

CAPSTONE PROJECT

최종 발표

162985_홍예진

164172_오정운

165410_이혜원

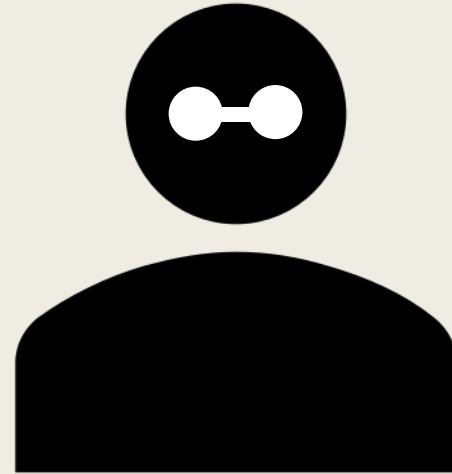


목차

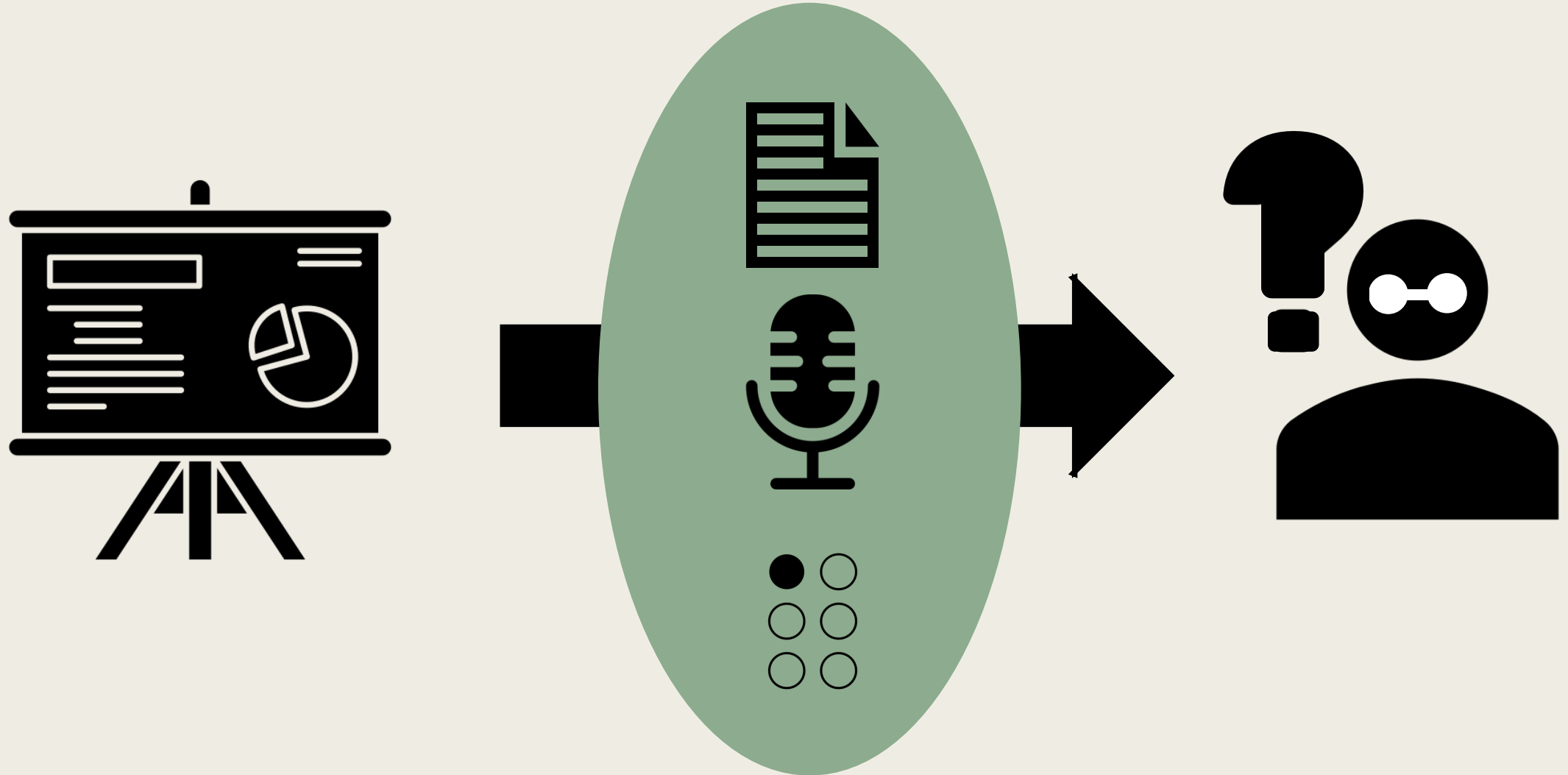
1. 추진 배경 및 프로젝트 개요
2. 사용 기술 및 도구
3. 기능 및 개발 과정
4. 사용자 평가
5. 프로그램에 대한 고찰
6. QnA



