
사이냅 OCR 서식 편집 매뉴얼

v2.0.2210



목차

사이넵 OCR 폼메이커 시작하기	1
개요	1
제품 구성 및 시스템 요구사항	3
사이넵 OCR 폼메이커 설치하기	4
의존성 설치	4
설치	4
디렉토리 구조	8
삭제	9
사이넵 OCR 폼메이커 – 사용 가이드	10
화면 구성	10
상단 메뉴	11
새로 만들기	15
엑셀 파일 열기, 엔진에서 서식 데이터 불러오기	16
샘플 이미지 지정하기 – 일반 이미지	17
샘플 이미지 지정하기 – PDF, TIFF	18
서식 데이터 편집하기	20
서식 데이터 테스트하기	32
서식 저장하기	35
서식 배치 테스트하기	43
사이넵 OCR 폼메이커 운영	48
로그 확인	48
수동 업데이트 및 테스트	49
단축키 가이드	57

사이냅 OCR 폼메이커 시작하기

개요

소개

사이냅 OCR 폼메이커는 사이냅 OCR에서 사용하는 서식 템플릿을 작성할 수 있는 GUI 편집 도구입니다. 사이냅 OCR의 사용권을 획득한 사용자는 사이냅 OCR 폼메이커를 통해 사용자 정의 서식 템플릿을 작성할 수 있습니다.

제품의 저작권

사이냅 OCR에 포함된 모든 프로그램과 부속물에 대한 저작권 및 지적소유권은 (주)사이냅소프트가 가지며, 이 권리는 대한민국의 저작권법과 국제 저작권 조약으로 보호받습니다.

(주)사이냅소프트의 사전 서면 동의 없이 프로그램이나 부속물의 내용을 수정, 변형, 복사할 수 없습니다.

사이냅 OCR에 포함된 프로그램이나 부속물을 사용하는 것은 사용자가 정상적인 유통 경로에 의하여 제품의 사용권을 획득하였음을 인정하는 것입니다.

책임의 한계와 법적 고지

(주)사이냅소프트는 사이냅 OCR에 포함된 프로그램 및 부속물이 고객의 특정 목적에 적합할 것이라는 보증을 하지 않으며, 본 프로그램의 사용으로 인하여 초래된 결과에 책임지지 않습니다.

제품 설치 수행자 정보

본 제품은 구매하신 사용자가 직접 설치 할 수 있도록 구성되어 있습니다. 설치 중 발생하는 문의사항은 전자우편주소 support@synapsoft.co.kr 으로 문의하여 주십시오.

고객지원

제품 구매 후 1 년 동안 무상으로 전화 및 온라인 기술 지원 서비스를 받으실 수 있습니다.

- 회사 주소: (07805) 서울특별시 강서구 마곡중앙4로 62
- 대표 전화: (02) 2039-3780
- 홈페이지: <https://www.synapsoft.co.kr>
- 전자우편주소: support@synapsoft.co.kr

제품 문의처

본 제품을 사용하는 중에 발생하는 문의사항은 (주)사이냅소프트 홈페이지를 방문하시거나,
전자우편주소 support@synapsoft.co.kr 으로 문의하여 주십시오.

제품 구성 및 시스템 요구사항

제품 구성

사이냅 OCR 프로는 아래와 같은 내용물로 구성되어있으며 사이냅 OCR 엔진과 연동되어 기능이 제공됩니다.

- 라이선스 파일 - 엔진 구동 및 SDK 수행 시 필요한 API 키
- 사이냅 OCR 엔진 - 실제 OCR 동작을 수행하며 REST API를 제공하는 데몬 프로세스
- 사이냅 OCR OCR2PDF - OCR 엔진과 연동하여 이미지 파일을 PDF로 변환하는 프로그램
- 폼메이커 - OCR 엔진과 연동하여 서식 템플릿을 작성하는 프로그램
- 사이냅 OCR SDK - OCR 엔진과 통신하여 어플리케이션 제작 시 사용할 수 있는 라이브러리 및 실행 파일
- 제품 매뉴얼 - 구성품의 설치 및 운영 가이드, REST API 가이드 제공

시스템 요구사항

폼메이커는 아래와 같은 환경에서 구동 및 사용이 가능합니다.

지원 운영 체제

- Windows 10 x86_64, Windows 2019 Server x86_64, Windows 2016 Server x86_64
- 기타 운영체제 별도 문의 필요

동작 환경

- Intel i5 이상의 x86 기반 CPU 아키텍처
- 최소 8GB 이상의 메인 메모리, 16GB 이상의 메인 메모리 권장
- 설치 및 운영을 위한 3GB 이상의 하드디스크 여유 공간 필요

지원 포맷

- BMP - Bitmap
- PNG - Portable Network Graphics
- JPG / JPEG - Joint Photographic Experts Group
- TIF / TIFF - Tagged Image File Format (old-style JPEG 제외)
- PDF - Portable Document Format (암호화된 PDF 제외)
- GIF - GIF89a, GIF87a, 애니메이션 GIF의 경우 첫 번째 프레임만 처리

제약 사항

- 이미지가 사이즈가 크거나 텍스트가 많은 경우 서식 매칭 시간이 길어질 수 있습니다.

사이넵 OCR 폼메이커 설치하기

의존성 설치

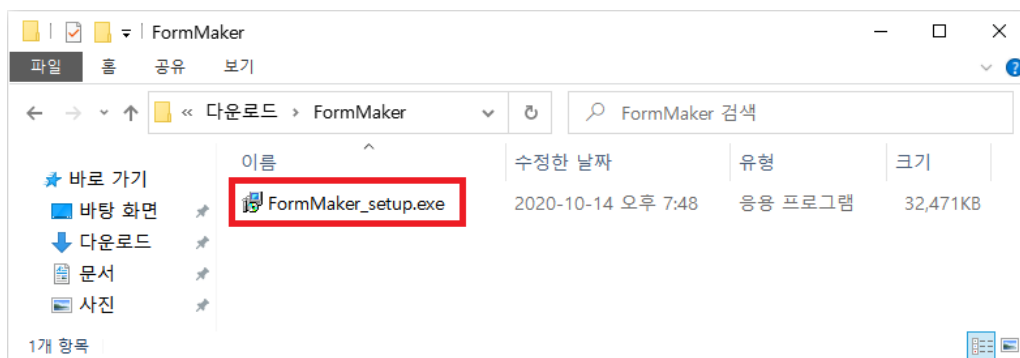
- 사이넵 OCR 엔진 2020.03 버전 이상이 설치되어 있어야합니다.

설치

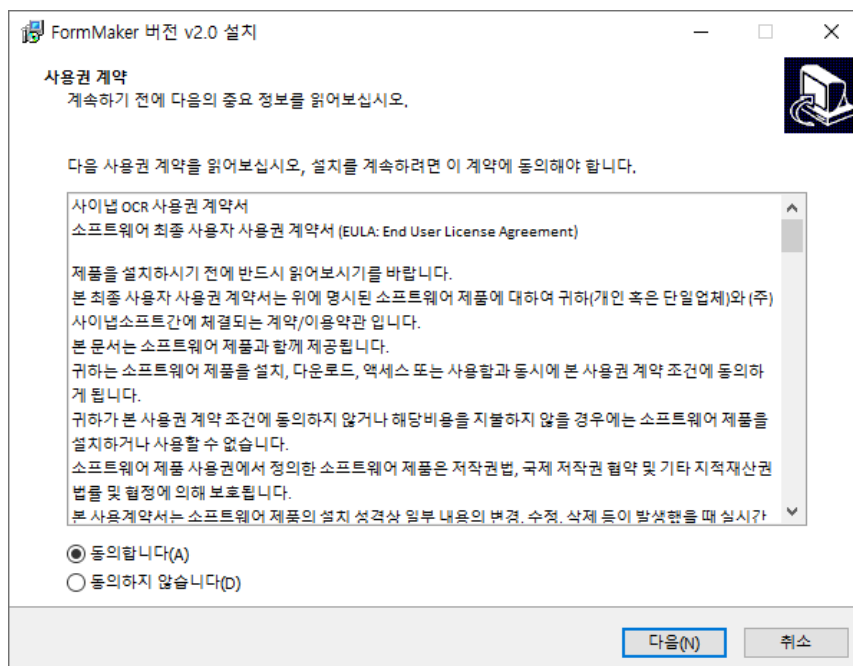
인스톨러 파일을 이용하여 폼메이커를 설치할 수 있습니다.

제공받은 압축 파일을 해제하면 아래와 같이 인스톨러 파일 FormMaker_setup.exe 가 생성됩니다.

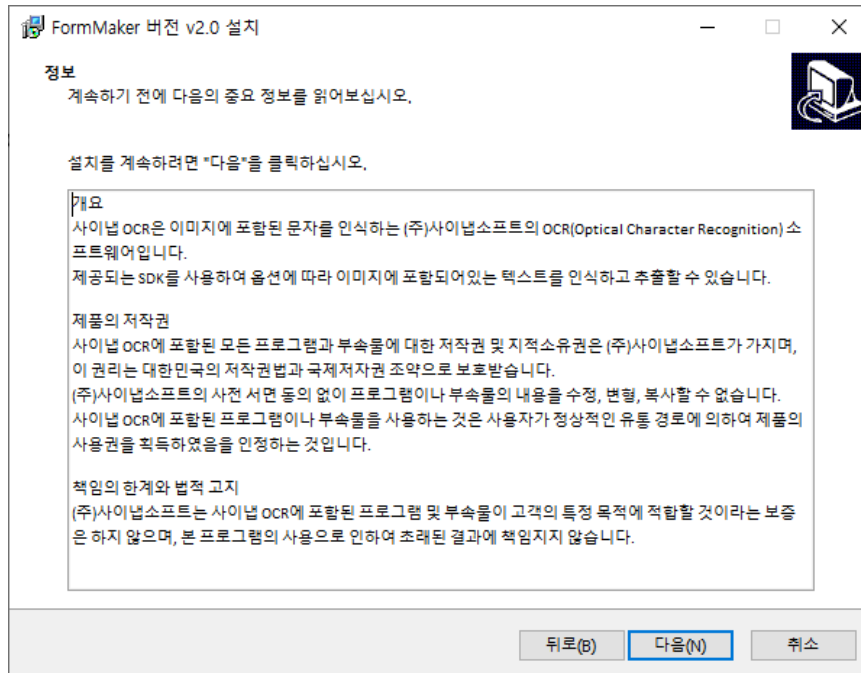
FormMaker_setup.exe 파일을 실행합니다.



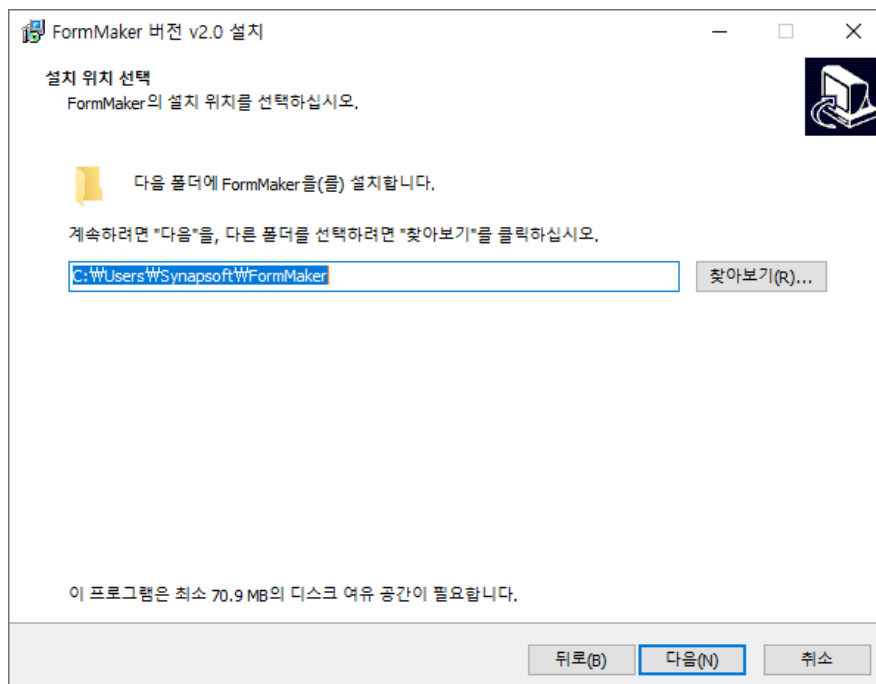
사용권 계약 내용과 계약 동의 여부를 확인합니다. '동의합니다'에 체크하고 다음 버튼을 클릭합니다.



사이넵 OCR 라이선스 정보입니다. 정보 확인 후 다음 버튼을 클릭합니다.

