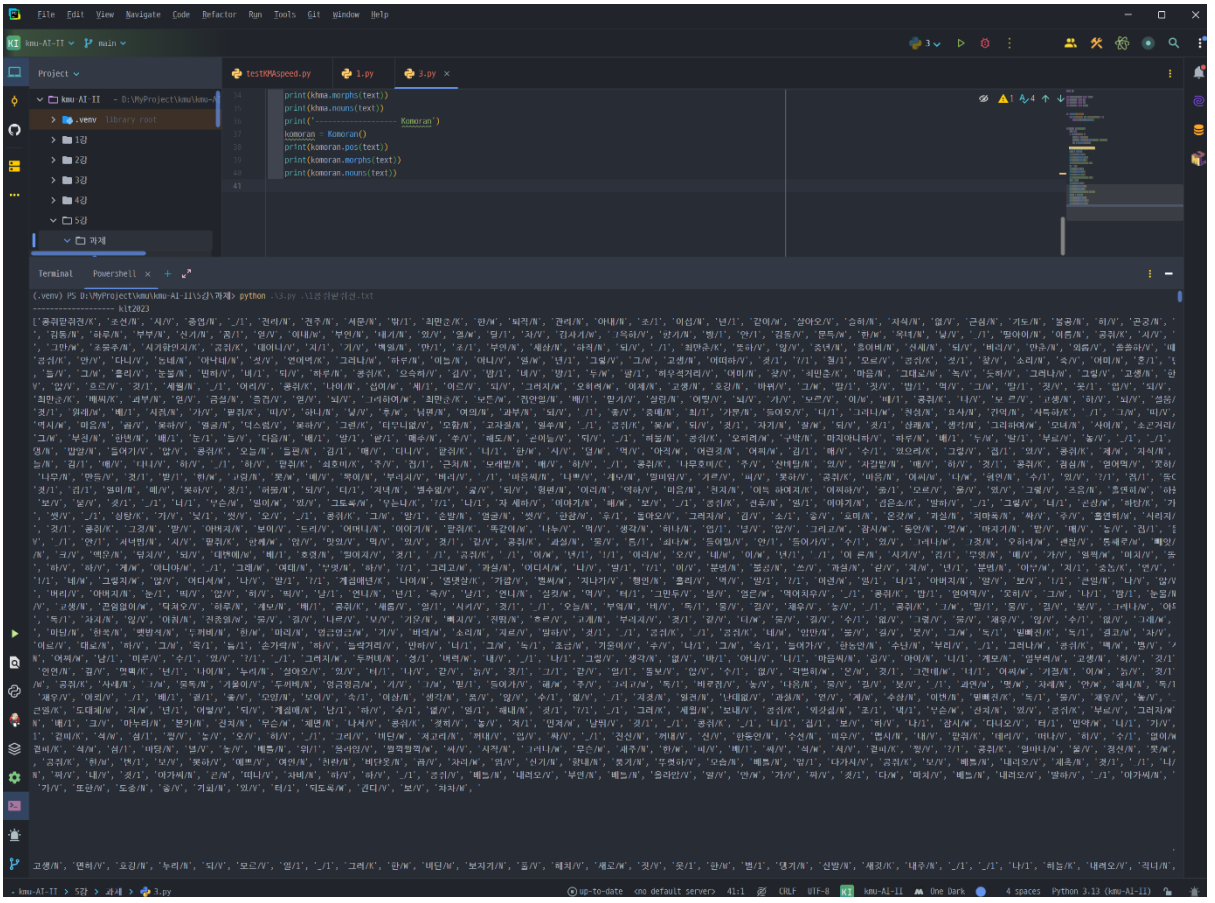
C. testSPM-ko.py

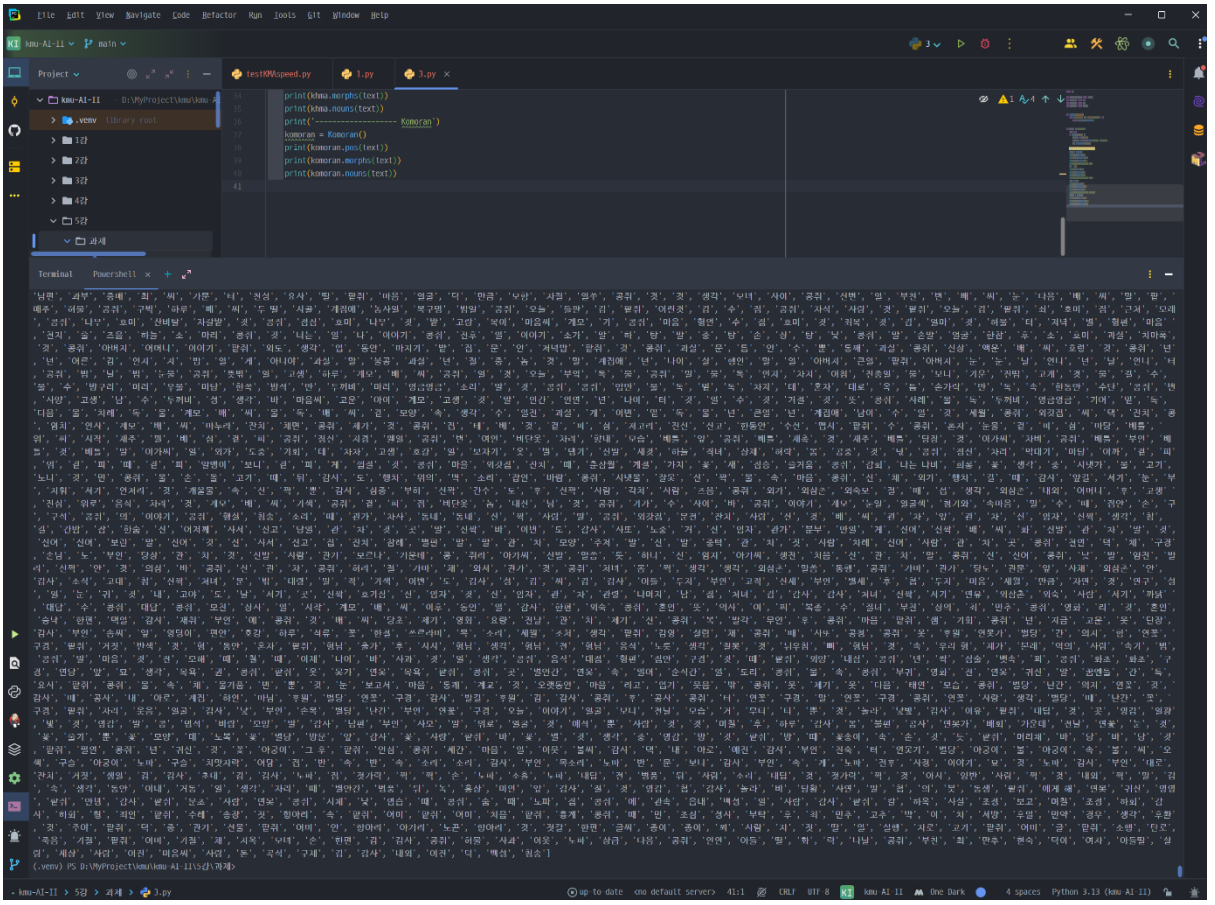


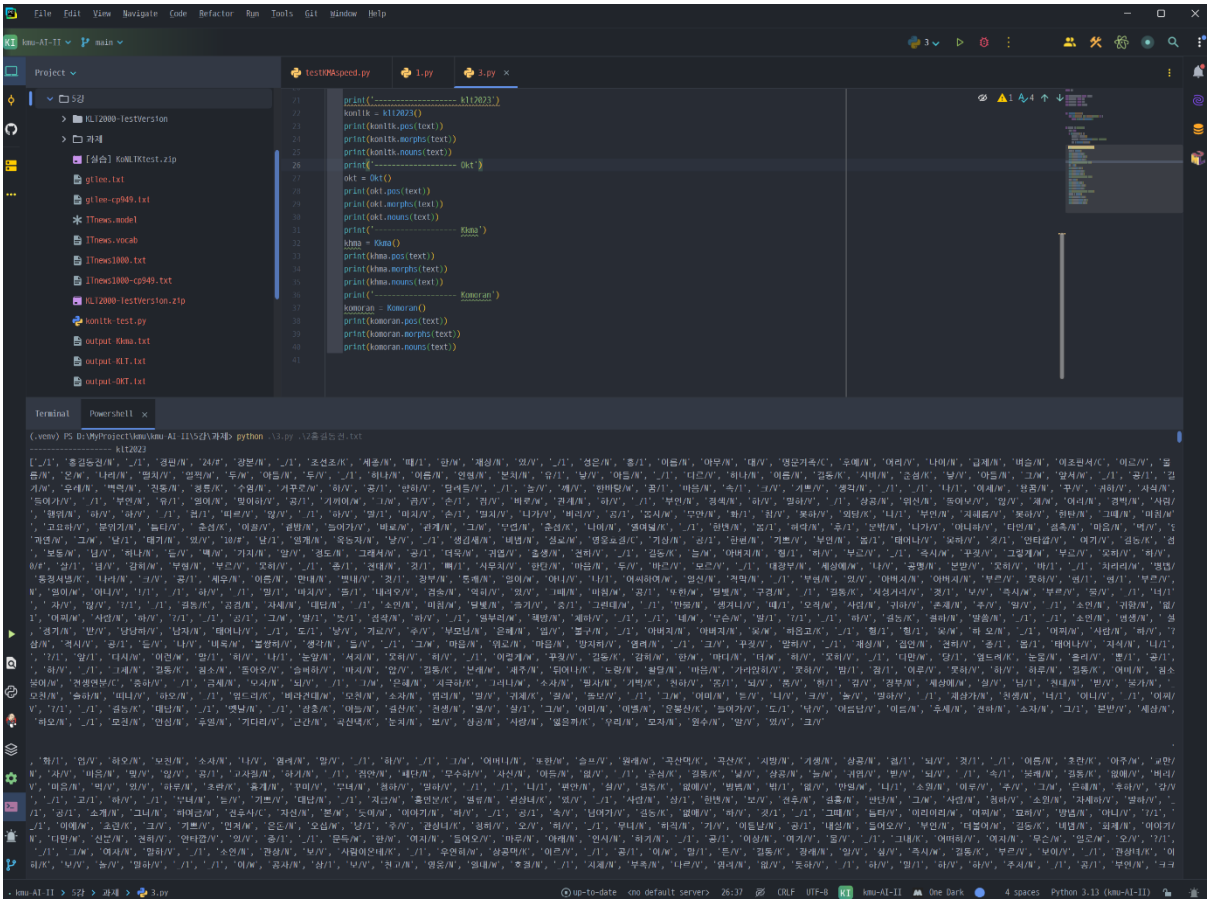

File Edit View Navigate Code Refactor Run Tools Git Window Help
kmu-AI-II main
testSPM-ko
Project
konlpy-test.py
testKMAspeed.py
testSPM-ko.py
output-Kkma.txt
output-KLT.txt
output-OKT.txt
test.txt
testKMAspeed.py
import sentencepiece as spm
sp = spm.SentencePieceProcessor()
spm.SentencePieceTrainer.Train('--input=ITnews1000.txt --model_prefix=ITnews --vocab_size=8000')