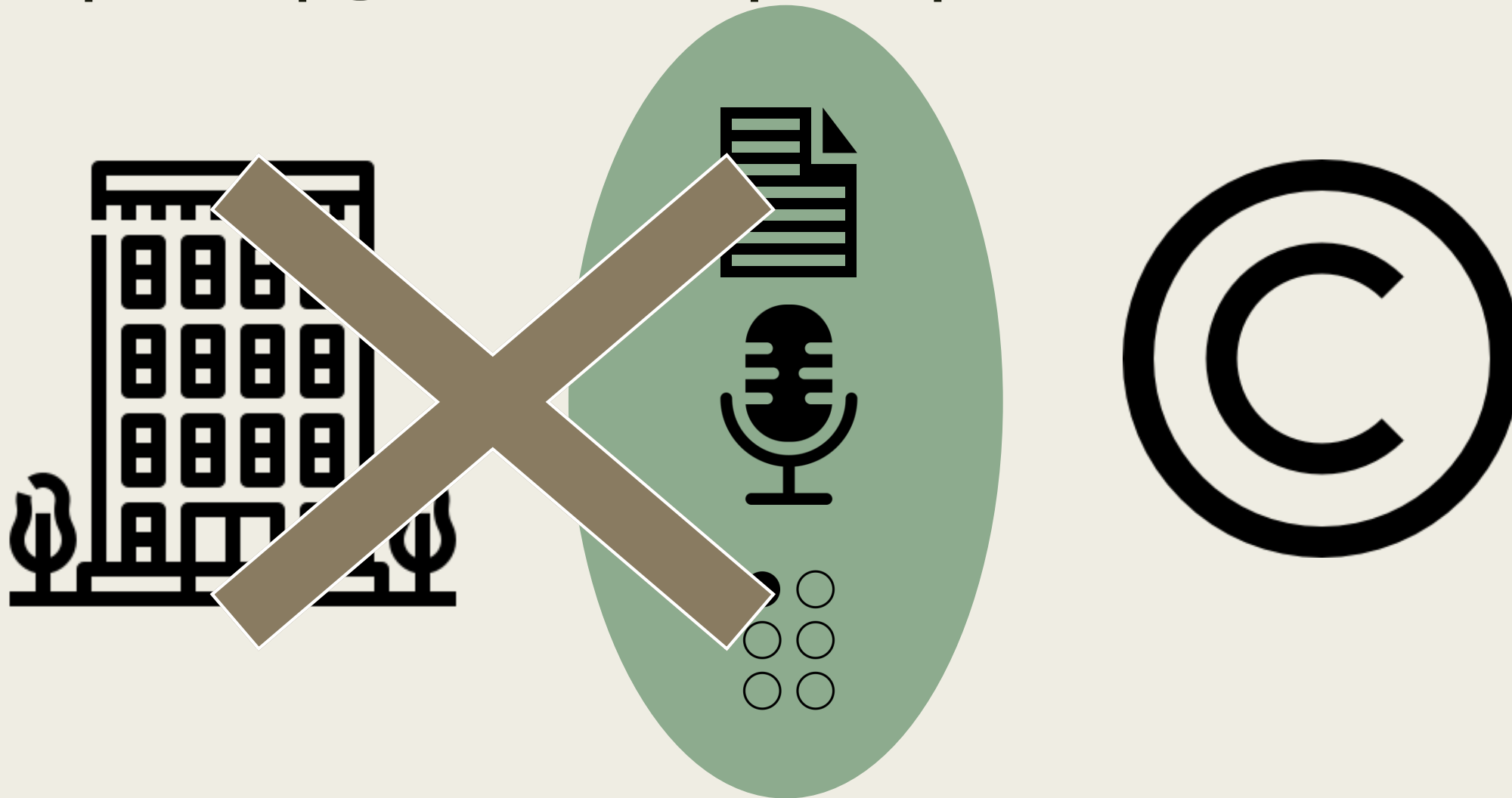
1. 추진 배경 및 프로젝트 개요



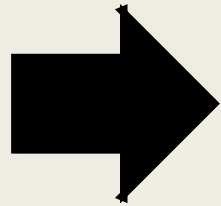
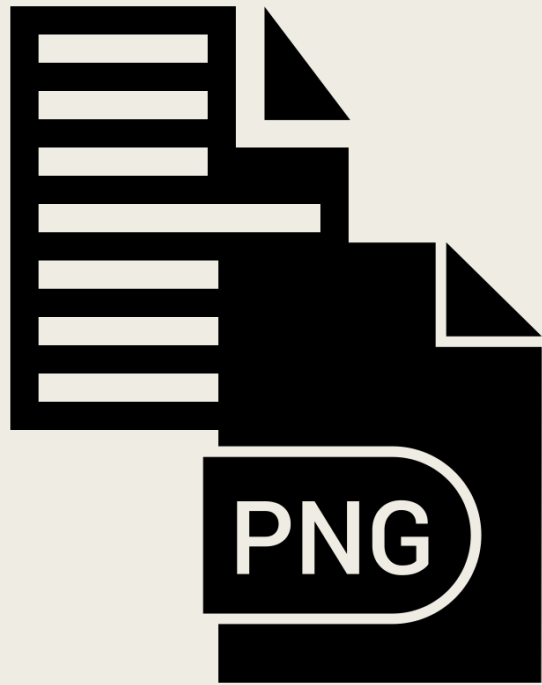
1. 추진 배경 및 프로젝트 개요



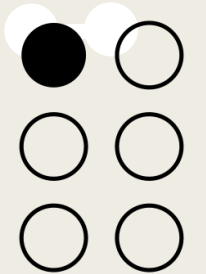
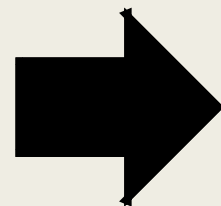
1. 추진 배경 및 프로젝트 개요



1. 추진 배경 및 프로젝트 개요



ABC..



2. 사용 기술 및 도구

- Pycharm – python IDE

:전체적 개발

- Tesseract-OCR

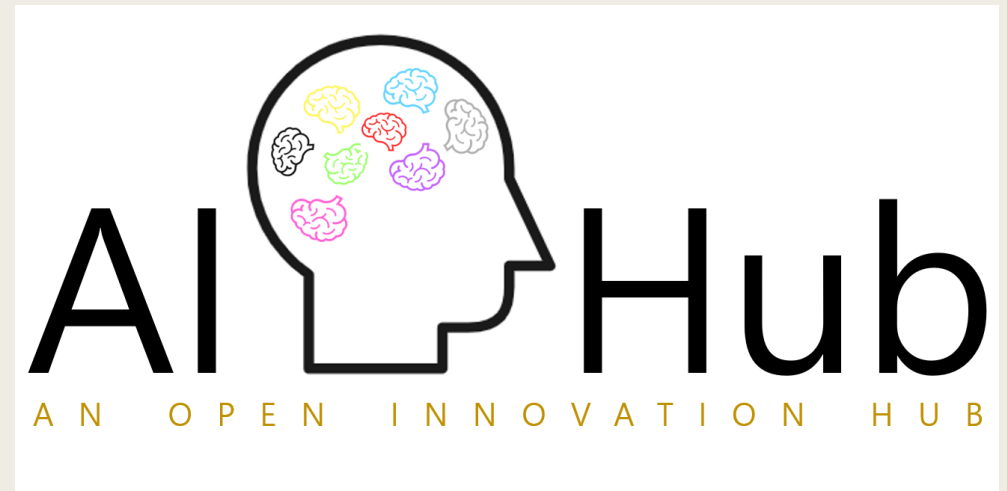
:이미지에서 텍스트 추출

- OpenCV

:이미지 전처리

- Pdfplumber, docx2txt, python-pptx, pyhwp

:문서에서 텍스트 추출



3. 기능 및 개발 과정 - 협업 과정



3. 기능 및 개발 과정 - 협업 과정

캡스톤 3명 · jnu.ac.kr [멤버 관리](#)

빠른 액세스

국립중앙도서관 2019-01-29
11-1371028-000759-03

컴퓨터 점자 규
지난주에 열어봄

이름 ↓

- 회의록
- 혜원
- 정운
- 예진

W	중간보고서.docx	2020. 5. 14. 나	415KB
P	0603 12자 회의록.pptx	오전 12:45 나	541KB
P	0527 11자 회의록.pptx	2020. 5. 28. 나	222KB
P	0520 10자 회의록.pptx	2020. 5. 20. 나	168KB
P	0512 9자 회의록.pptx	2020. 5. 14. 나	123KB
P	0506 8자 회의록.pptx	2020. 5. 6. 나	250KB
P	0429 7자 회의록.pptx	2020. 4. 29. 나	602KB
P	0422 6자 회의록.pptx	2020. 4. 24. 이혜원	237KB
P	0408_4자회의.pptx	2020. 4. 8. 나	217KB
P	0401_3자회의.pptx	2020. 4. 8. 나	52KB
W	0329_2자회의(카톡).docx	2020. 3. 30. 나	29KB
P	0328_1자회의.pptx	2020. 4. 1. 나	70KB

YeJinHong / Capstone [Unwatch](#)

<> Code 0 Issues 0 Pull requests 1 Actions Projects 0 Wiki Security 0

No description, website, or topics provided.