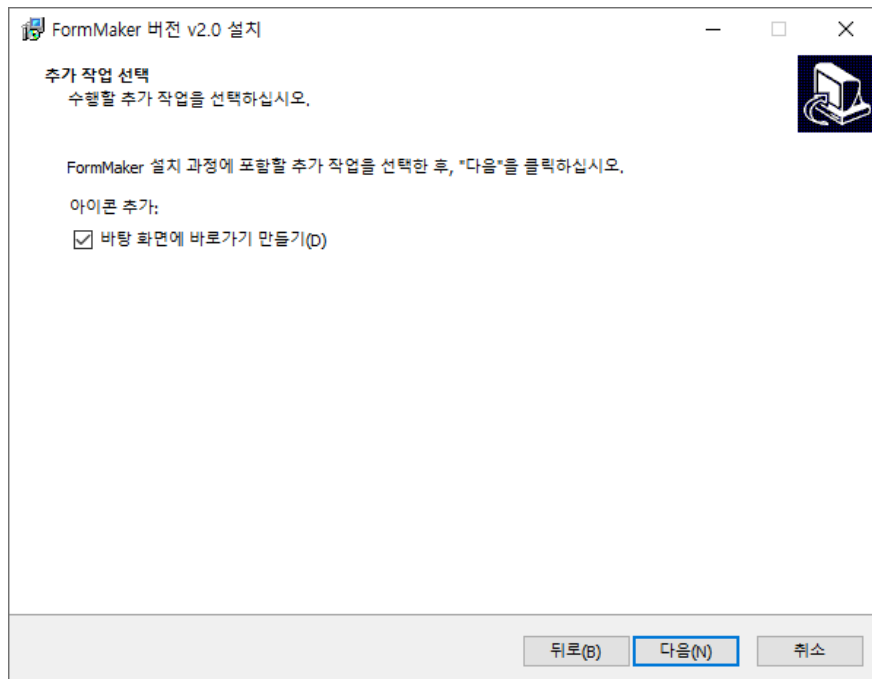


사이냅 OCR 폼메이커를 설치할 경로를 지정합니다. 기본값은 사용자 폴더의 하위 폴더에 설치됩니다. 경로 지정 후 다음 버튼을 클릭합니다.

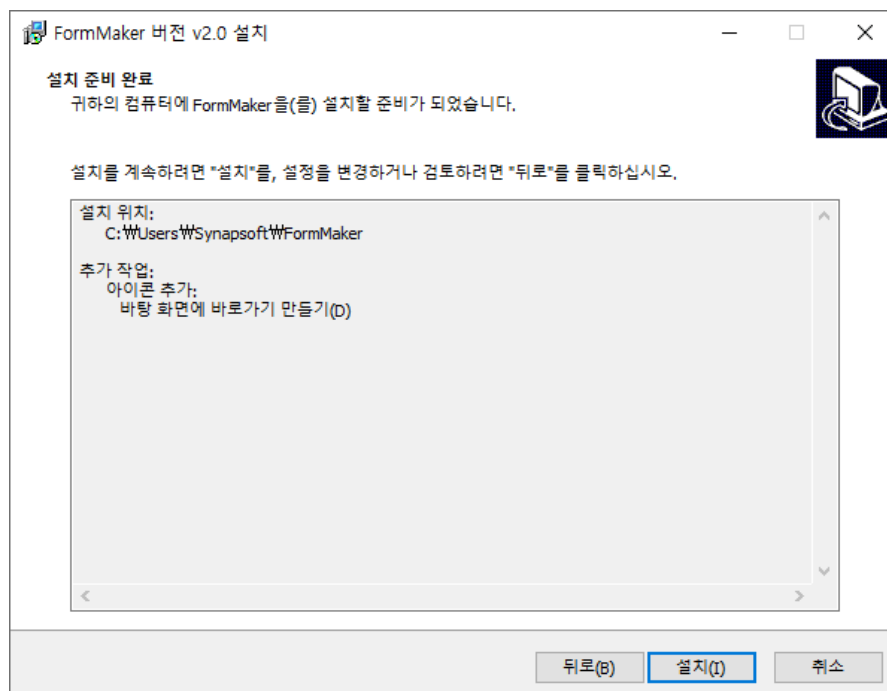


바탕화면에 바로가기 생성 여부를 지정합니다. 체크할 경우 설치 후 바탕화면에 폼메이커 바로가

기가 생성됩니다. 생성 여부 지정 후 다음 버튼을 클릭합니다.

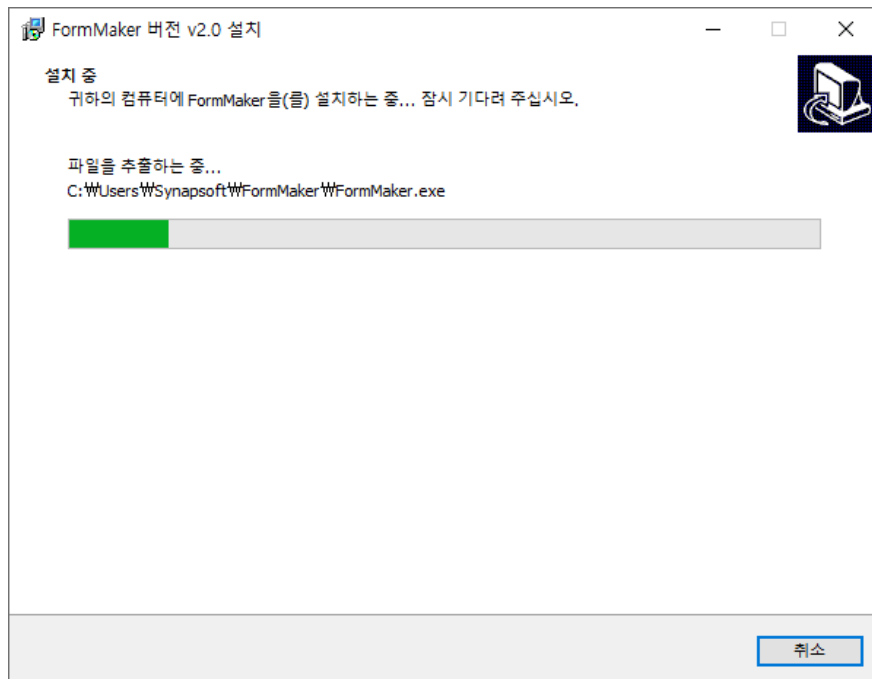


지정한 내용을 설치 전에 점검합니다. 잘못 지정한 내용이 있을 경우 뒤로 버튼을 클릭하여 되돌아갑니다. 내용 확인 후 설치 버튼을 클릭하면 설치가 진행됩니다.

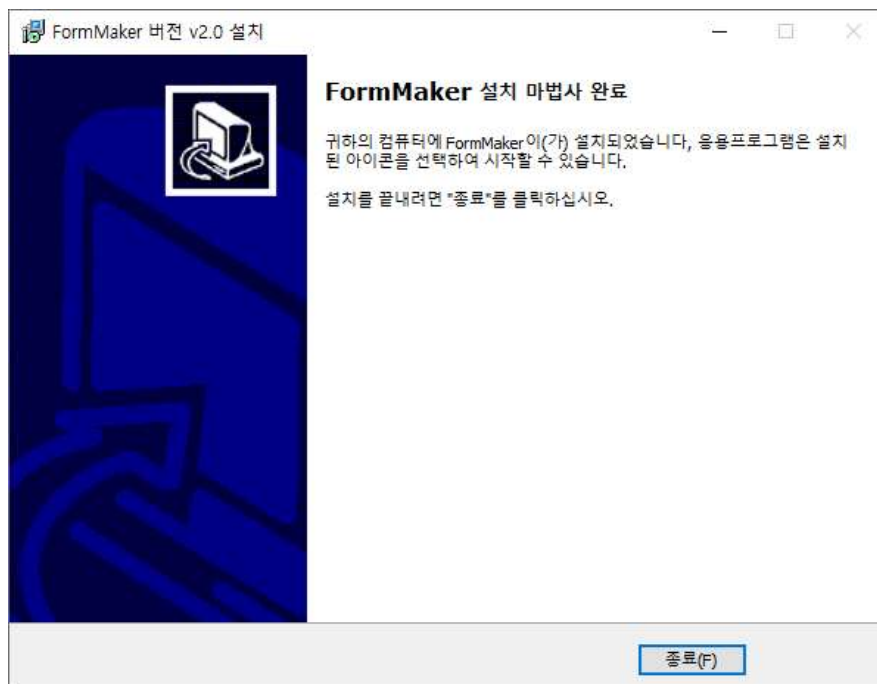


설치 진행 내용 및 진행도를 확인할 수 있습니다. 설치가 완료되면 자동으로 다음 페이지로 전환

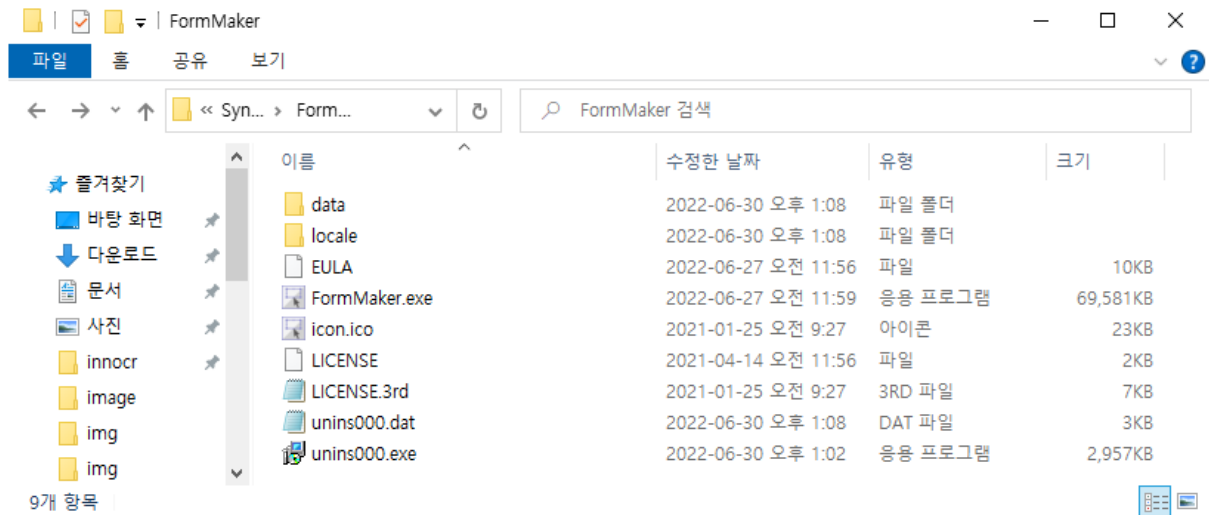
됩니다.



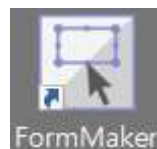
설치가 완료되었습니다. 종료 버튼을 클릭하여 인스톨러를 종료합니다.



설치 완료 후 지정한 경로에 설치된 것을 확인할 수 있습니다.



바탕화면 바로가기 만들기 옵션을 지정했을 경우 아래와 같이 폼메이커 바로가기가 생성됩니다. 설치 경로의 폼메이커.exe 또는 바로가기를 이용하여 폼메이커를 실행할 수 있습니다.



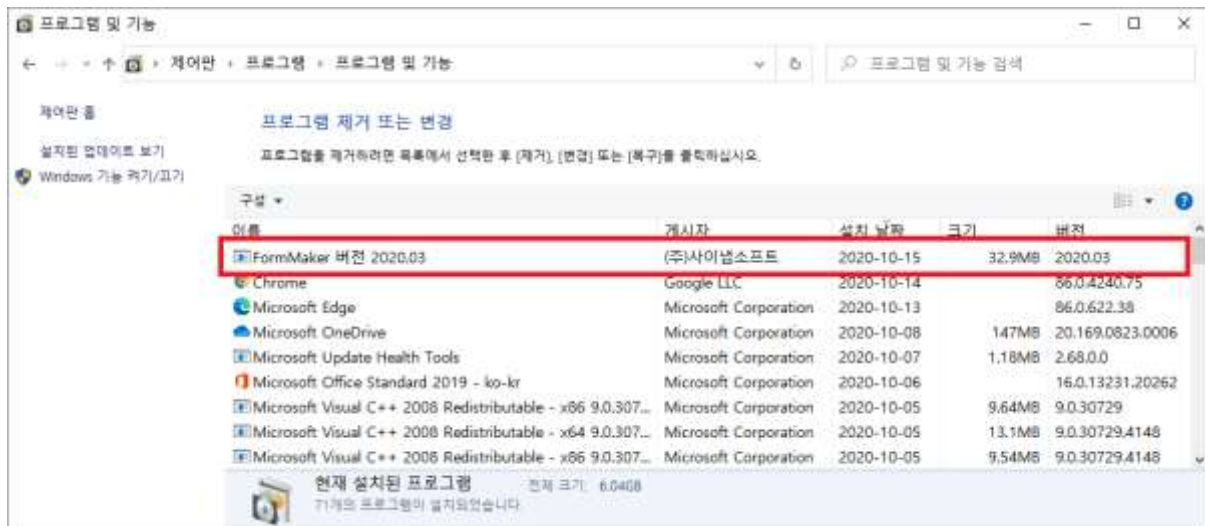
디렉토리 구조

설치된 사이냅 OCR 폼메이커의 구성은 아래와 같습니다.