#spm.SentencePieceTrainer.Train('--input=gtlec.txt --model_prefix=gtlec --vocab_size=16000')
#spm.SentencePieceTrainer.Train('--input=ko_wiki_text.txt --model_prefix=kowiki --vocab_size=64000')
sp.load('ITnews.model') Unresolved attribute reference 'load' for class 'SentencePieceProcessor'
Terminal
Powershell
}
normalizer_spec {
name: nmt_nfkc
add_dummy_prefix: 1
remove_extra_whitespaces: 1
escape_whitespaces: 1
normalization_rule_tsv:
}
denormalizer_spec {}
trainer_interface.cc(353) LOG(INFO) SentenceIterator is not specified. Using MultiFileSentenceIterator.
trainer_interface.cc(185) LOG(INFO) Loading corpus: ITnews1000.txt
trainer_interface.cc(409) LOG(INFO) Loaded all 13708 sentences
trainer_interface.cc(425) LOG(INFO) Adding meta_piece: <unk>
trainer_interface.cc(425) LOG(INFO) Adding meta_piece: <s>
trainer_interface.cc(425) LOG(INFO) Adding meta_piece: </s>
trainer_interface.cc(430) LOG(INFO) Normalizing sentences...
trainer_interface.cc(539) LOG(INFO) all chars count=1240970
trainer_interface.cc(550) LOG(INFO) Done: 99.95% characters are covered.