Manage topics

161 commits 4 branches 0 packages 0 releases

Branch: master New pull request Create new file Upload

YeJinHong symbol 에러처리	
.idea	부분 추출 오류 수정
example	확장자 blb로 변경, utf-8로 인코딩(결과 확인 필요)
img	symbol 에러처리
map	좌측에 인덱스가 존재하는 TextArea.py 추가
.gitignore	Initial commit
Aeyeicon.png	Initial commit
i2twidget.py	i2t로 writetext 이동
Mainwindow.py	부분 인식 후 문서 텍스트 추출 불가 오류 수정
T2Bwidget.py	tab 분리
TableWidget.py	symbol 에러처리
TextArea.py	부분 추출 오류 수정
braille_standard.py	점자 파일 형식 지정하는 부분을 다른 클래스로 분리
convertor.py	symbol 에러처리
text_extract.py	table_widget -> tabs으로 리팩터링



3. 기능 및 개발 과정 - 협업 과정

1. 목적

- 각 개인 조사 현황 보고 및 방향 재검토
- 현재

예진 : python gui 조사

혜원 : 텍스트 추출 api 조사(python연계방향)

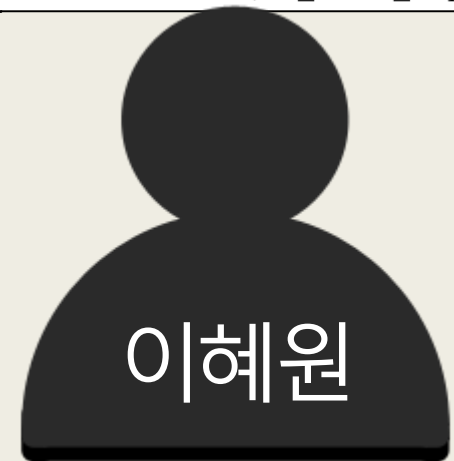
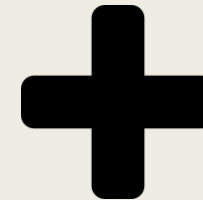
정운 : 기존 점역기 문제 등 개선 방향 설명의 계획서 배경 갈무리
+(이미지 패턴인식 조사)

3. 차후 일정 계획

- 홍예진
 - 불러온 파일의 미리보기 제공과 확대/축소 기능 구현
 - 인쇄 기능 구현
 - 설문지 작성
- 이혜원
 - 새 문서 작성과 저장 기능 구현
 - 페이지 목차 구현
 - 금주 회의록 작성
- 오정운
 - 추출한 텍스트를 점자로 바꾸는 점역 기능 구현

4. 다음주 활동 계획

- 혜원
 - 다양한 이미지 텍스트 추출 결과 공유
 - 점자 파일 형식 추가 기능 구현 (Convertor 개선)
- 정운
 - 유니코드 -> 아스키코드 수정 (kor_braille_map 수정)
 - char 단위 점역기능 구현(약어 구현)
- 예진
 - 약어 아스키코드 삽입 (kor_braille_map 수정)