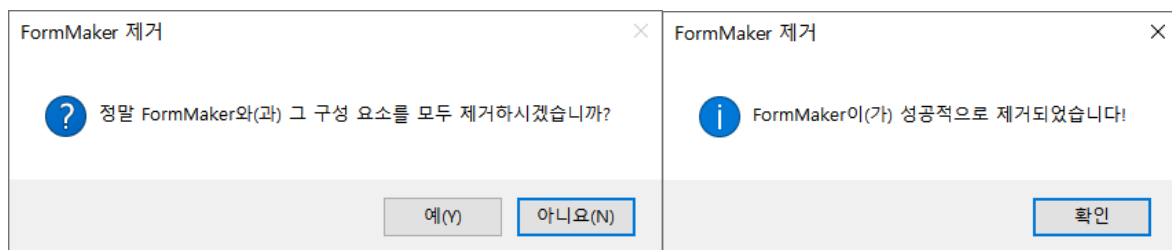
data	# 로그, 모델, 임시파일 등 FormMaker에서 관리하는 파일
├ image	# 서식 타겟 이미지가 저장되는 폴더
├ logs	# FormMaker의 로그가 쓰여지는 파일
├ result	# 서식 인식 테스트 결과 파일이 저장되는 폴더
├ temp	# FormMaker이 실행되는 도중 임시로 사용하는 폴더
├ connection_info	# 연결된 OCR 엔진 정보가 쓰여지는 파일
└ DroidSansFallback.ttf	# FormMaker 배치 테스트 결과 이미지 생성 시 사용되는 폰트
locale	# locale 정보
EULA	# 사용권계약서
FormMaker.exe	# FormMaker 실행파일
icon.ico	# FormMaker 아이콘파일
LICENSE	# 사용권 지침
LICENSE.3rd	# 3rd. party 사용권 지침
unins000.dat	# FormMaker 삭제에 사용되는 데이터 파일
unins000.exe	# FormMaker 삭제에 사용되는 실행 파일

삭제

"제어판>프로그램>프로그램 및 기능"에서 사이냅 OCR 폼메이커를 삭제할 수 있습니다.



해당 항목을 우클릭하여 '제거'를 클릭하면 제거 확인 여부를 묻는 창이 나타나고, '예' 버튼 클릭 시 삭제가 진행됩니다.



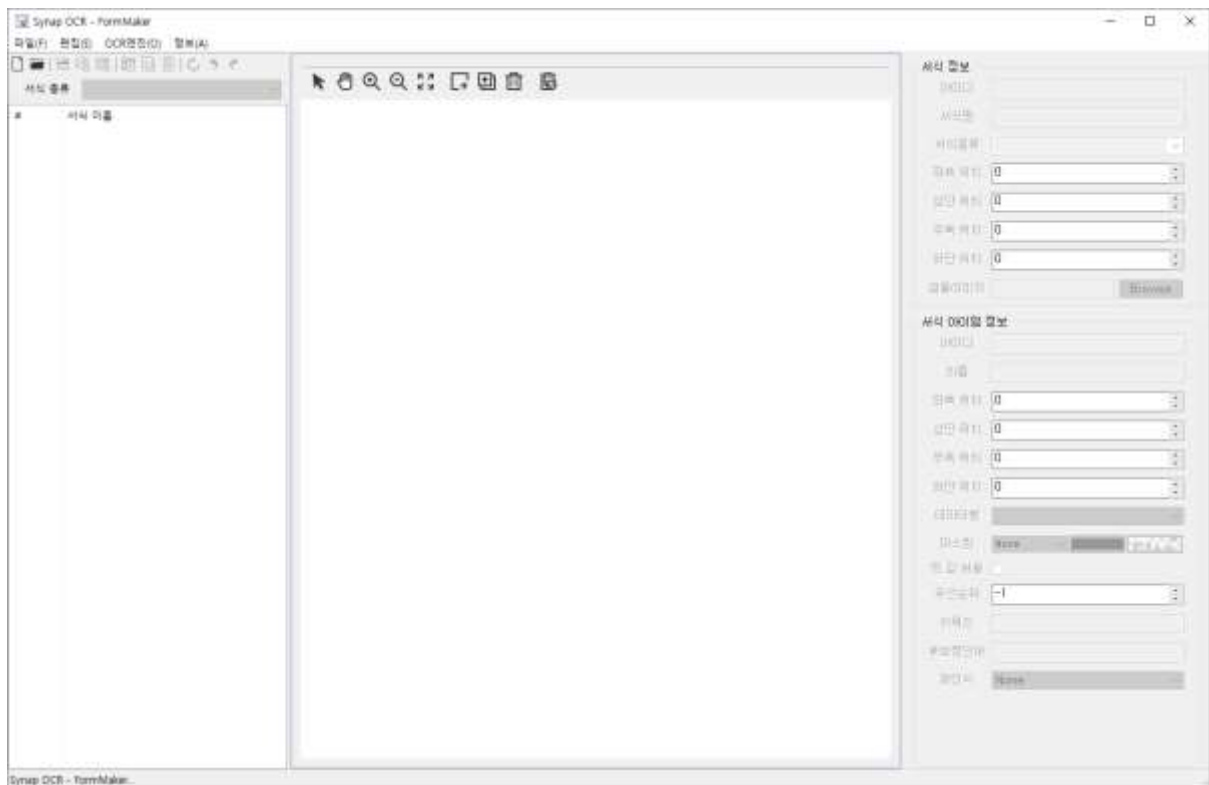
"제어판>프로그램>프로그램 및 기능"에서 폼메이커를 제거해도 로그 파일 같은 임시 파일 일부가 설치된 경로에 남아있을 수 있습니다.

완전하게 제거하기 위해서는 설치된 경로에서 "FormMaker" 폴더를 제거해야 합니다.

사이냅 OCR 폼메이커 – 사용 가이드

화면 구성

시작 화면은 아래와 같습니다. 화면은 왼쪽부터 컨트롤 패널, 캔버스 패널, 속성 패널로 구성되어 있습니다.



- 컨트롤 패널에는 새 서식, 열기, 저장, 다른 이름 엑셀로 저장, 엔진으로 저장, 서식 추가/복사/제거, 다시 불러오기, 되돌리기, 다시실행 버튼과 서식 종류 선택란이 있으며, 작업 중인 서식 목록이 출력됩니다.
- 캔버스 패널에는 편집에 사용되는 도구 모음(포인터, 패닝, 확대, 축소, 화면맞춤)과 서식 아이템 추가/복사/제거, 서식 테스트 버튼이 있으며, 작업 영역에는 편집 가능한 서식 영역이 표시됩니다. 우측 속성 패널에서 샘플 이미지를 선택하여 이미지를 보면서 서식 편집을 수행할 수 있습니다.
- 속성 패널에는 작업 중인 서식 데이터의 속성이 출력됩니다.
캔버스 패널에서 작업한 내용이 속성 패널에 반영되며, 속성 패널에서 값을 수정할 경우 캔버스 패널에 반영됩니다.






상단 메뉴

패널 상단에는 기본 메뉴들이 있습니다. 순서대로 파일, 편집, OCR 엔진, 정보가 있습니다.

파일

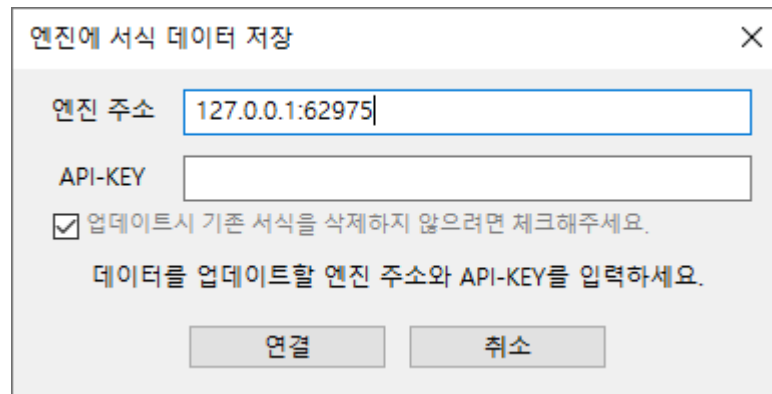
새로 만들기, 열기, 저장 등의 메뉴가 포함되어 있습니다.

파일(F)	편집(E)	OCR엔진(O)	정보(A)
새 서식(N)			Ctrl+N
열기(O)			Ctrl+O
저장(S)			Ctrl+S
다른 이름 엑셀로 저장			Ctrl+Shift+S
엔진으로 저장			Ctrl+Shift+E
종료(E)			Ctrl+Q

- 새 서식
 - 새로 만들기(단축키 Ctrl+N)
 - 컨트롤 패널의  버튼과 동일하게 동작합니다.
- 열기
 - 서식 엑셀 파일 열기(단축키 Ctrl+O)
 - 컨트롤 패널의  버튼과 동일하게 동작합니다.
- 저장
 - 저장하기(단축키 Ctrl+S)
 - 엑셀파일을 열었을 때는 엑셀파일 저장이 수행되고
엔진에서 서식을 불러왔을 때는 엔진으로 저장이 수행됩니다.
 - 컨트롤 패널의  버튼과 동일하게 동작합니다.
- 다른 이름 엑셀로 저장
 - 엑셀 파일로 저장하기(단축키 Ctrl+Shift+S)
 - 컨트롤 패널의  버튼과 동일하게 동작합니다.
- 엔진으로 저장
 - 엔진에 업데이트하기(단축키 Ctrl+Shift+E)
 - 컨트롤 패널의  버튼과 동일하게 동작합니다.

데이터를 업데이트할 엔진 주소와 API KEY를 입력하는 다이얼로그가 나타납니다.
엔진 주소와 API KEY를 입력하고 연결 버튼을 선택하면 서식 저장이 수행됩니다.
체크를 해제하고 저장할 경우 서식 삭제 정보도 함께 엔진에 저장됩니다. 서식 삭제는 폼메이커에서 삭제한 이력이 있는 서식 아이디와 동일한 서식이 엔진에서

삭제됩니다. 따라서 엔진에 등록된 서식을 삭제하기 위해서는 엔진에 등록된 서식 아이디와 동일한 서식아이디를 갖는 서식을 폼메이커에서 삭제한 이력이 있어야 삭제가 가능합니다.



엔진에 서식 데이터 저장

엔진 주소: 127.0.0.1:62975

API-KEY:

☒ 업데이트시 기존 서식을 삭제하지 않으려면 체크해주세요.

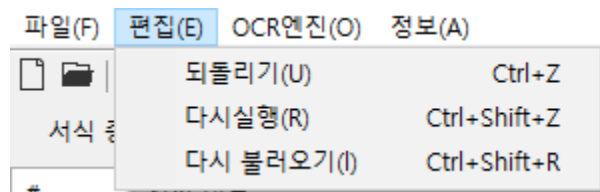
데이터를 업데이트할 엔진 주소와 API-KEY를 입력하세요.

연결 취소

- 다이얼로그 내 각 입력 항목은 아래와 같은 제약사항이 있습니다.
 - 엔진 주소: URL 형식으로 작성되며 프로토콜 생략 시 'http://'로 지정됩니다.
 - API-KEY: 영문, 숫자, 하이픈만 포함하며 8~64글자로 이루어져야 합니다.
- 종료
 - 서식 설정 도구 종료하기(단축키 Ctrl+Q)

편집

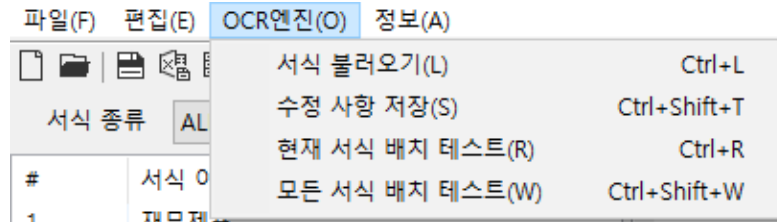
되돌리기/다시실행, 다시 불러오기 등 편집에 관련된 메뉴가 포함되어 있습니다.



- 되돌리기
 - 이전 작업 내용을 취소합니다(단축키 Ctrl+Z)
 - 컨트롤 패널의 ↶ 버튼과 동일하게 동작합니다.
- 다시 실행
 - 작업 취소한 내용을 되돌립니다(단축키 Ctrl+Shift+Z)
 - 컨트롤 패널의 ↷ 버튼과 동일하게 동작합니다.
- 다시 불러오기
 - 서식 데이터를 다시 불러와서 작업한 내용을 모두 초기화합니다(단축키 Ctrl+Shift+R)
 - 컨트롤 패널의 ↺ 버튼과 동일하게 동작합니다.