trainer_interface.cc(560) LOG(INFO) Alphabet size=1480
trainer_interface.cc(561) LOG(INFO) Final character coverage=0.9995
trainer_interface.cc(592) LOG(INFO) Done! preprocessed 11990 sentences.
unigram_model_trainer.cc(265) LOG(INFO) Making suffix array...
unigram_model_trainer.cc(269) LOG(INFO) Extracting frequent sub strings... node_num=478614
unigram_model_trainer.cc(312) LOG(INFO) Initialized 70556 seed sentencepieces
trainer_interface.cc(598) LOG(INFO) Tokenizing input sentences with whitespace: 11990
trainer_interface.cc(609) LOG(INFO) Done! 90198
unigram_model_trainer.cc(602) LOG(INFO) Using 90198 sentences for EM training
unigram_model_trainer.cc(618) LOG(INFO) EM sub_iter=0 size=37902 obj=16.3766 num_tokens=221299 num_tokens/piece=5.83872
unigram_model_trainer.cc(618) LOG(INFO) EM sub_iter=1 size=35037 obj=14.9972 num_tokens=222472 num_tokens/piece=6.34963
unigram_model_trainer.cc(618) LOG(INFO) EM sub_iter=0 size=26274 obj=15.2302 num_tokens=232880 num_tokens/piece=8.86352
unigram_model_trainer.cc(618) LOG(INFO) EM sub_iter=1 size=26262 obj=15.1597 num_tokens=232976 num_tokens/piece=8.87122
unigram_model_trainer.cc(618) LOG(INFO) EM sub_iter=0 size=19696 obj=15.5268 num_tokens=246868 num_tokens/piece=12.5339
unigram_model_trainer.cc(618) LOG(INFO) EM sub_iter=1 size=19694 obj=15.4518 num_tokens=246874 num_tokens/piece=12.5355
unigram_model_trainer.cc(618) LOG(INFO) EM sub_iter=0 size=14769 obj=15.8639 num_tokens=261438 num_tokens/piece=17.7018
unigram_model_trainer.cc(618) LOG(INFO) EM sub_iter=1 size=14769 obj=15.7748 num_tokens=261463 num_tokens/piece=17.7035
unigram_model_trainer.cc(618) LOG(INFO) EM sub_iter=0 size=11076 obj=16.2608 num_tokens=277148 num_tokens/piece=25.0224
unigram_model_trainer.cc(618) LOG(INFO) EM sub_iter=1 size=11076 obj=16.1676 num_tokens=277147 num_tokens/piece=25.0223
unigram_model_trainer.cc(618) LOG(INFO) EM sub_iter=0 size=8800 obj=16.6585 num_tokens=289807 num_tokens/piece=32.9326
unigram_model_trainer.cc(618) LOG(INFO) EM sub_iter=1 size=8800 obj=16.5811 num_tokens=289986 num_tokens/piece=32.9439
trainer_interface.cc(687) LOG(INFO) Saving model: ITnews.model
trainer_interface.cc(699) LOG(INFO) Saving vocabs: ITnews.vocab
['_행경', '언', '진', '부', '는', '_오는', '_20', '열부터', '_', '버스', '-', '전', '철', '_등', '_대', '중', '교통', '과', '_', '바드', '_등', '_대형', '시설', '_안', '의', '_개방', '형', '_약', '국', '에서', '_마', '스', '크', '_', '작', '묘', '_의무', '을', '_해', '제', '한다고', '_보했', '습니다', '.']
[2797, 228, 130, 124, 7, 453, 442, 1015, 3, 3274, 24, 130, 697, 25, 257, 171, 3305, 23, 3, 4594, 25, 715, 2405, 428, 6, 2184, 189, 490, 328, 32, 570, 54, 407, 3, 1989, 106, 3790, 8, 513, 147, 1619, 4296, 2714, 4]
['_견세계적으로', '_-', '햇', '보', '_돌풍', '-', '을', '_', '왈으긴', '_생성', '현', '_언공지능', '(', 'A', 'I', ')', '_', '햇', 'G', 'P', 'T', '_개발', '사', '언', '_오른', 'A', 'I', '가', '_14', '열', '(', '현지', '시건', ')', '_더욱', '_강력', '해', '신', '_세로', '운', '_언공지능', '_들', '언', '_G', 'P', 'I', '-', '4', '를', '_검개', '했다', '.', '_A', 'I', '_기술', '을', '_', '놓', '고', '_구글과', '_두', '한', '_경쟁', '에', '_돌입', '한', '_마이크로소프트', '(', 'MS', ')', '는', '_즉각', '_사서', '_검색엔진', '에', '_G', 'P', 'T', '-', '4', '_탑재', '_', '_', '선언', '.', '_구글과', '의', '_격차', '_벌', '리', '거', '에', '_나섰다', '.']