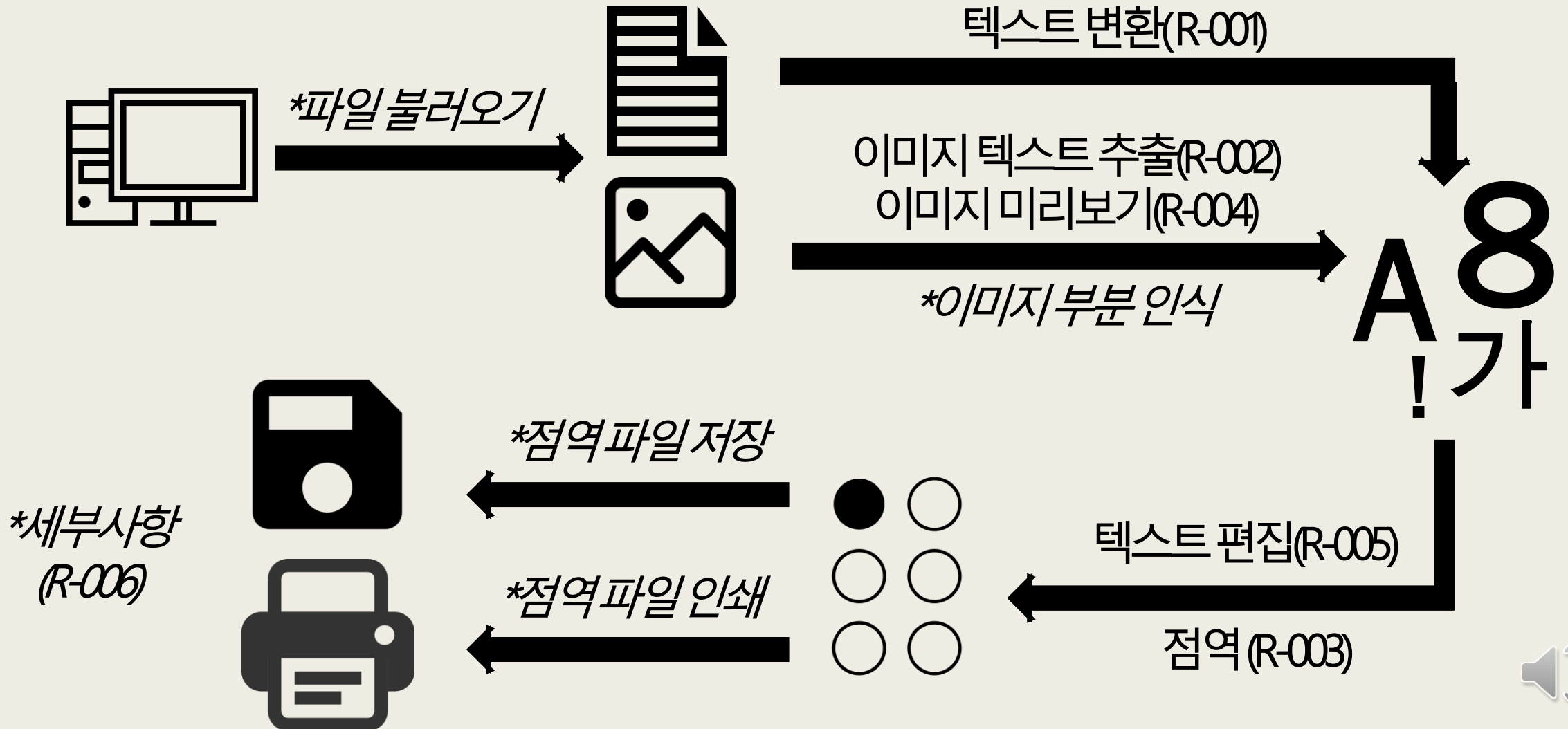
GUI 및 세부 모듈

텍스트->점자

파일->텍스트



3. 기능 및 개발 과정 - 프로그램 순서도

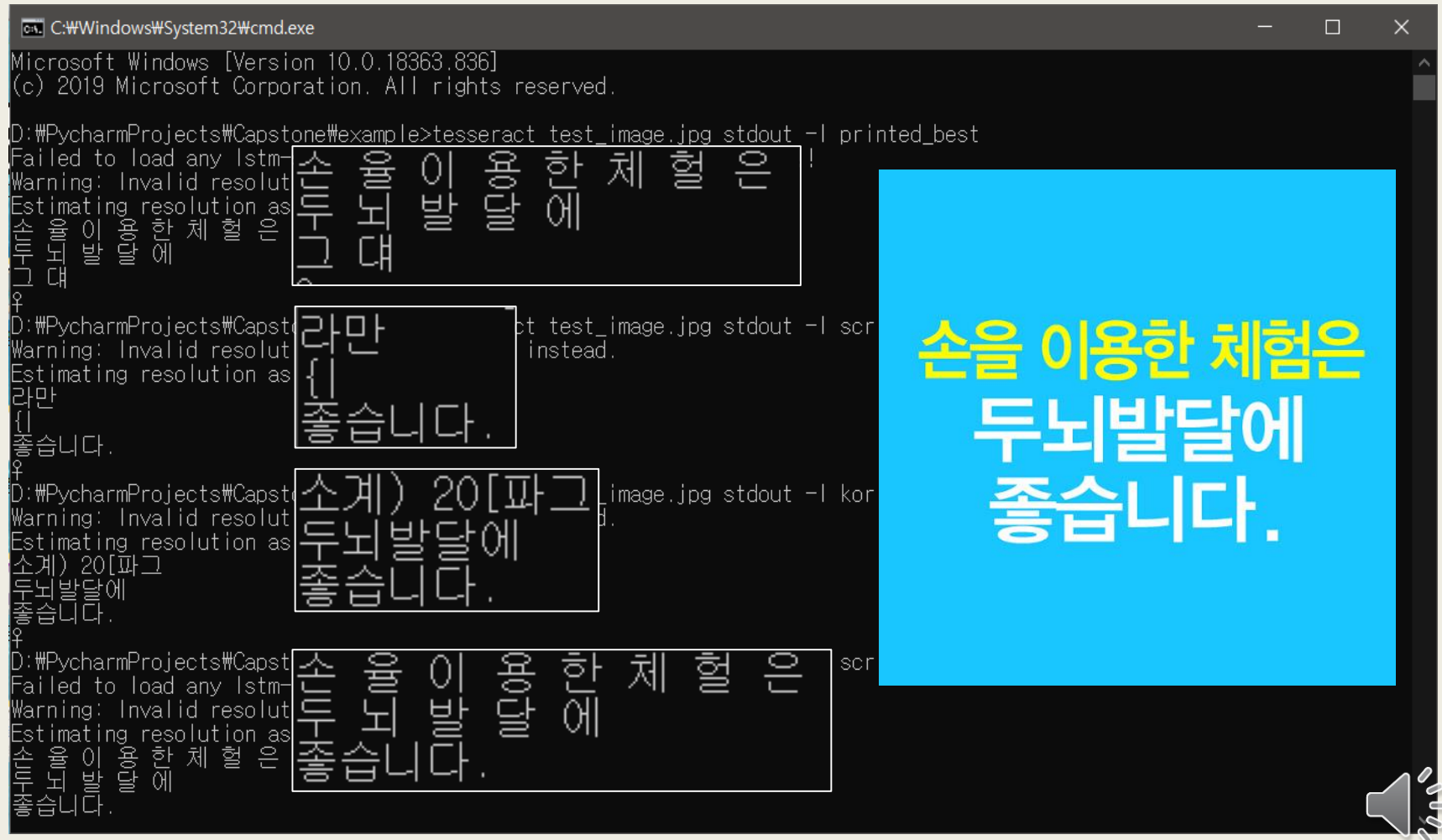


3. 기능 및 개발 과정 - 데이터 학습

학습시킨 데이터

기존의 한글 데이터
1,2

기존 데이터
+
학습시킨 데이터



DEMO

Text -> Braille



DEMO

텍스트 변환 (R-001)

D:/PycharmProjects/Capstone/qkfrdmsQu.txt - Aeye

File Edit Document

image -> text text -> braille text

파일 입력창 ☐ 이어쓰기 텍스트 결과창

qkfrdmsQu.txt - Windows 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

밝은 뺨
김 희 숙

깜빡거리다 꺼져버린
백열전구를 돌려 빼다가 그 속에
가느다랗고 바르르 떠는 뺨
둥근 몸을 밝혀 온 것이
이렇게 작은 뺨였다는 것을 알았다
아마도 이 밝은 뺨는
너무 밝아서 그동안 아무도
눈 뜨고는 보지 못했던 것
그 속이 꺼지고 난 다음에야
그 끊어진 뺨를
들여다 볼 수 있었다는 것도 알았다
숨을 참고 내 속에 뺨를 들여다보기 위해
찍은 엑스레이사진 속에서
하얗게 보이던 그 뺨들