OCR엔진

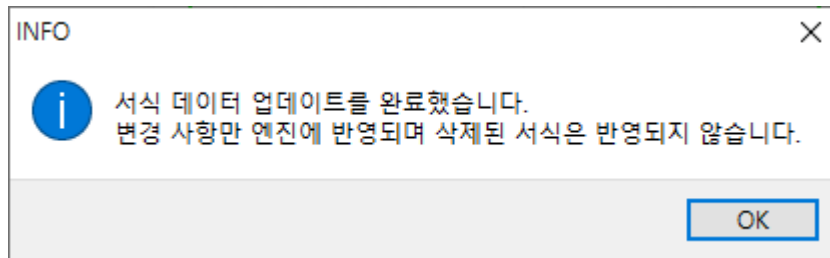
엔진에서 서식 불러오기, 엔진에 수정 사항 저장 등의 메뉴를 포함합니다.



- 서식 불러오기
 - 엔진에서 서식 데이터를 불러옵니다(단축키 Ctrl+L)
 - 데이터를 불러올 엔진 주소와 API KEY를 입력하는 다이얼로그가 나타납니다.

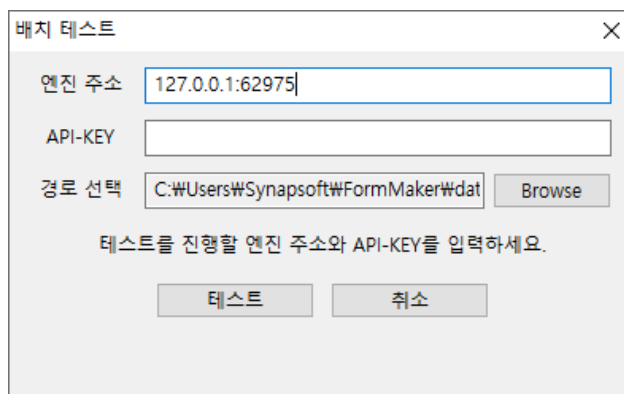
- 다이얼로그 내 각 입력 항목은 아래와 같은 제약사항이 있습니다.
 - 엔진 주소: URL 형식으로 작성되며 프로토콜 생략 시 'http://'로 지정됩니다.
 - API-KEY: 영문, 숫자, 하이픈만 포함하며 8~64글자로 이루어져야 합니다.
- 수정 사항 저장
 - 서식 데이터를 엔진에 업데이트합니다(단축키 Ctrl+Shift+T)
 - 연결할 엔진 정보가 없는 경우(서식을 불러오지 않거나 서식을 엔진에 저장한 이력이 없는 경우) 파일 메뉴의 엔진으로 저장 메뉴와 동일한 다이얼로그가 나타납니다.

- 다이얼로그 내 각 입력 항목은 아래와 같은 제약사항이 있습니다.
 - 엔진 주소: URL 형식으로 작성되며 프로토콜 생략 시 'http://'로 지정됩니다.
 - API-KEY: 영문, 숫자, 하이픈만 포함하며 8~64글자로 이루어져야 합니다.
- 서식 불러오기나 저장을 통해 엔진에 연결한 이후에는 다이얼로그가 나타나지 않고 엔진에 바로 업데이트됩니다. 이 경우에는 서식 삭제 내역은 엔진에 반영되지 않고 서식 편집, 추가 내역만 업데이트 됩니다.



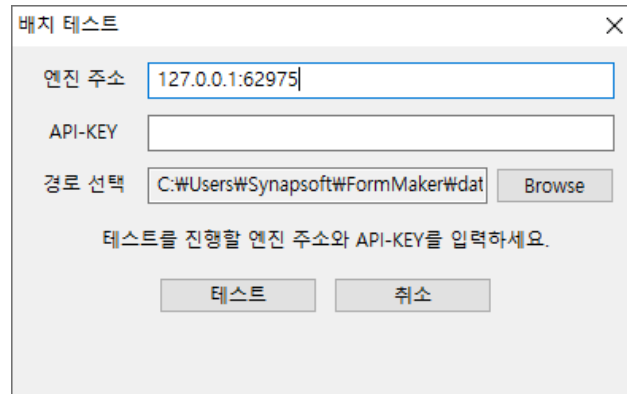
- 현재 서식 배치 테스트
 - 현재 선택된 서식을 이용해 배치 테스트를 수행합니다(단축키 Ctrl+R)
 - 서식 테스트를 수행할 엔진 주소와 API KEY, 테스트할 이미지들이 있는 경로를 입력할 다이얼로그가 나타납니다.

자세한 내용은 '서식 배치 테스트하기' 목차를 참고 부탁드립니다.



- 다이얼로그 내 각 입력 항목은 아래와 같은 제약사항이 있습니다.
 - 엔진 주소: URL 형식으로 작성되며 프로토콜 생략 시 'http://'로 지정됩니다.
 - API-KEY: 영문, 숫자, 하이픈만 포함하며 8~64글자로 이루어져야 합니다.
- 모든 서식 배치 테스트
 - 엔진에 등록된 모든 서식을 이용해 배치 테스트를 수행합니다(단축키 Ctrl+Shift+W)
 - 서식 테스트를 수행할 엔진 주소와 API KEY, 테스트할 이미지들이 있는 경로를 입력할 다이얼로그가 나타납니다.

자세한 내용은 '서식 배치 테스트하기' 목차를 참고 부탁드립니다.



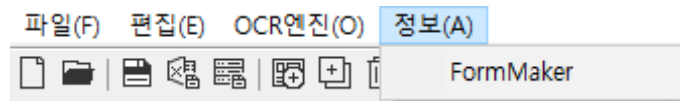
The 'Batch Test' dialog box contains the following fields and controls:

- 엔진 주소 (Engine Address):** A text input field containing '127.0.0.1:62975'.
- API-KEY:** An empty text input field.
- 경로 선택 (Path Selection):** A text input field containing 'C:\Users\Synapsoft\FormMaker\dat' and a 'Browse' button.
- Instructions:** A line of text stating '테스트를 진행할 엔진 주소와 API-KEY를 입력하세요.' (Please enter the engine address and API-KEY to proceed with the test).
- Buttons:** '테스트' (Test) and '취소' (Cancel) buttons at the bottom.

- 다이어로그 내 각 입력 항목은 아래와 같은 제약사항이 있습니다.
 - 엔진 주소: URL 형식으로 작성되며 프로토콜 생략 시 'http://'로 지정됩니다.
 - API-KEY: 영문, 숫자, 하이픈만 포함하며 8~64글자로 이루어져야 합니다.

정보

폼메이커 버전 정보를 제공하는 메뉴가 포함됩니다.



- 폼메이커: 제품 버전 정보를 제공합니다.

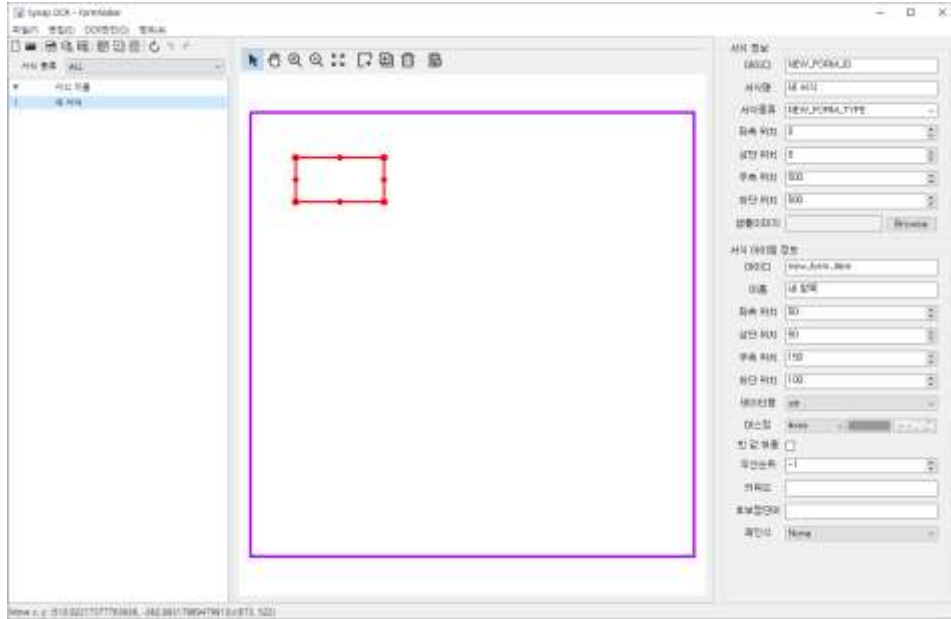


새로 만들기

새로 만들기를 실행하면 아래와 같이 새 서식 데이터가 생성됩니다.

컨트롤 패널에 '새 서식' 항목이 추가되고, 캔버스 패널에 기본 서식 영역(보라색 박스)과 기본 서식 아이템(빨간색 박스)이 출력됩니다.

속성 패널에는 기본 서식 정보와 서식 아이템 정보가 출력됩니다. 서식 정보와 아이템 정보는 사용자가 직접 수정할 수 있습니다.



엑셀 파일 열기, 엔진에서 서식 데이터 불러오기

서식 엑셀 파일을 열거나 엔진에서 서식 데이터를 불러오면 아래와 같이 불러온 서식 데이터가 출력됩니다.

기본 예제 서식이 제공되며, 예제 서식은 {사이냅OCR}/data/form_data.xlsx에서 확인할 수 있습니다. 해당 엑셀 파일에는 시트별로 주민등록증, 운전면허증, 여권, 주민등록등·초본, 사업자등록증, 법인사업자등록증 서식 예제가 포함되어 있습니다. 엑셀 상세 내용은 '서식 저장하기-엑셀 파일로 저장' 목차를 참고 부탁드립니다.

서식 아이템들 중 현재 작업 중인 아이템 박스는 빨간색으로 표시되며, 나머지 아이템 박스들은 초록색으로 표시됩니다.



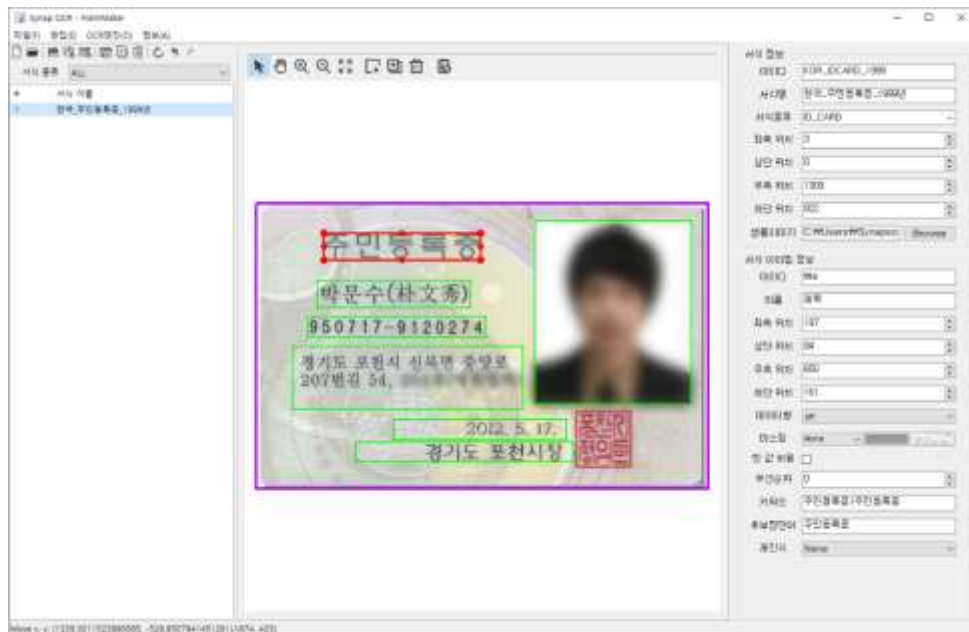
샘플 이미지 지정하기 – 일반 이미지

새 서식 만들기 버튼(📄)을 클릭하여 새 서식 데이터를 생성하거나 편집할 서식 데이터를 불러옵니다. 속성 패널의 서식 정보에서 'Browse' 버튼을 클릭하여 샘플 이미지로 사용할 이미지를 선택하면, 샘플 이미지 지정 다이얼로그가 나타납니다.