[4227, 30, 0, 2500, 4902, 33, 5, 3, 7024, 3081, 189, 4863, 17, 406, 491, 18, 3, 0, 650, 461, 621, 92, 62, 21, 1596, 406, 491, 11, 894, 36, 17, 5040, 720, 18, 557, 892, 40, 320, 216, 89, 4863, 2791, 21, 2256, 461, 621, 206, 108, 8, 459, 44, 4, 1827, 491, 120, 5, 3, 7971, 14, 2051, 359, 15, 434, 10, 1879, 15, 468, 17, 503, 18, 7, 4104, 591, 376, 10, 2256, 461, 621, 206, 108, 1700, 8, 3, 2783, 13, 2051, 6, 2906, 2631, 53, 41, 10, 1066, 4]
(.venv) PS D:\MyProject\kmu\kmu-AI-II\5강>
kmu-AI-II > 5강 > testSPM-ko.py
up-to-date
<no default server>
1:1
CRLF
UTF-8
kmu-AI-II
One Dark
4 spaces
Python 3.13 (kmu-AI-II)

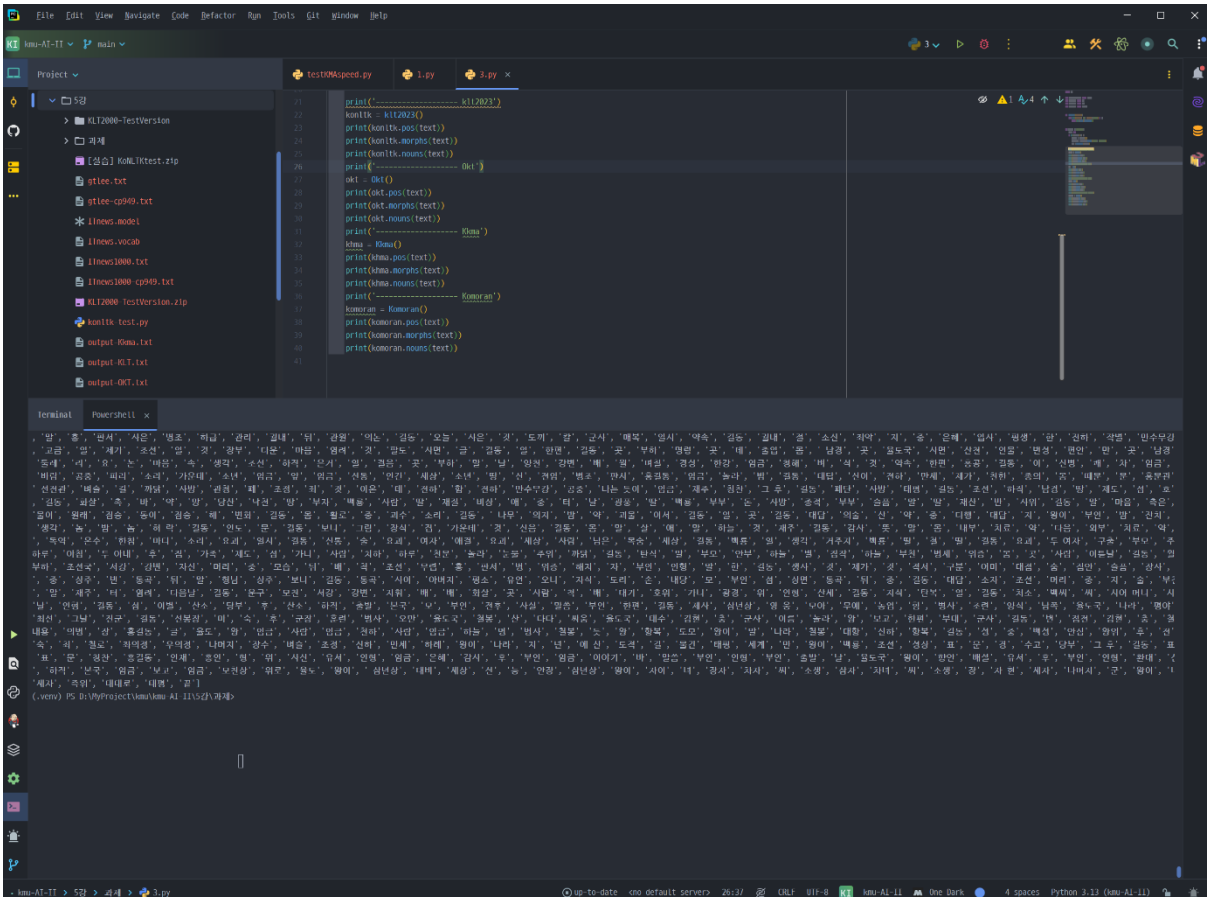
3. [과제1]에서 준비한 텍스트 파일에 대해 파이썬/윈도용 형태소 분석기 실습

A. 파이썬 형태소 분석기(KoNLPy, KoNLTK) 실습 내용 및 실습화면 스샷

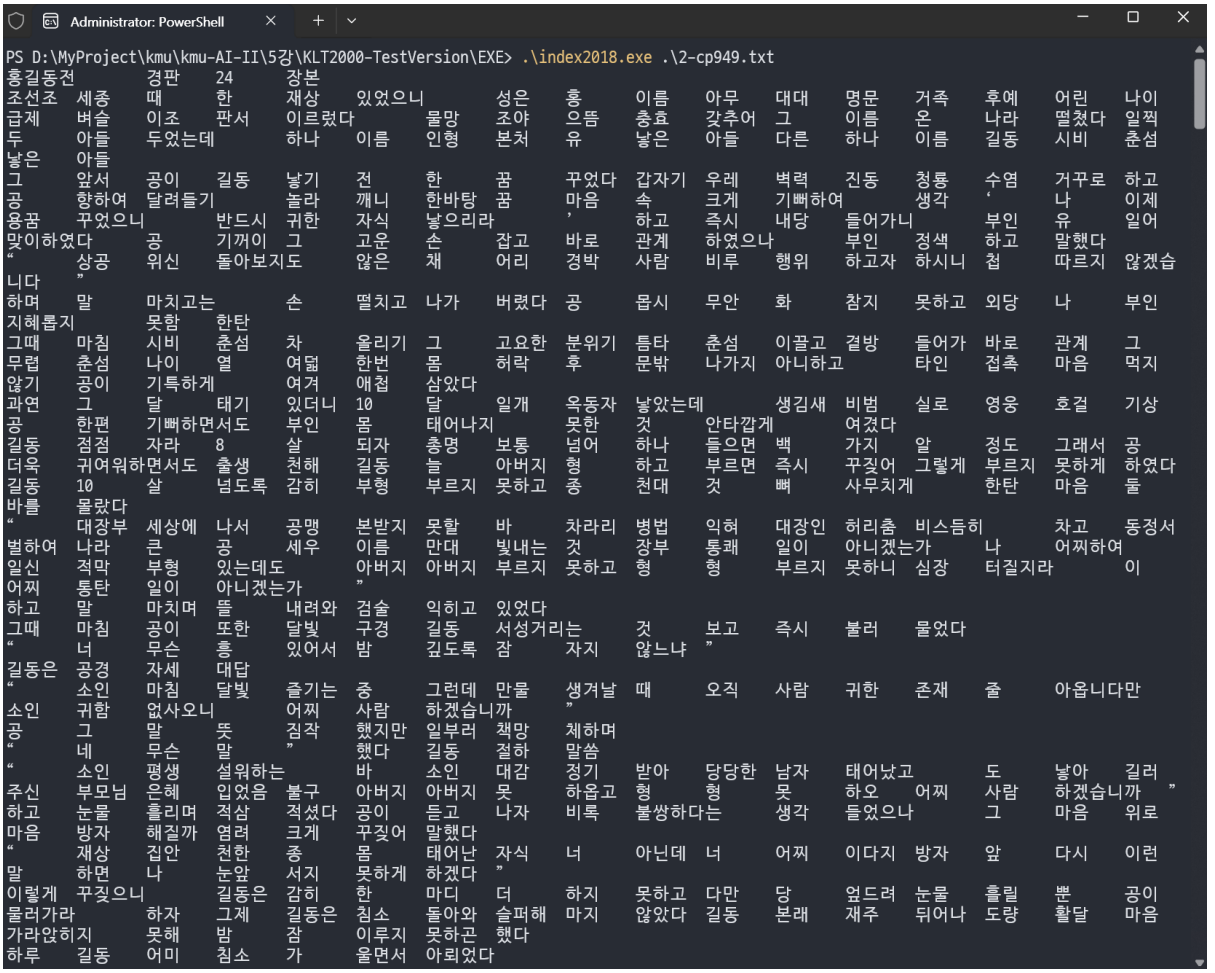



File Edit View Navigate Code Refactor Run Tools Git Window Help
kmu-AI-II
main
Project
test0KWspeed.py
1.py
3.py
print(khma.morphs(text))
print(khma.nouns(text))
print("------------------ Komoran")
komoran = Komoran()
print(komoran.pos(text))
print(komoran.morphs(text))
print(komoran.nouns(text))
Terminal
Powershell
(.venv) PS D:\MyProject\kmu\kmu-AI-II\5강\과제>