Ln 1, Col 1 100% Windows (CRLF) UTF-8

이미지 부분 인식

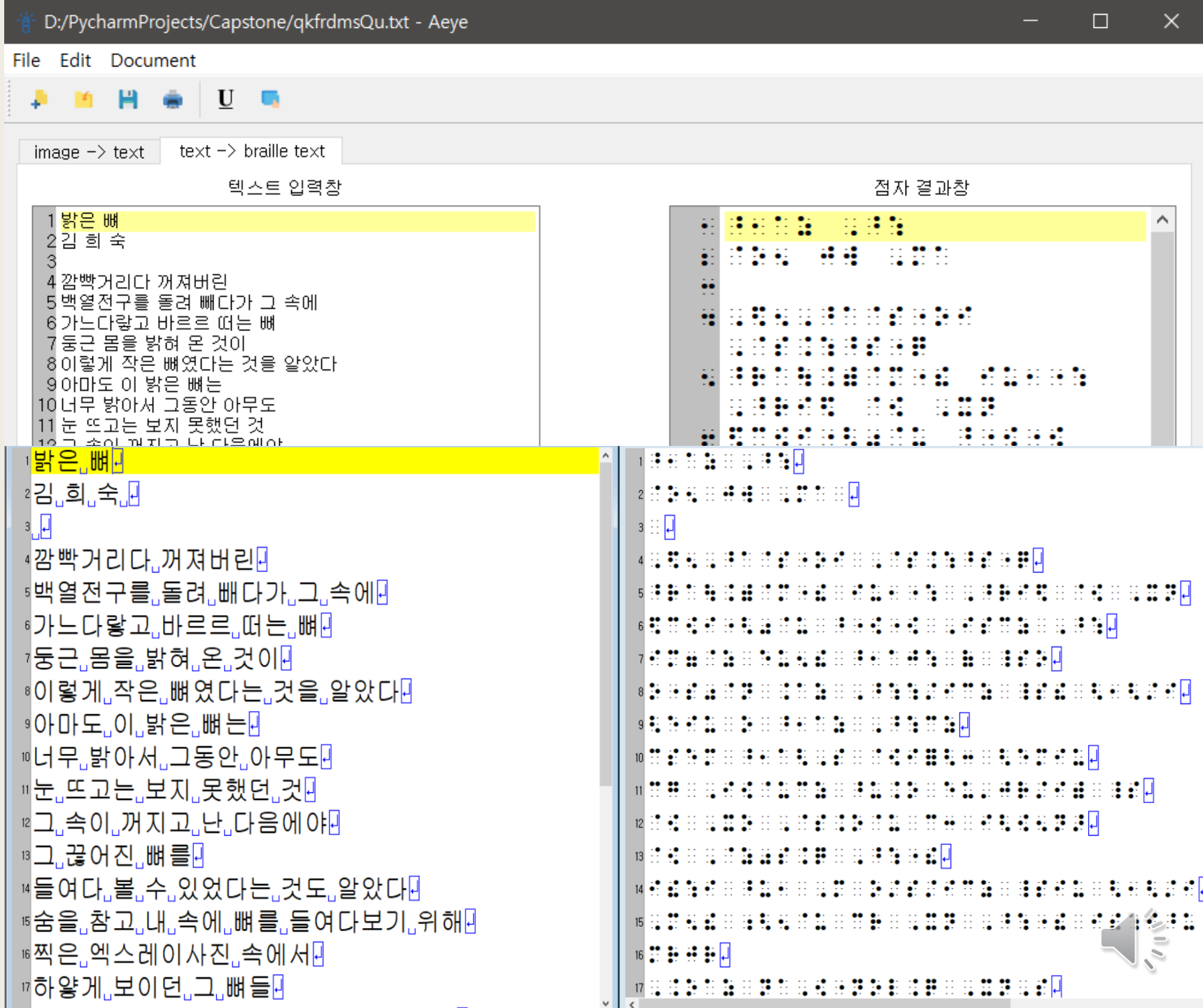
1 밝은 뺨
2 김 희 숙
3
4 깜빡거리다 꺼져버린
5 백열전구를 돌려 빼다가 그 속에
6 가느다랗고 바르르 떠는 뺨
7 둥근 몸을 밝혀 온 것이
8 이렇게 작은 뺨였다는 것을 알았다
9 아마도 이 밝은 뺨는
10 너무 밝아서 그동안 아무도
11 눈 뜨고는 보지 못했던 것
12 그 속이 꺼지고 난 다음에야
13 그 끊어진 뺨를
14 들여다 볼 수 있었다는 것도 알았다
15 숨을 참고 내 속에 뺨를 들여다보기 위해
16 찍은 엑스레이사진 속에서
17 하얗게 보이던 그 뺨들
18 몸 한 곳이 볼 꺼진 뒤에야 볼 수 있었던
19 앙상한 뺨는
20 그동안 나를 밝히고 있었다는 것
21 누구나 제 뺨를 보는 일이
22 어떤 위치를 찾기 위한 것이라면
23 하얀 뺨들 사이 볼 꺼진
24 검은 흔적을 찾는 일은 많이 더듬어야 할 일.
25 눈 뜨고는 볼 수 없는 밝은 뺨
26 뺨만 들어있어도 눈부시게 밝았던 전구
27 바람 사이에는 수많은
28 나뭇가지들이 뺨로 들어 있다

초기화

DEMO

점역 (R-003)

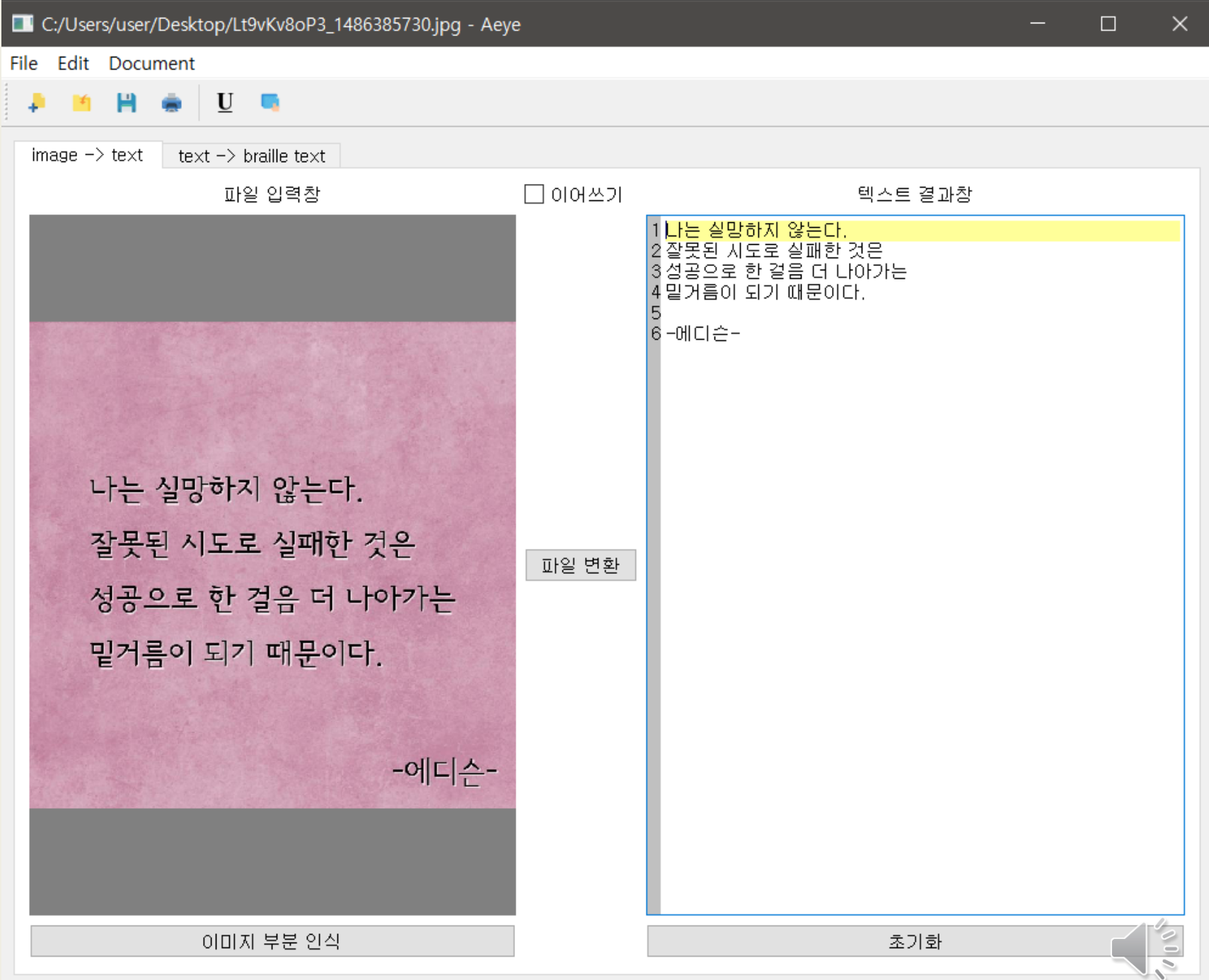
텍스트 편집 (R-005)



DEMO

이미지 텍스트 추출 (R-002)