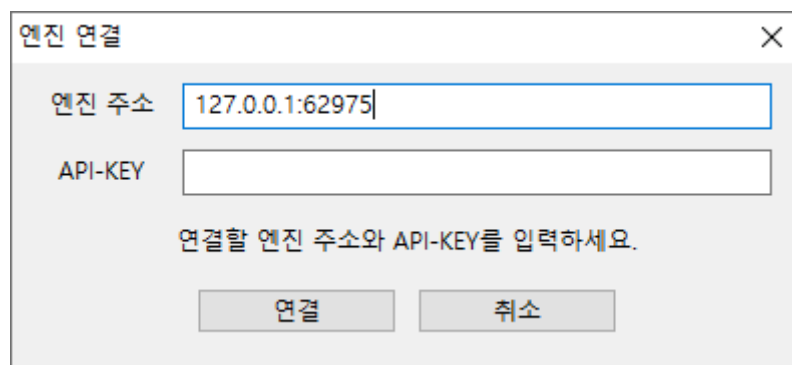
만약 이미지가 회전되어 있다면 다이얼로그 좌측 하단의 슬라이더를 이용하여 이미지의 회전 각도를 보정할 수 있습니다. 서식 샘플 이미지는 정방향으로 설정해야 정확한 서식 테스트 결과를

확인 하실 수 있습니다. 정방향에 수평하게 보정한 후 저장 버튼을 클릭하면, 아래와 같이 캔버스 패널에 지정한 샘플 이미지가 표시됩니다.



샘플 이미지 지정하기 – PDF, TIFF

새 서식 만들기 버튼(📄)을 클릭하여 새 서식 데이터를 생성하거나 편집할 서식 데이터를 불러옵니다. 속성 패널의 서식 정보에서 'Browse' 버튼을 클릭하여 샘플 이미지로 사용할 이미지를 선택하면, 엔진 연결 다이얼로그가 나타납니다.



연결할 엔진 주소와 API KEY를 입력한 후 연결 버튼을 클릭합니다. 엔진과 연결되면 샘플 이미지 지정 다이얼로그가 나타납니다.

다이얼로그 내 각 입력 항목은 아래와 같은 제약사항이 있습니다.

- 엔진 주소: URL 형식으로 작성되며 프로토콜 생략 시 'http://'로 지정됩니다.

- API-KEY: 영문, 숫자, 하이픈만 포함하며 8~64글자로 이루어져야 합니다.



이미지 하단에 앞, 뒤 페이지로 이동할 수 있는 버튼이 있으며, 페이지 번호를 직접 입력하실 수 있습니다. 원하는 페이지를 지정합니다. 만약 이미지가 회전되어 있다면 이미지의 회전 각도를 정방향에 수평하게 보정합니다. 샘플 이미지 지정 후 저장 버튼을 클릭하면, 캔버스 패널에 아래와 같이 지정한 샘플 이미지가 표시됩니다.



서식 데이터 편집하기

서식 데이터는 서식 전체의 설정을 의미하는 서식 정보와 서식 매칭 및 데이터 추출 시 사용되는 서식 아이템으로 이루어져 있습니다. 설정된 서식 정보와 서식 아이템 정보는 OCR 수행 시 입력 받은 이미지의 OCR 결과와 매칭되어 서식 인식에 사용됩니다.

서식 아이템은 반드시 1개 이상 설정되어 있어야 하며 우선순위와 키워드 설정 여부에 따라 서식 매칭 용 서식 아이템과 서식 추출 용 서식 아이템으로 구분됩니다. 서식 매칭 용 서식 아이템은 서식 매칭에 직접적으로 활용되고 서식데이터에 꼭 포함되어야 합니다. 빈 값 허용을 설정하지 않은 서식 추출 용 서식 아이템은 OCR 결과에서 값을 추출하지 못할 경우 서식 신뢰도에 영향을 주고 서식 신뢰도가 일정 수준 이하로 내려가면 서식 매칭이 이루어지지 않습니다.

서식 정보

- 서식을 식별하는데 사용하는 아이디, 서식 명, 서식 종류를 지정할 수 있습니다. 서식 종류를 같은 종류로 지정하면 서식 그룹화된 인식 결과를 받아볼 수 있습니다.
- 이미지 에서 서식의 범위를 나타내는 서식 영역 정보(좌측 위치, 상단 위치, 우측 위치, 하단 위치)를 지정할 수 있습니다.
- 서식 작성 및 테스트에 사용하는 샘플이미지를 지정할 수 있습니다.


서식 아이템 정보

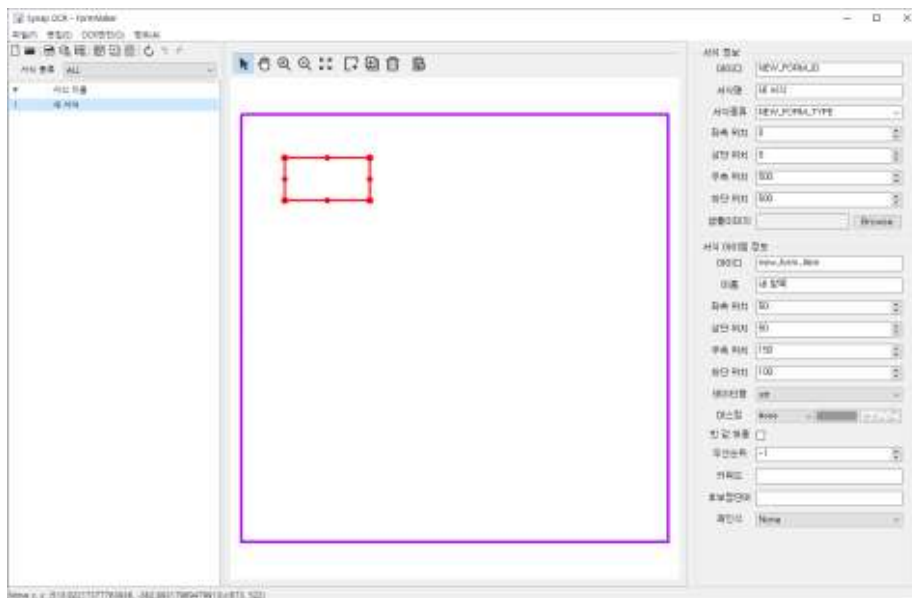
- 각 서식 아이템을 식별하는데 사용하는 아이디, 이름을 지정할 수 있습니다.
- 서식 아이템으로 서식 매칭 또는 서식 추출 시 대상 영역이 되는 서식 아이템 영역 정보

(좌측 위치, 상단 위치, 우측 위치, 하단 위치)를 지정할 수 있습니다

- OCR로 인식한 값을 보정하거나 OCR 추출 대상이 아닌 값을 인식하도록 데이터 타입을 지정할 수 있습니다.
- 서식 아이템 영역을 비 식별화 할 수 있는 마스킹 옵션을 지정할 수 있습니다.
- 해당 아이템에 OCR 인식 결과가 없어도 되는지를 설정하는 빈 값 허용을 지정할 수 있습니다. 빈 값이 허용되는 영역의 경우 서식 매칭 여부 판단 시에 영향을 주지 않는 영역이 되며 빈 값이 들어온 경우 해당 서식 아이템은 인식 결과를 제공하지 않습니다.
- 우선순위와 키워드 설정 여부에 따라 서식 매칭 용 서식 아이템과 서식 추출 용 서식 아이템으로 구분됩니다. 우선순위가 0 이상으로 설정한 경우 서식 매칭에 활용할 수 있는 서식 매칭 용 서식 아이템이 되고 함께 설정한 키워드를 이용해 OCR 인식 결과와 매칭을 수행하게 됩니다.
- 인식 결과에 일부 오 인식이 있을 경우를 대비해 후 보정 단어를 지정하여 인식 결과를 보정할 수 있습니다.
- 정확한 인식이 필요한 영역의 인식률을 높일 수 있도록 서식 아이템 영역 재인식을 지정할 수 있습니다.

서식 편집하기

새 서식 만들기 버튼()을 클릭하여 새 서식 데이터를 생성합니다.




속성 패널의 서식 정보에 만들고 싶은 서식 정보를 입력하고 샘플 이미지를 지정합니다.



서식 정보의 각 항목은 아래와 같으며 적절하게 선택합니다.

- 아이디: 서식 아이디, 서식 구분에 사용되는 아이디 값
 - 영문, 숫자, 하이픈(-), 언더스코어(_)만 허용됩니다.
 - 최대 31 글자까지 지정 가능합니다.
 - 허용하지 않은 문자를 입력할 경우 허용하지 않는 문자는 자동 제거되며, 이때 허용하는 문자가 남아있지 않은 경우 입력 전 값으로 자동 변경됩니다.
 - 필수로 입력해야 하는 항목입니다.
- 서식명: 서식 명칭
 - 최대 40 글자까지 지정 가능합니다.
- 서식 종류: 서식 유형
 - 서식 인식 결과는 서식 종류를 기준으로 그룹핑 되어 제공됩니다.
 - ID_CARD, DRIVE_LICENSE, PASSPORT, FAMILY_REGISTER 네 종류가 예시로 제공되며, 예시 외에 사용자 정의 서식 종류를 입력하실 수 있습니다.
 - 최대 40 글자까지 지정 가능합니다.
- 좌측/상단/우측/하단 위치: 서식 전체 영역의 좌측/상단/우측/하단 좌표
 - 캔버스 패널에서 서식 영역 박스 크기 변경 시 자동으로 반영됩니다.
 - 속성 패널에서 해당 항목 값을 변경 시 캔버스 패널에 자동으로 반영됩니다.
 - 수동 입력이 가능하며 음수 기호(-)와 숫자만 입력 가능합니다.
 - 입력한 값이 서식 아이템 영역보다 작게 입력된 경우 서식 아이템 영역에 맞게 크기가 자동 변경됩니다.
 - -32768 에서 32767 사이의 값으로 입력 가능하고 입력 가능한 값을 넘으면 지정 가능한 범위로 자동 변경됩니다.
 - 정상적이지 않은 형태의 값이 입력될 경우 좌표 값으로 설정 가능한 최솟값으로 자동 변경됩니다.

축소 버튼()을 이용해 캔버스 화면을 축소한 후 서식 영역 박스 크기를 이미지에 맞게 조절합니다.



캔버스의 화면맞춤 버튼(화면맞춤 아이콘)을 클릭하면 아래처럼 캔버스 화면에 딱 차는 비율로 조정됩니다.



기본으로 생성된 서식 아이템 박스를 원하는 위치로 이동하여 크기를 조절하고, 속성 패널에서 서식 아이템 정보를 작성합니다.



서식 아이템 정보의 각 항목은 아래와 같이 작성합니다.

- 아이디: 서식 아이템을 구분하는 아이디
 - 영문, 숫자, 하이픈(-), 언더스코어(_)만 허용됩니다.
 - 최대 40 글자까지 지정 가능합니다.
 - 허용하지 않은 문자를 입력할 경우 입력 칸에서 포커스가 나오거나, 엔터키를 입력했을 때 제거됩니다.
 - 필수로 입력해야 하는 항목입니다.
 - 이름: 서식 아이템의 명칭
 - 최대 40 글자까지 지정 가능합니다.
- ex. 아이디(fid)는 'title', 이름(label)은 '제목'으로 지정



- 좌측/상단/우측/하단 위치: 서식 아이템의 좌측/상단/우측/하단 좌표
 - 캔버스 패널에서 서식 아이템 박스 크기 조절 시 속성 패널에 자동으로 반영됩니다.
 - 속성 패널에서 해당 항목 값 변경 시 캔버스 패널에 자동으로 반영됩니다.

- 속성 패널의 입력란은 미세조정 용도이기 때문에 서식 영역은 캔버스 패널을 통해 지정하시기를 권장해 드립니다.
- 수동 입력이 가능하며 음수 기호(-)와 숫자만 입력 가능합니다.
- 입력한 값이 서식 영역보다 크게 입력된 경우 서식 영역에 맞게 크기가 자동 변경됩니다.
- 상단을 하단보다 큰 값으로 입력하면 하단 크기에 맞게 크기가 자동 변경됩니다.
- 하단을 상단 보다 낮은 값으로 입력하면 상단 크기에 맞게 크기가 자동 변경됩니다.
- 좌측을 우측보다 큰 값으로 입력하면 우측 크기에 맞게 크기가 자동 변경됩니다.
- 우측을 좌측 보다 낮은 값으로 입력하면 좌측 크기에 맞게 크기가 자동 변경됩니다.
- -32768 에서 32767 사이의 값으로 입력 가능하고 입력 가능한 값을 넘으면 지정 가능한 범위로 자동 변경됩니다.
- 정상적이지 않은 형태의 값이 입력될 경우 좌표 값으로 설정 가능한 최솟값으로 자동 변경됩니다.
- 데이터형: 서식 아이템의 데이터 유형
 - num(숫자), char(한 글자), str(문자와 숫자가 섞여있는 텍스트), image(그림 영역), currency(통화 기호가 붙어있는 금액 영역), checkbox(체크박스), table(표)
 - char 로 지정 시
 - 매칭된 영역에 여러 글자가 있어도 한 글자로 보정합니다.
 - 우선순위, 키워드를 지정할 수 없습니다.
 - 데이터 형을 char 로 변경할 경우 우선순위 값이 -1 로 자동 변경됩니다.
 - 데이터 형을 char 로 변경할 경우 키워드 값이 빈 문자열로 자동 변경됩니다.
 - num 으로 지정 시
 - 알파벳 'O', 'o'로 인식된 결과값을 숫자 '0'으로 보정합니다.
 - 알파벳 'I', 'i'로 인식되거나 한글 'ㅣ', 특수문자 '|'로 인식된 결과값을 숫자 '1'로 보정합니다.
 - 우선순위, 키워드를 지정할 수 없습니다.
 - 데이터 형을 num 으로 변경할 경우 우선순위 값이 -1 로 자동 변경됩니다.
 - 데이터 형을 num 으로 변경할 경우 키워드 값이 빈 문자열로 자동 변경됩니다.
 - str 로 지정 시
 - 별도 보정작업을 수행하지 않고 인식 한 결과를 사용합니다.
 - Image 로 지정 시
 - 매칭된 영역에 인식된 글자가 있어도 인식 결과가 없도록 보정합니다.