kmu-AI-II > 5강 > 과제 > 3.py
up to date
Python 3.13 (kmu-AI-II)


File Edit View Navigate Code Refactor Run Tools Git Window Help
kmu-AI-TT main
Project
testKMAspeed.py
1.py
3.py
print('---------- klt2023')
komlk = klt2023()
print(komlk.pos(text))
print(komlk.morphs(text))
print(komlk.nouns(text))
print('---------- Okt')
okt = Okt()
print(okt.pos(text))
print(okt.morphs(text))
print(okt.nouns(text))
print('---------- Kkma')
kkma = Kkma()
print(kkma.pos(text))
print(kkma.morphs(text))
print(kkma.nouns(text))
print('---------- Komoran')
komoran = Komoran()
print(komoran.pos(text))
print(komoran.morphs(text))
print(komoran.nouns(text))
Terminal
Powershell
(.venv) PS D:\MyProject\kmu\kmu-AI-II\5강\과제> python .\3.py .\2홍길동전.txt
---------- klt2023


File Edit View Navigate Code Refactor Run Tools Git Window Help
kmu-AI-IT
main
Project
testKMAspeed.py
1.py
3.py
5강
KLT2000-TestVersion
과제
[샘플] KoNLTKtest.zip
gtlee.txt
gtlee-cp949.txt
llnews.model
llnews.vocab
llnews1000.txt
llnews1000-cp949.txt
KLT2000-TestVersion.zip
konltk-test.py
output-Kkma.txt
output-KLT.txt
output-OKT.txt
print('-------------------- klt2023')
konltk = klt2023()
print(konltk.pos(text))