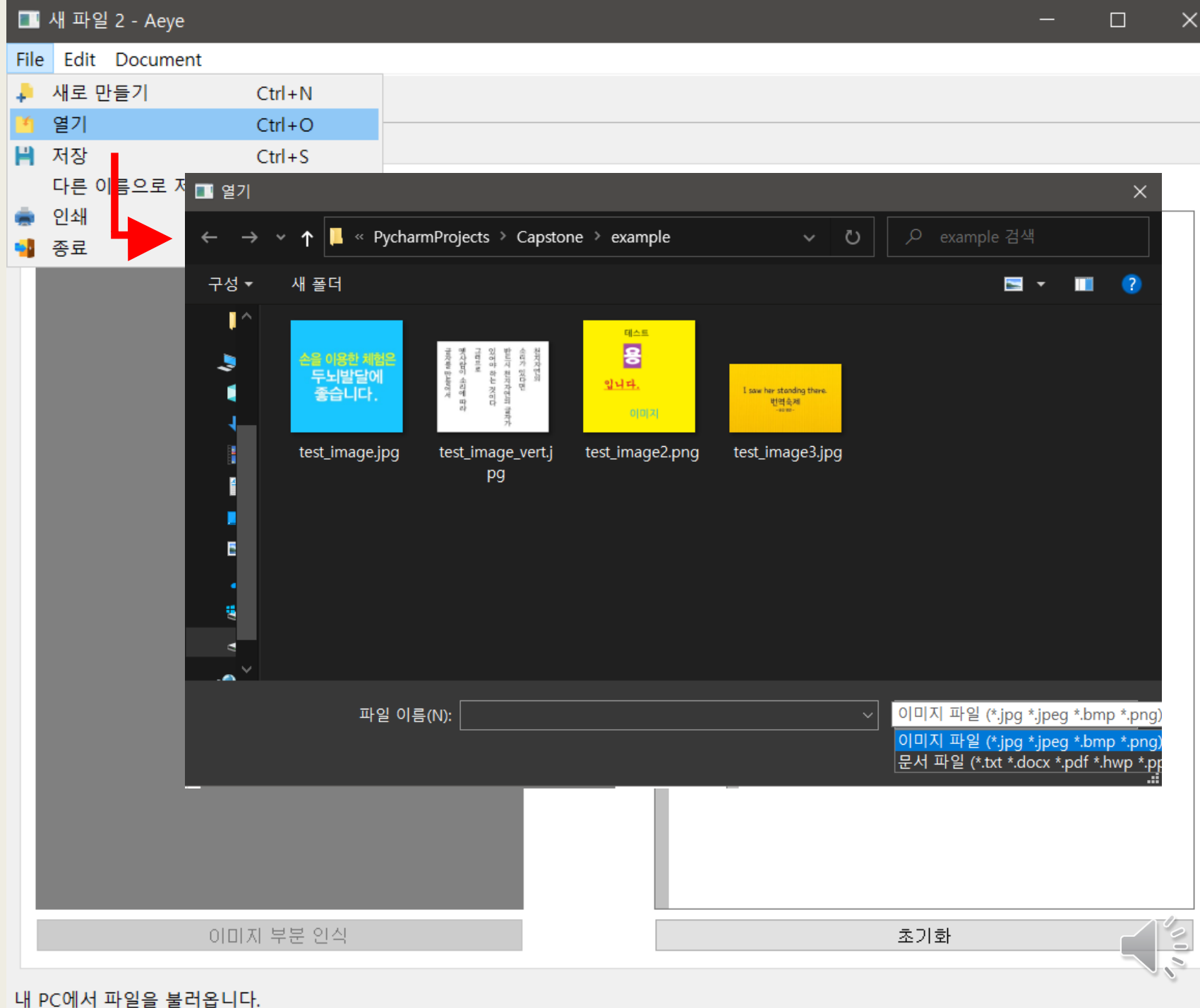
이미지 미리 보기 (R-004)



DEMO

세부 기능 (R-006)

- 파일
불러오기



DEMO

세부 기능 (R-006)

- 이미지
부분 인식

D:/PycharmProjects/Capstone/unnamed.jpg - Aeye

File Edit Document

image -> text text -> braille text

파일 입력창 ☐ 이어쓰기 텍스트 결과창

1 '해커스랩' 13단계 첫 통과
2 신출귀몰 숨씨 다시 과시

Drag to crop image

Cropped image

‘해킹 명예의 전당’ 1호 영예

오대호씨(20·포항공대 컴퓨터공학과 3). 오씨는 이틀밤을 꼬박 새운 끝에 최고 난이도인 13번째를 뚫는 데 성공, 해킹자유지대 ‘명예의 전당’ 1호로 등록되는 ‘영광’을 누렸다.

포항공대 해킹연구동호회 PLUS에 들어가면서 시작된 오씨의 해킹 이력. 굳이 해킹자유지대 첫 통과자란 것을 들먹이지 않더라도 그의 해킹 솜씨는 ‘알 만한 이들은 다 알 만큼’ 화려하다.

지난 98년 이탈리아에서 ‘dark’란 별명을 가진 해커가 포항공대에 침투한 사실을 발견, 해커수(7,492) 신고한 것을 비롯해 교내 시스템에 누군

‘해커스랩’ 13단계 첫
신출귀몰 숨씨 다시 과시

가 침입할 때마다 전산실에 정보해킹은 참을 모
람이었다. 또 교내 전산관리자를 대상으로 보안
관련 세미나를 벌일 만큼 보안에 관한 한 타의
추종을 불허한다.

지난 5월 모 신문사 사이트가 해킹당했을 편
일요일이라 전산담당자가 없었던지라 그 자신이
직접 문제를 해결하기도 했다.

오씨가 처음 컴퓨터를 만지기 시작한 것은 고
등학교 때(서울 구경고). 친구 소개로 가장 ‘재미
있다’는 C언어를 공부하면서 컴퓨터에 재미를
붙이기 시작. 당시로는 드문 홈페이지를 이미 제
작했다. 그러던 중 신문을 통해 해킹사고를 접하

‘해커스랩’ 13단계 첫 통과
신출귀몰 숨씨 다시 과시

실질적 컴퓨터 앞에서 포즈를 취한 ‘저명 해커’ 오대호씨.

인의 관심을 끌지는 못하고 있지만 아주 중요한
분야를 시도하고 싶어요.”
그래서 그는 장래 꿈을 꿔있을 때 ‘컴퓨터 전문
가’ 대신 ‘체계 정복이요’라고 말한다.

존경하는 인물에 대해서도 “누군가를 본받고
싶다면 2인자밖에 될 수 없잖아요. 그래서 저는
제 자신만을 생각해요”라고 말하는 당찬 신세대
해커 오대호.

170cm의 아담한 키에 계구정이 끼가 물론 물어
나는 안경 자람이지만 툭툭뚫뚫한 눈이 믿음직
스럽기만 하다.