- 우선순위, 키워드, 빈 값 허용, 후 보정 사전,재인식을 지정할 수 없습니다.
- 데이터 형을 image 로 변경할 경우 우선순위 값이 -1 로 자동 변경됩니다.
- 데이터 형을 image 로 변경할 경우 키워드 값이 빈 문자열로 자동 변경됩니다.
- 데이터 형을 image 로 변경할 경우 빈 값 허용 값이 자동 체크 해제됩니다.
- 데이터 형을 image 로 변경할 경우 후 보정 사전 값이 빈 문자열로 자동 변경됩니다.
- 데이터 형을 image 로 변경할 경우 재인식 값이 빈 none 으로 자동 변경됩니다.
- currency 로 지정 시
 - 알파벳 'S', 's'로 인식된 결과값을 '\$'로, 'W', 'w'로 인식된 결과값을 '₩'로 보정합니다.
 - 우선순위, 키워드를 지정할 수 없습니다.
 - 데이터 형을 currency 로 변경할 경우 우선순위 값이 -1 로 자동 변경됩니다.
 - 데이터 형을 currency 로 변경할 경우 키워드 값이 빈 문자열로 자동 변경됩니다.
- checkbox 로 지정 시
 - 해당 영역의 체크 여부를 반환합니다.
 - 우선순위, 키워드, 빈 값 허용, 후 보정 사전,재인식을 지정할 수 없습니다.
 - 데이터 형을 checkbox 로 변경할 경우 우선순위 값이 -1 로 자동 변경됩니다.
 - 데이터 형을 checkbox 로 변경할 경우 키워드 값이 빈 문자열로 자동 변경됩니다.
 - 데이터 형을 checkbox 로 변경할 경우 빈 값 허용 값이 자동 체크 해제됩니다.
 - 데이터 형을 checkbox 로 변경할 경우 후 보정 사전 값이 빈 문자열로 자동 변경됩니다.
 - 데이터 형을 checkbox 로 변경할 경우 재인식 값이 빈 none 으로 자동 변경됩니다.
- table 로 지정 시
 - 해당 영역의 표를 인식하여 결과를 반환합니다.
 - 우선순위, 키워드, 후 보정 사전,재인식을 지정할 수 없습니다.
 - 데이터 형을 table 로 변경할 경우 우선순위 값이 -1 로 자동 변경됩니다.

- 데이터 형을 table 로 변경할 경우 키워드 값이 빈 문자열로 자동 변경됩니다.
- 데이터 형을 table 로 변경할 경우 후 보정 사전 값이 빈 문자열로 자동 변경됩니다.
- 데이터 형을 table 로 변경할 경우 재인식 값이 빈 none 으로 자동 변경됩니다.
- ex. 데이터형(dtype)을 아이템의 특성에 맞게 지정



- 마스킹: 서식 아이템 영역에 마스킹 할 범위와 방식 지정
 - 마스킹 영역은 none, left, right, full 로 지정할 수 있습니다.
None 지정 시 마스킹 되지 않으며, left, right 지정 시 인식 영역의 절반이 마스킹됩니다.
 - 마스킹 방식은 컬러 마스킹과 모자이크 중 하나를 선택할 수 있습니다.

- ex. 첫 번째처럼 마스킹 지정 시 두 번째처럼 마스킹 결과를 얻을 수 있습니다.



- 빈 값 허용: 빈 내용 허용 여부 지정
 - 해당 영역의 인식된 텍스트가 없어도 되는지 여부를 지정합니다.
 - 우선순위 값이 0 이상인 경우 지정할 수 없습니다.
 - 데이터형이 image 또는 checkbox 인 경우 지정할 수 없습니다.
 - 데이터형을 image 또는 checkbox 로 변경할 경우 빈 값 허용이 자동 체크 해제됩니다.
 - 우선순위 값을 -1 이 아닌 값으로 변경할 경우 빈 값 허용이 자동 체크 해제됩니다
 - ex. 주민등록증의 주소 영역은 총 3 행으로 작성되지만, 실제 주소지에 따라 1 행~3 행으로 변동될 수 있습니다. 따라서 1 행은 반드시 상세 정보가 존재해야

하지만, 2 행과 3 행은 상세 정보에 빈 값이 허용됩니다.



- 우선순위: 서식 종류 구분에 사용되는 상세 영역 여부 지정
 - 0 부터 순차적으로 지정할 수 있으며, 0 이상의 항목이 반드시 한 개 이상 존재해야 합니다.
 - 우선 순위가 0 인 영역의 경우는 서식의 크기 결정에 영향을 주기 때문에 텍스트에 최대한 밀접하게 설정해야 합니다.
 - 같은 우선 순위의 값을 여러 개로 지정이 가능하며 서식 종류 구분에 사용되지 않는 항목은 -1 로 지정합니다.
 - -1 에서 10 사이의 값으로 입력 가능합니다.
 - 데이터형이 str 인 경우에만 지정할 수 있습니다.
 - 데이터형을 str 이 아닌 값으로 변경할 경우 우선 순위 값이 -1 로 자동 변경됩니다.
 - 우선순위 값이 0 인 서식 아이템 하나만 남아있는 경우 해당 서식아이템의 데이터형을 str 이 아닌 다른 타입으로 변경 시도하면 우선순위 값이 -1 로 자동 변경되어 우선순위가 0 인 박스가 남지 않기 때문에 데이터형을 변경할 수 없습니다.
 - 우선순위 값이 0 이상인 경우 빈 값 허용을 지정할 수 없고 키워드를 지정할 수 있습니다.
 - 우선순위 값이 -1 인 경우 키워드를 지정할 수 없고 빈 값 허용을 지정할 수 있습니다.
 - 우선순위 값을 -1 로 변경할 경우 키워드 값이 빈 문자열로 자동 변경됩니다.
 - 우선순위 값을 -1 이 아닌 값으로 변경할 경우 빈 값 허용이 자동 체크 해제됩니다.
 - 동일한 우선순위 값을 갖는 서식 아이템이 없으면서 우선순위 값이 더 큰 서식 아이템이 있는 경우 서식 아이템을 삭제 할 수 없습니다. 이 경우 동일한 우선순위를 갖는 서식 아이템을 추가해야 삭제할 수 있습니다.
 - 숫자 외에 다른 값은 입력할 수 없습니다.

- ex. 서식 종류 구분의 첫 번째 기준이 될 수 있는 '주민등록증'을 priority 0 으로 지정



- 키워드: 서식 종류 구분에 사용되는 상세 영역에 포함되는 텍스트 지정
 - Python 정규식 문법으로 작성합니다.
 - 데이터 형이 str 이며 우선순위 값이 0 이상인 경우 필수로 작성되어야 하는 속성입니다.
 - 데이터형을 str 이 아닌 값으로 변경할 경우 키워드 값이 빈 문자열로 자동 변경됩니다.
 - 우선순위 값을 -1 로 변경할 경우 키워드 값이 빈 문자열로 자동 변경됩니다.
 - OCR 엔진이 인식한 텍스트의 최소 영역(raw box) 단위로 매칭이 이루어집니다.
 - ex. '주민등록증|주민등록증|주민등록증'으로 작성시 '주민등록증'으로 인식되어도 서식 인식이 가능합니다.



- 후보정단어: 인식된 상세 정보 보정에 사용되는 단어 목록
 - 인식된 상세 정보에 오인식이 있을 경우 보정할 결과값 목록을 작성합니다.

- 보정 시 유사도는 자모를 분리하여 측정하며 유사도가 3 이하(대략, 한글 기준 1 글자, 영문 기준 3 글자)인 것은 제외되며 유사도가 높은 단어를 선택하여 보정합니다.
- 데이터형이 image, table, checkbox 인 경우 지정할 수 없습니다.
- 데이터형을 image, table, checkbox 로 변경할 경우 후보정단어 값이 빈 문자열로 자동 변경됩니다.
- ex. 후보정단어에 '주민등록증' 작성시 '주민등록증'으로 오인식이 발생했을 때 '주민등록증'으로 보정됩니다.



- 재인식: 서식 인식 결과를 활용한 재인식 모드 지정
 - 서식 인식 결과 박스를 활용해 글자를 재인식하여 개선된 결과를 얻기 위해 사용합니다.
 - basic, advanced 두 종류로 지정할 수 있습니다. 재인식 모드 설정 시 일반적으로 advanced 설정을 권장합니다.
 - basic: 글자 위치 재검출은 수행하지 않고 글자 재인식만 진행. 단어 인식에 적합.
 - advanced: 글자 위치 재검출 및 글자 재인식 진행. 문장 또는 여러 줄 인식에 적합.
 - 데이터형이 image, table, checkbox 인 경우 지정할 수 없습니다.
 - 데이터형을 image, table, checkbox 로 변경할 경우 재인식 값이 None 으로 자동 변경됩니다.


캔버스의 새 서식 아이템 추가 버튼(+)을 클릭하거나 서식 아이템 복사 버튼(+)을 클릭하여 서식 아이템을 추가하고, 위 작업을 반복합니다.

잘못 작업하였거나 불필요한 서식 아이템은 삭제 버튼(🗑️)을 클릭하여 삭제할 수 있습니다.



서식 데이터 테스트하기

서식 테스트 기능은 서식 작성시 사용한 샘플이미지를 이용하여 실제로 사이냅 OCR 엔진에 OCR을 요청하고 서식 매칭을 수행하여 서식 데이터 편집을 잘 진행했는지 확인할 수 있는 기능입니다.

테스트 전에 엑셀 파일로 서식을 저장하기를 권장합니다. 캔버스의 테스트 버튼()을 클릭하면 아래와 같이 다이얼로그가 나타납니다. 다이얼로그에 사이냅 OCR 엔진 주소와 API-KEY를 입력한 후 연결 버튼을 클릭합니다.

서식 인식 테스트

엔진 주소

127.0.0.1:62975

API-KEY

테스트를 진행할 엔진 주소와 API-KEY를 입력하세요.

연결

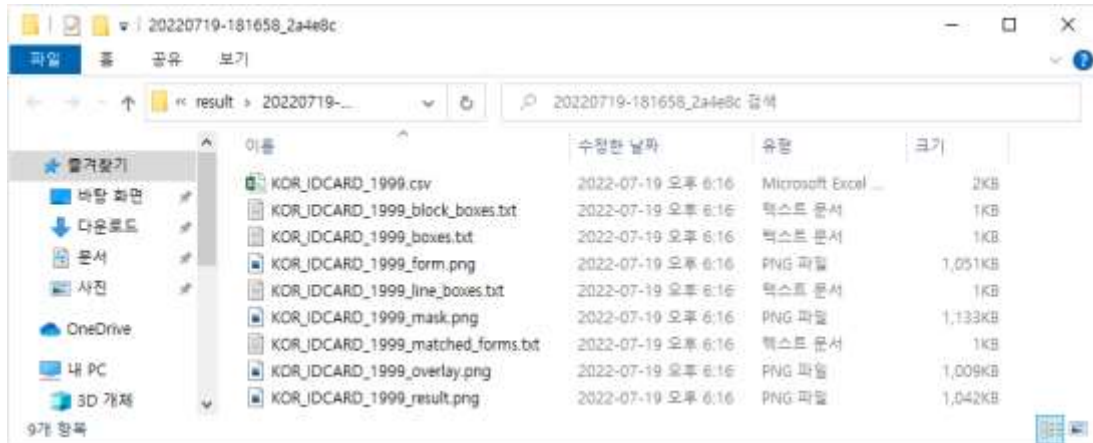
취소

다이얼로그 내 각 입력 항목은 아래와 같은 제약사항이 있습니다.

- 엔진 주소: URL 형식으로 작성되며 프로토콜 생략 시 'http://'로 지정됩니다.
- API-KEY: 영문, 숫자, 하이픈만 포함하며 8~64글자로 이루어져야 합니다.

사이냅 OCR 엔진에서 서식 인식 테스트가 진행된 후 테스트 결과 파일이 저장된 폴더가 나타납니다. 결과 파일들을 기반으로 서식 설정을 보정할 수 있습니다.