print(konltk.morphs(text))
print(konltk.nouns(text))
print('-------------------- Okt')
okt = Okt()
print(okt.pos(text))
print(okt.morphs(text))
print(okt.nouns(text))
print('-------------------- Kkma')
kkma = Kkma()
print(kkma.pos(text))
print(kkma.morphs(text))
print(kkma.nouns(text))
print('-------------------- Komoran')
komoran = Komoran()
print(komoran.pos(text))
print(komoran.morphs(text))
print(komoran.nouns(text))
terminal
Powershell
(.venv) PS D:\MyProject\kmu\kmu AI II\5강\과제>

B. index2018.exe를 이용하여 윈도 cmd창에서 실습 내용 및 실습화면 스샷


Administrator: PowerShell
PS D:\MyProject\kmu\kmu-AI-II\5강\KLT2000-TestVersion\EXE> .\index2018.exe .\2-cp949.txt
홍길동전 경판 24 장본
조선조 세종 때 한 재상 있었으니 성은 홍 이름 아무 대대 명문 거족 후예 어린 나이
급제 벼슬 이조 판서 이르렀다 물망 조야 으뜸 충효 갖추어 그 이름 온 나라 떨쳤다 일찍
두 아들 두었는데 하나 이름 인형 본처 유 낳은 아들 다른 하나 이름 길동 시비 춘섬
낳은 아들
그 앞서 공이 길동 낳기 전 한 꿈 꾸었다 갑자기 우레 벽력 진동 청룡 수염 거꾸로 하고
공 향하여 달려들기 놀라 깨니 한바탕 꿈 마음 속 크게 기뻐하여 생각 ‘ 나 이제
용꿈 꾸었으니 반드시 귀한 자식 낳으리라 ’ 하고 즉시 내당 들어가니 부인 유 일어
맞이하였다 공 기꺼이 그 고운 손 잡고 바로 관계 하였으나 부인 정색 하고 말했다
“ 상공 위신 돌아보지도 않은 채 어리 경박 사람 비루 행위 하고자 하시니 첩 따르지 않겠습
니다 ”
하며 말 마치고는 손 떨치고 나가 버렸다 공 몹시 무안 화 참지 못하고 외당 나 부인
지혜롭지 못함 한탄