[임성연 기자]

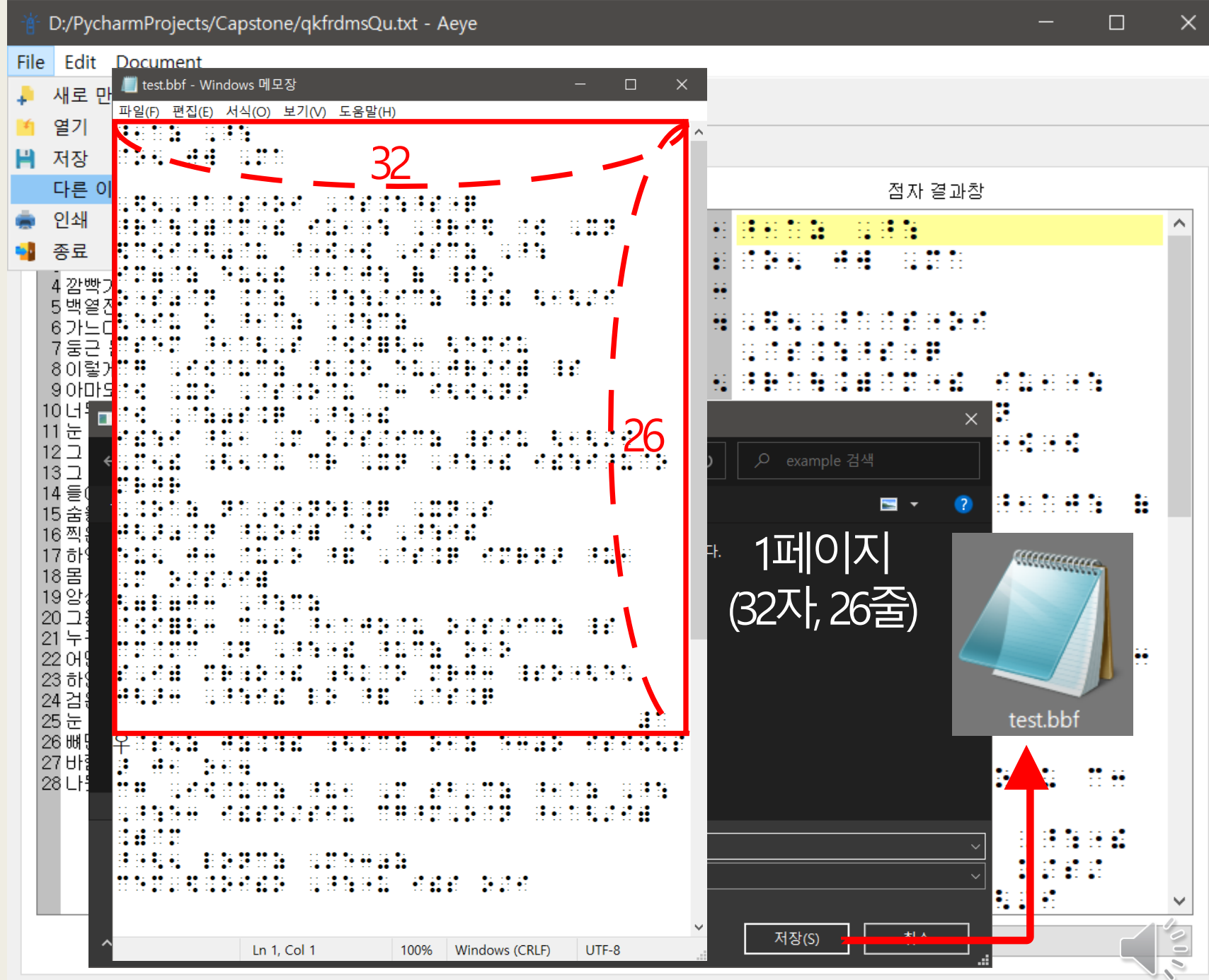
이미지 부분 인식

초기화

DEMO

세부 기능 (R-006)

- 점역 파일
저장

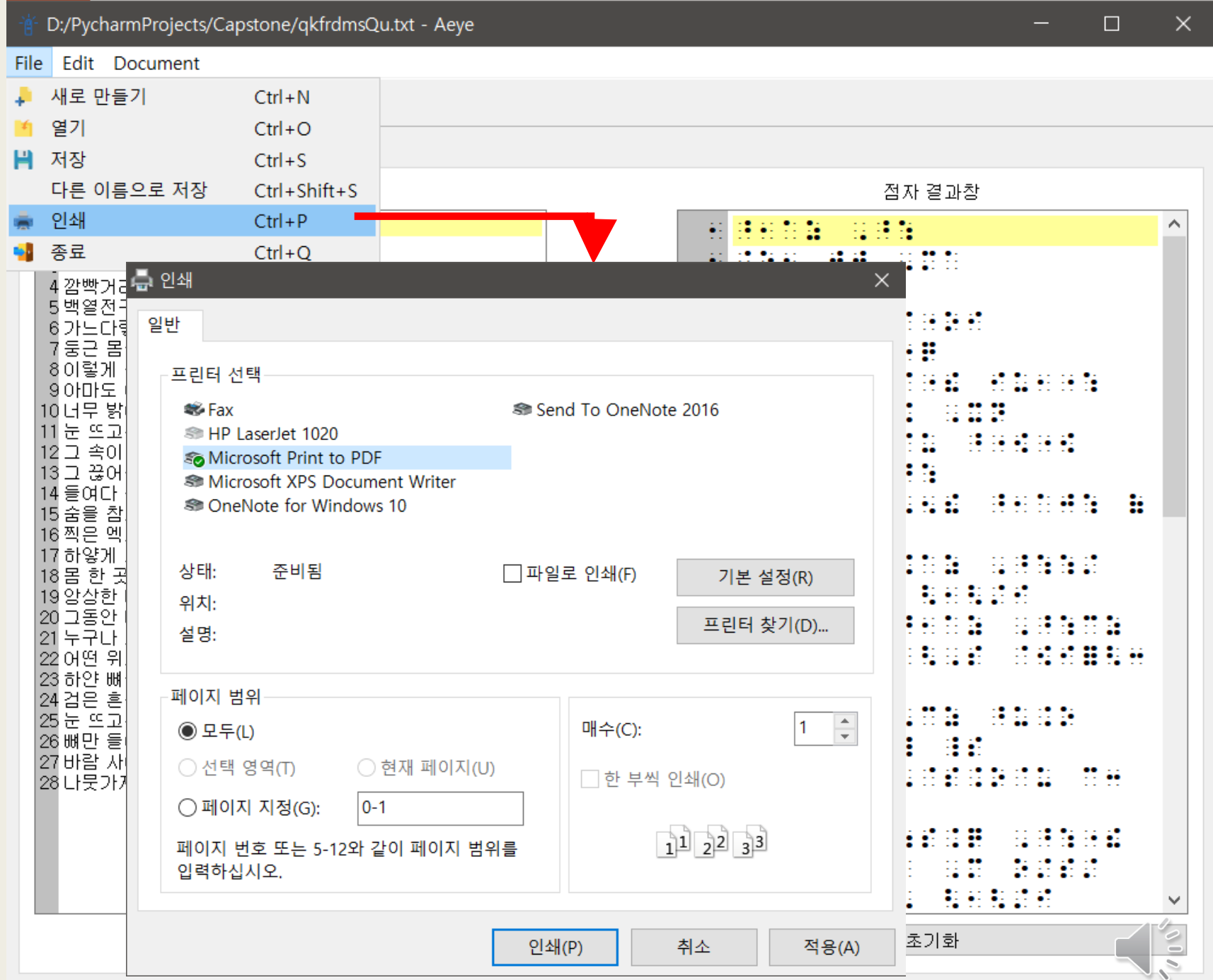


내 PC에 파일을 다른 이름으로 저장합니다.