사이넵 OCR 엔진에서 서식 인식 테스트가 진행된 후 테스트 결과 파일이 저장된 폴더가 나타납니다. 결과 파일들을 기반으로 서식 설정을 보정할 수 있습니다.



- *.txt: OCR 결과 텍스트 파일
- *.csv: 서식 인식 결과 csv
- *_form.png: 서식 템플릿 박스가 그려진 이미지



- *_mask.png: 지정한 방식으로 마스킹된 이미지



- *_overlay.png: 서식 템플릿 박스와 OCR 결과 박스가 그려진 이미지



- *_result.png: 최종 서식 인식 결과가 그려진 이미지

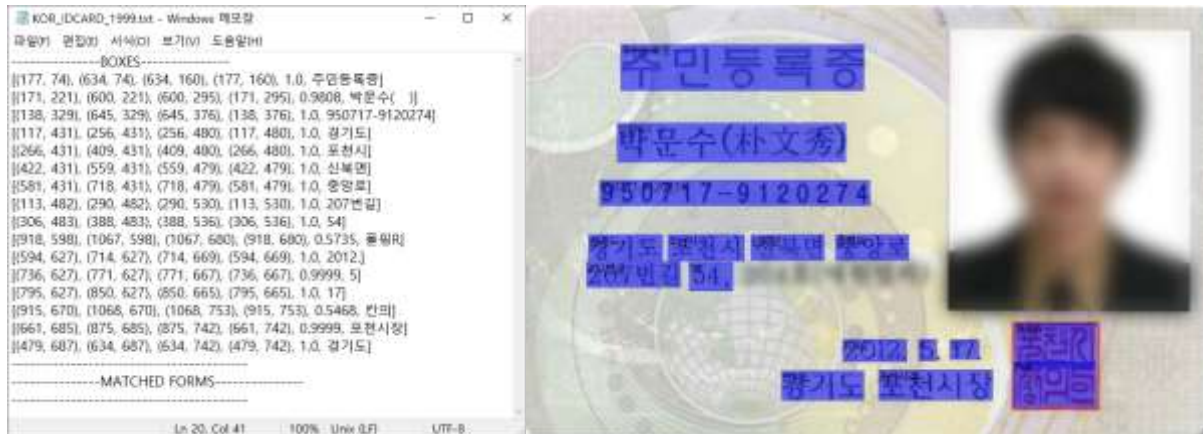


서식 테스트 실패 시 데이터 수정 작업하는 법

서식 인식 테스트 결과, 설정한 데이터로 서식이 인식되지 않았을 경우에 결과 파일들을 이용하여 원인을 파악할 수 있습니다. *.txt 파일에서 OCR 인식 결과 텍스트를 확인하고, *_overlay.png 파일에서 각 텍스트의 위치를 확인할 수 있습니다.

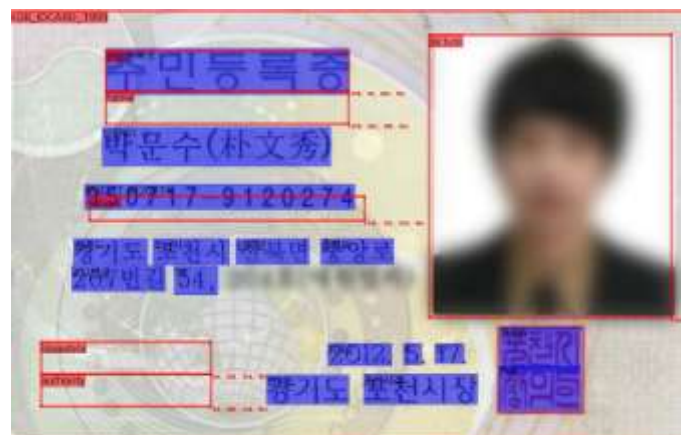
우선순위에 해당하는 텍스트가 인식되지 않았을 경우

설정한 서식 데이터의 우선순위 0에 해당하는 텍스트가 OCR 인식 결과에 없어 서식 인식을 진행하지 못했을 경우, *_overlay.png 파일에 OCR 결과 박스(파란색)만 그려지고, 서식 템플릿 박스(빨간색)은 그려지지 않습니다. 이 경우, 우선순위 0에 해당하는 텍스트가 인식되지 않은 원인을 파악하고 해당 내용을 수정합니다.



서식 데이터의 오류로 서식을 인식하지 못했을 경우

우선순위 0에 해당하는 텍스트를 찾아 서식 인식을 진행했으나 서식 데이터의 오류로 인해 서식을 인식하지 못했을 경우, *_overlay.png 파일에 OCR 결과 박스(파란색)와 서식 템플릿 박스(빨간색)가 그려집니다. 이 경우, 서식 템플릿 박스가 그려진 결과를 활용하여 데이터의 오류를 파악하고 해당 내용을 수정합니다.




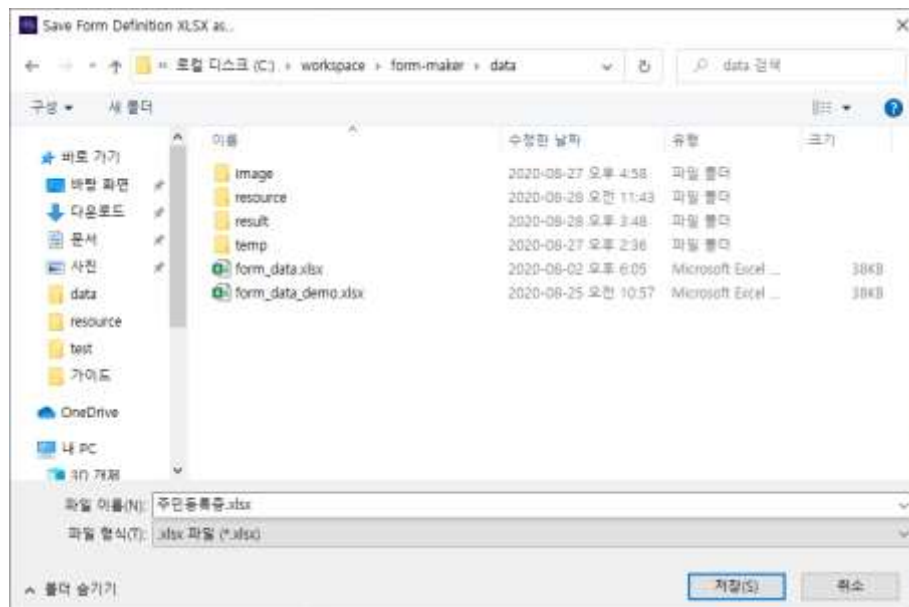
서식 저장하기

서식 인식이 되는 것까지 확인이 됐으면 서식 데이터를 엑셀 파일로 저장하거나 엔진에 업데이트합니다. 저장한 엑셀 파일과 엔진에 포함된 스크립트를 이용하여 수동으로 업데이트 및 테스트를

수행할 수 있습니다. 수동 업데이트 및 테스트는 매뉴얼의 '수동 업데이트 및 테스트' 목차를 참고 부탁드립니다.

엑셀 파일로 저장

컨트롤 패널의 다른 이름 엑셀로 저장 버튼()을 클릭하거나 파일 메뉴에서 다른 이름 엑셀로 저장(단축키 Ctrl+Shift+S)을 클릭합니다.



엑셀 파일은 아래와 같이 저장되며, 저장한 엑셀 파일을 이용하여 사이냅 OCR 엔진에 직접 서식을 업데이트할 수 있습니다. 저장한 엑셀 파일을 폼메이커로 다시 열어 편집하거나, 엑셀 파일을 직접 수정할 수 있습니다.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	id	name	type	left	top	right	bottom	img	img_center				
2	KOR_IDCARD	한국_주민등록증	ID_CARD		2	0	1282	814 ldcard.png	640.0, 407.0				
3													
4	fid	label	left	top	right	bottom	dtype	nullable	masking	priority	keyword	words	re_recog
5	title	제목	176	80	633	158	str	False			0	주민등록증	
6	name	이름	160	220	610	296	str	False			-1		
7	registration_number	주민등록번호	129	322	652	383	str	False	mosaic_right		-1		
8	address	주소	96	424	763	597	str	False			-1		
9	picture	사진	789	45	1241	575	image	False			-1		
10	issue_date	발급일자	537	620	885	677	str	False			-1		
11	authority	발급기관	403	684	886	745	str	False			-1		

1 행은 서식 정보에 대한 컬럼명이 작성됩니다. 고정 값으로 id, name, type, left, top, right, bottom, img, img_center 로 지정됩니다.

각 컬럼에 대한 데이터는 2 행에 기록되며 각각의 내용은 아래와 같습니다.

- id: 서식 아이디, 서식 구분에 사용되는 아이디 값
 - 영문, 숫자, 하이픈(-), 언더스코어(_)만 허용되며 최대 31 글자까지 지정 가능
- name: 서식 명칭

- type: 서식 유형
 - 서식 인식 결과는 type 을 기준으로 그룹핑되어 제공됩니다.
 - ID_CARD, DRIVE_LICENSE, PASSPORT, FAMILY_REGISTER 네 종류가 예시로 제공됩니다.
- left, top, right, bottom: 서식의 좌측, 상단, 우측, 하단 좌표
 - 폼메이커를 이용하여 편집하는 경우가 아니라면 엑셀 파일에서 직접 수정하지 않는 것을 권장합니다.
- img: 서식 작성 및 테스트에 사용한 샘플 이미지
 - 폼메이커에서 엑셀 파일을 열었을 때 해당 이미지를 보여줍니다.
- img_center: 샘플 이미지의 중점 좌표
 - 폼메이커에서 서식 데이터와 샘플 이미지 출력에 사용되는 값으로, 엑셀 파일에서 직접 수정하지 않는 것을 권장합니다.

3 행은 서식 정보와 서식 아이템 정보를 구분하는 빈 행입니다.

4 행은 서식 아이템 정보에 대한 컬럼명이 작성됩니다. 고정값으로 fid, label, left, top, right, bottom, dtype, nullable, masking, priority, keyword, words, re_recog 로 지정됩니다.

각 컬럼에 대한 데이터는 5 행부터 기록되며 각각의 내용은 아래와 같습니다.

- fid: 서식 아이템 아이디, 서식 아이템 구분에 사용되는 아이디 값
- label: 서식 아이템의 명칭
ex. 아이디(fid)는 'title', 이름(label)은 '제목'으로 지정



- left, top, right, bottom: 서식 아이템의 좌측, 상단, 우측, 하단 좌표
 - 폼메이커를 이용하여 편집하는 경우가 아니라면 엑셀 파일에서 직접 수정하지 않는 것을 권장합니다.
- dtype: 서식 아이템의 데이터 유형
 - num(숫자), char(한 글자), str(문자와 숫자가 섞여있는 텍스트), image(그림 영역), currency(통화 기호가 붙어있는 금액 영역), checkbox(체크박스), table(표)

- char 로 지정 시
 - 매칭된 영역에 여러 글자가 있어도 한 글자로 보정합니다.
- num 으로 지정 시
 - 알파벳 'O', 'o'로 인식된 결과값을 숫자 '0'으로 보정합니다.
 - 알파벳 'I', 'l'로 인식되거나 한글 'ㅣ', 특수문자 'l'로 인식된 결과값을 숫자 '1'로 보정합니다.
- str 로 지정 시
 - 별도 보정작업을 수행하지 않고 인식 한 결과를 사용합니다.
- Image 로 지정 시
 - 매칭된 영역에 인식된 글자가 있어도 인식 결과가 없도록 보정합니다.
- currency 로 지정 시
 - 알파벳 'S', 's'로 인식된 결과값을 '\$'로, 'W', 'w'로 인식된 결과값을 '₩'로 보정합니다.
- checkbox 로 지정 시
 - 해당 영역의 체크 여부를 반환합니다. (2021.10 추가)
- table 로 지정 시
 - 해당 영역의 표를 인식하여 결과를 반환합니다. (2022.01 추가)
- ex. 데이터형(dtype)을 아이템의 특성에 맞게 지정



- nullable: 빈 값 허용 여부
 - 해당 영역의 인식된 텍스트가 없어도 되는지 여부를 True, False 로 지정합니다.
 - 빈 값이 허용되는 영역의 경우 서식 종류 판단 시에 사용되지 않는 영역이 됩니다.
 - ex. 주민등록증의 주소 영역은 총 3 행으로 작성되지만, 실제 주소지에 따라 1 행~3 행으로 변동될 수 있습니다.
따라서 1 행은 반드시 상세 정보가 존재해야 하지만, 2 행과 3 행은 상세 정보에

빈 값이 허용됩니다.