그때 마침 시비 춘섬 차 올리기 그 고요한 분위기 틈타 춘섬 이끌고 결방 들어가 바로 관계 그
무렵 춘섬 나이 열 여덟 한번 몸 허락 후 문밖 나가지 아니하고 타인 접촉 마음 먹지
않기 공이 기특하게 여겨 애첩 삼았다
과연 그 달 태기 있더니 10 달 일개 옥동자 낳았는데 생김새 비범 실로 영웅 호걸 기상
공 한편 기뻐하면서도 부인 몸 태어나지 못한 것 안타깝게 여겼다
길동 점점 자라 8 살 되자 총명 보통 넘어 하나 들으면 백 가지 알 정도 그래서 공
더욱 귀여워하면서도 출생 천해 길동 늘 아버지 형 하고 부르면 즉시 꾸짖어 그렇게 부르지 못하게 하였다
길동 10 살 넘도록 감히 부형 부르지 못하고 종 천대 것 뼈 사무치게 한탄 마음 둘
바를 몰랐다
“ 대장부 세상에 나서 공맹 본받지 못할 바 차라리 병법 익혀 대장인 허리춤 비스듬히 차고 동정서
벌하여 나라 큰 공 세우 이름 만대 빛내는 것 장부 통쾌 일이 아니겠는가 나 어찌하여
일신 적막 부형 있는데도 아버지 아버지 부르지 못하고 형 형 부르지 못하니 심장 터질지라 이
어찌 통탄 일이 아니겠는가 ”
하고 말 마치며 뜰 내려와 검술 익히고 있었다
그때 마침 공이 또한 달빛 구경 길동 서성거리는 것 보고 즉시 불러 물었다
“ 너 무슨 흥 있어서 밤 깊도록 잠 자지 않느냐 ”
길동은 공경 자세 대답
“ 소인 마침 달빛 즐기는 중 그런데 만물 생겨날 때 오직 사람 귀한 존재 줄 아옵니다만
소인 귀함 없사오니 어찌 사람 하겠습니까 ”
공 그 말 뜻 짐작 했지만 일부러 책망 체하며
“ 네 무슨 말 ” 했다 길동 절하 말씀
“ 소인 평생 설워하는 바 소인 대감 정기 받아 당당한 남자 태어났고 도 낳아 길러
주신 부모님 은혜 입었음 불구 아버지 아버지 못 하옵고 형 형 못 하오 어찌 사람 하겠습니까 ”
하고 눈물 흘리며 적삼 적셨다 공이 듣고 나자 비록 불쌍하다는 생각 들었으나 그 마음 위로
마음 방자 해질까 염려 크게 꾸짖어 말했다
“ 재상 집안 천한 종 몸 태어난 자식 너 아닌데 너 어찌 이다지 방자 앞 다시 이런
말 하면 나 눈앞 서지 못하게 하겠다 ”
이렇게 꾸짖으니 길동은 감히 한 마디 더 하지 못하고 다만 당 엎드려 눈물 흘릴 뿐 공이
물러가라 하자 그제 길동은 침소 돌아와 슬퍼해 마지 않았다 길동 본래 재주 뛰어나 도량 활달 마음
가라앉히지 못해 밤 잠 이루지 못하곤 했다
하루 길동 어미 침소 가 울면서 아뢰었다