DEMO

세부 기능 (R-006)

- 점역 파일
인쇄



점역결과를 프린터합니다.

4. 사용자 평가

- 사용자-시각장애인 1급 남성

- 주제에 관해

- 필요성, 당위성은 있을 것 같다. 건강검진 결과표같은 개인정보가 담긴 서류의 경우 요즘은 점자 파일로 바뀌어서 보내주는 등 나날이 시각장애인을 위해 나아지고 있는 편이지만 부족한 곳도 많다. 개인이 프로그램으로 손쉽게 사용할 수 있게 만들어준다는 점에서 고평가를 하고 싶다.

- 사용후기

- 다른 문서들은 잘 되는 것 같고 인쇄된 책 스캔한 이미지에서 문장을 뽑아낼 수 있어서 좋았다. 그리고 오타를 수정할 수 있어서 틀린 부분을 고칠 수 있긴 하지만 많은 페이지수의 책을 생각한다면 정확성은 아쉬운 부분이라 생각한다.



5. 고찰

- 점자가 표현할 수 있는 정보의 한계성으로 한국어와 영어에 점역기능이 한정됨
- 기존 OCR의 한글 인식의 결과를 향상시키려 모델을 학습시켰으나 학습 과정의 부족으로 결과물이 눈에 띄게 좋아지지 않는음



5. 고찰

- 기존 OCR의 한글 인식의 결과를 향상시키려 모델을 학습시켰으나 학습 과정의 부족으로 결과물이 눈에 띄게 좋아지지 않는

```
At iteration 431953/1500000/1500367, Mean rms=0.829%, delta=2.42%,  
wrote checkpoint.
```

```
Finished! Error rate = 1.022
```

```
lstmtraining #
```

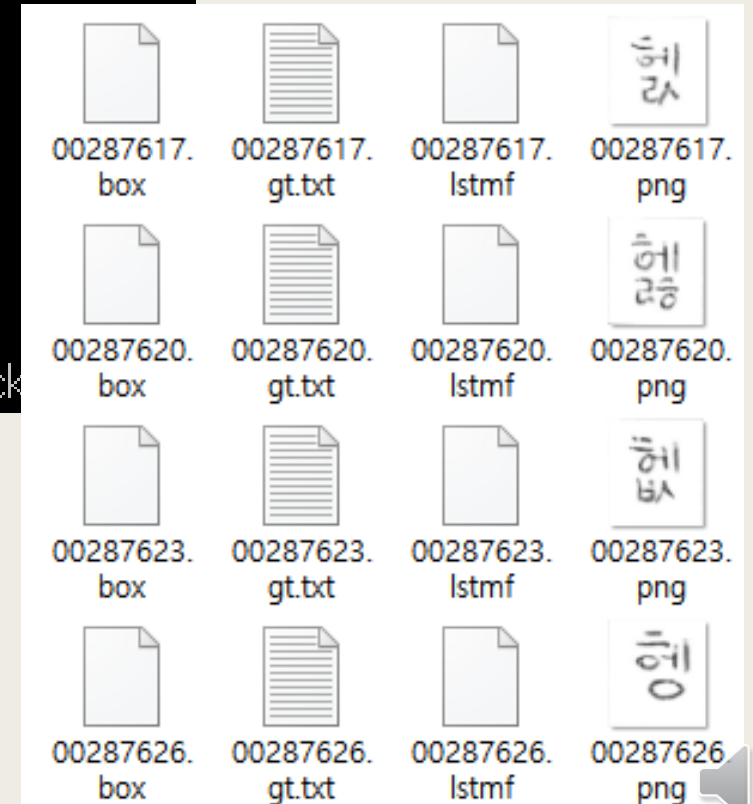
```
--stop_training #
```

```
--continue_from data/small/checkpoints/small_checkpoint #
```

```
--traineddata data/small/small.traineddata #
```

```
--model_output data/small.traineddata
```

```
Loaded file data/small/checkpoints/small_checkpoint, unpack
```



5. 고찰

- 기존 OCR의 한글 인식의 결과를 향상시키려 모델을 학습시켰으나 학습 과정의 부족으로 결과물이 눈에 띄게 좋아지지는 않음

예비역 육군대위 윤영주(43·가명)씨는 여군사관을 거쳐 2000년대초 소위 계급장을 달던 순간의 떨림을 잊지 못한다. "여자가 웬 군인이냐"고 주변에서 말렸지만 남들과 다른 삶을 살고 싶어 선택한 길이었다. 윤씨는 2017년 7월까지 15년 1개월 간 용인 김포 양평 서울 철원 등지의 부대에서 중대장과 참모로 일했다. 복무기간 중 표창장 등 크고 작은 상을 29차례 수상할 만큼 군 생활도 모범적이었다고 자부했지만, 소령 진급에 실패해 원치 않게 군복을 벗어야 했다. 그는 "첫째와 둘째 아이를 낳았을 때 미더두타매 외 도고우라가물때 뜨머그산 올 그저츠채헛매 조고 외 재극 랑 마 환 우 르

뭉 뽀지 돈 존 다 내 까 자 된 크 된 미 나 고 구 원 에서 맘 구 환 확 등 파 다 롱 하 롱 감 그

원 택 핀 퀴 미 머 다 올 때 호 과 여 견 :

예비역 육군대위 윤영주(43·가명)씨는 여군사관을 거쳐 2000년대초 소위 계급장을 달던 순간에서 급 대 감 과 참 모 도 일 티 다 녹 을 잊지 못한다. "여자가 웬 군인이냐"고 주변에서 말렸지만 남들과 다른 삶을 살고 싶

들 도 모 복 퍼 미 아 그 자 루 퀴 지 아 선택한 길이었다. 윤씨는 2017년 7월까지 15년 1개월 간 용인 김포 양평 철원 등)

나 모 호 라 나 나 도 오 나 르 나 으 쿠 에서 중대장과 참모로 일했다. 복무기간 중 표창장 등 크고 작은 상을 29차례 수상할 만큼 군

무 롬 챔 줘 지 요 품 호 모 그 지 들 보 활도 모범적이었다고 자부했지만, 소령 진급에 실패해 원치 않게 군복을 벗어야 사스 귀

냥 첫째와 둘째 아이를 낳았을 때는 육아휴직도 가지 않고 출산휴가 3개월만 쓰고

학습시킨 모델

무를 챙겼지만, 출산으로 자리를 비운 것이 진급심사에 영향을 준 것 아닌가 1위라고 했다.

복 기존 kor.traineddata



6. QnA



감사합니다.