- masking: 서식 아이템 영역에 마스킹할 범위와 방식 지정
 - {마스킹 종류}_{마스킹 영역}의 형식
 - 마스킹 종류는 mosaic 또는 RGB, RGBA 값으로 지정 가능합니다.
 - 마스킹 영역은 left, right, full 로 지정 가능합니다.
 - ex. mosaic_left, #RRGGBB_right, #RRGGBBAA_full

- ex. 첫 번째처럼 마스킹 지정 시 두 번째처럼 마스킹 결과를 얻을 수 있습니다.



- priority: 서식 종류 구분에 사용되는 상세 영역 여부 지정
 - 0 부터 순서대로 지정하며, 0 이상의 항목이 반드시 한 개 이상 존재해야 합니다.
 - 같은 우선순위의 값을 여러 개로 지정 가능하며 서식 종류 구분에 사용되지 않는 항목은 -1 로 지정합니다.
 - ex. 서식 종류 구분의 기준이 될 수 있는 '주민등록증'을 priority 0 으로 지정



- keyword: 서식 종류 구분에 사용되는 상세 영역에 포함되는 텍스트 지정
 - Python 정규식 문법으로 작성합니다.
 - priority 값이 0 이상인 경우 필수로 작성해야하는 속성입니다.
 - OCR 엔진이 인식한 텍스트의 최소 영역(raw box) 단위로 매칭이 이루어집니다.
 - ex. '주민등록증|주민등록증|주민등록증'으로 작성시 '주민등록증'으로 인식되어도 서식 인식이 가능합니다.



- words: 인식된 상세 정보 보정에 사용되는 단어 목록
 - 인식된 상세 정보에 오인식이 있을 경우 보정할 결과값 목록을 작성합니다.
 - 보정 시 유사도는 자모를 분리하여 측정하며 유사도가 3 이하(대략, 한글 기준 1 글자, 영문 기준 3 글자)인 것은 제외되며 유사도가 높은 단어를 선택하여 보정합니다.
 - ex. 후보정단어에 '주민등록증' 작성시 '주민등록증'으로 오인식이 발생했을 때 '주민등록증'으로 보정됩니다.




- re_recog: 서식 인식 결과를 활용한 재인식 모드 지정
 - 서식 인식 결과 박스를 활용해 글자를 재인식하여 개선된 결과를 얻기 위해 사용합니다.

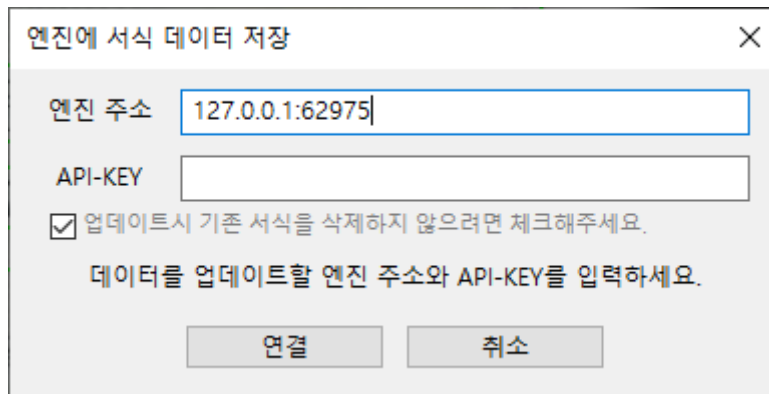
- basic, advanced 두 종류로 지정할 수 있습니다. 재인식 모드 설정 시 일반적으로 advanced 설정을 권장합니다.
 - basic: 글자 위치 재검출은 수행하지 않고 글자 재인식만 진행. 단어 인식에 적합.
 - advanced: 글자 위치 재검출 및 글자 재인식 진행. 문장 또는 여러 줄 인식에 적합.

엔진에 업데이트

작성한 서식을 엔진에 업데이트할 수 있으며, 서식 업데이트 시 사이냅 OCR 엔진에 즉시 반영되므로 사용 시 주의가 필요합니다.

컨트롤 패널의 엔진으로 저장 버튼()을 클릭하거나 파일 메뉴에서 엔진으로 저장(단축키 Ctrl+Shift+E) 또는 OCR엔진 메뉴에서 수정 사항 저장(단축키 Ctrl+Shift+T)을 클릭합니다.

아래와 같이 다이얼로그가 나타나면, 다이얼로그에 사이냅 OCR 엔진 주소와 API-KEY를 입력한 후 연결 버튼을 클릭합니다.



엔진에 서식 데이터 저장

엔진 주소

API-KEY

☒ 업데이트시 기존 서식을 삭제하지 않으려면 체크해주세요.

데이터를 업데이트할 엔진 주소와 API-KEY를 입력하세요.

다이얼로그 내 각 입력 항목은 아래와 같은 제약사항이 있습니다.

- 엔진 주소: URL 형식으로 작성되며 프로토콜 생략 시 'http://'로 지정됩니다.
- API-KEY: 영문, 숫자, 하이픈만 포함하며 8~64글자로 이루어져야 합니다.

업데이트에 성공했을 경우 다이얼로그가 닫히고, 실패했을 경우 다이얼로그에 에러 메시지가 출력됩니다.

각 에러 메시지는 다음과 같습니다.

에러 메시지	원인
엔진 상태 조회 중 에러가 발생했습니다. 자세한 내용은 로그를 확인하세요.	엔진이 구동 중이 아니거나 엔진에서 에러가 발생한 경우
엔진 상태를 조회하지 못 했습니다. API-KEY를 확인하세요.	입력한 API KEY가 유효하지 않은 경우
데이터를 저장하지 못 했습니다.	서식 데이터에 오류가 있는 경우
엔진 주소가 올바르지 않습니다. 확인 후 다시 시도하세요.	입력한 엔진 주소가 유효하지 않은 경우
API-KEY가 올바르지 않습니다. 확인 후 다시 시도하세요.	입력한 API-KEY 가 유효하지 않은 경우

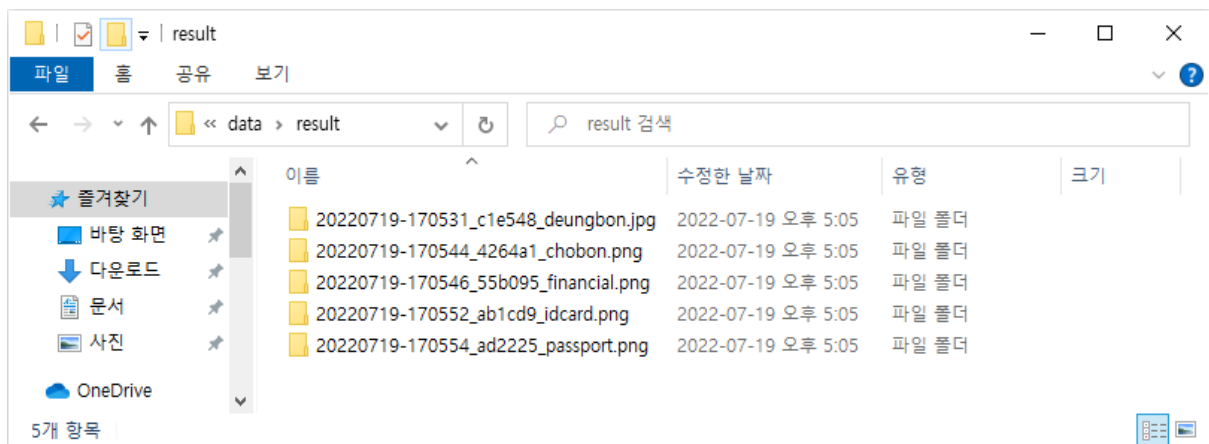
서식 배치 테스트하기

작업 중인 서식 또는 엔진에 업데이트된 서식을 이용해 여러 이미지들을 대상으로 배치 테스트를 수행할 수 있습니다.

패널 상단 메뉴 중 'OCR엔진' 탭에서 '현재 서식 배치 테스트' 또는 '모든 서식 배치 테스트'를 클릭하면 아래와 같이 다이얼로그가 나타납니다. 다이얼로그에 사이냅 OCR 엔진 주소와 API-KEY를 입력하고 테스트할 이미지들이 있는 경로를 선택한 후 '테스트' 버튼을 클릭합니다.

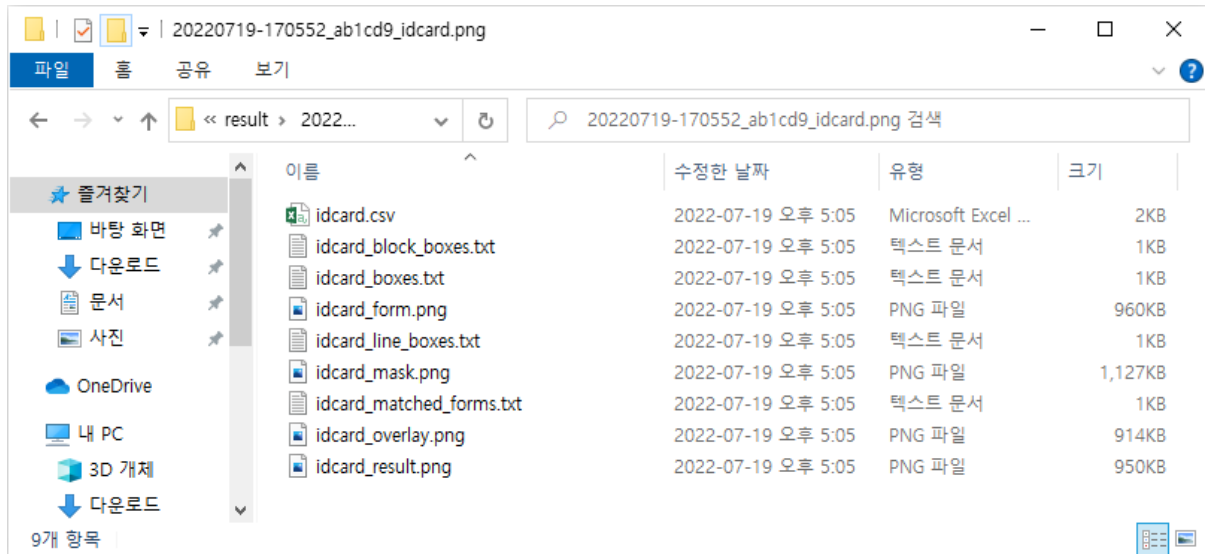
'테스트' 버튼을 클릭하면 아래와 같이 테스트 진행 상황을 보여주는 다이얼로그가 나타납니다.

테스트 완료 후 'Close' 버튼을 클릭하면 각 테스트 파일 별 배치 테스트 결과 파일들이 저장된 폴더가 나타납니다. 각 결과 파일들에 대한 설명은 '현재 서식 배치 테스트'와 '모든 서식 배치 테스트' 항목을 참고 부탁드립니다.



현재 서식 배치 테스트

현재 선택된 서식을 이용하여 테스트를 수행합니다. 테스트 결과 파일은 아래와 같습니다.



- *.csv: 서식 인식 결과 CSV
- *.txt: OCR 결과 텍스트 파일
- *_form.png: 서식 템플릿 박스가 그려진 이미지



- *_mask.png: 지정한 방식으로 마스킹된 이미지



- *_overlay.png: 서식 템플릿 박스(빨간색)와 OCR 결과 박스(파란색)가 그려진 이미지



- *_result.png: 최종 서식 인식 결과가 그려진 이미지



모든 서식 배치 테스트

엔진에 등록된 모든 서식을 이용하여 테스트를 수행합니다. 테스트 결과 파일은 아래와 같습니다.

관리 20220719-171415_62d214_idcard.png				
파일	폴더	공유	보기	사진 도구
← → ↕ ⬅ 20220719-... 🔍 20220719-171415_62d214_idcard.png 검색				
즐거찾기	이름	날짜	유형	크기
바탕 화면	b5b976d31571946675e0ba2a36fb5e9f_0.csv	2022-07-19 오후 5:14	Microsoft Excel ...	2KB
다운로드	idcard.png	2022-07-19 오후 5:14	PNG 파일	1,681KB
문서	KOR_IDCARD_1999_idcard.png.jpg	2022-07-19 오후 5:14	JPG 파일	195KB
사진				
3개 항목				

- *.csv: 서식 인식 결과 CSV

- *.png: 테스트에 사용된 이미지



- {인식된 서식 명}*.jpg: 인식된 서식 결과 이미지



사이냅 OCR 폼메이커 운영

로그 확인

사이냅 OCR 폼메이커의 로그는 "data/logs/FORM_MAKER.log" 파일에 작성됩니다. 서식 저장 및 업데이트, 장애 상황 등에 관련된 정보가 기록됩니다.

```
#> cat data/log/FORM_MAKERlog

2020-09-24    11:29:06,931    -    FORM_MAKER    -    INFO    :    MainThread(10616)
get_template_in_xlsx@modswform_xlsx.py:280 Verify succeeded. All form data from xlsx is valid.

2020-09-24    11:29:13,601    -    FORM_MAKER    -    INFO    :    MainThread(10616)
get_engine_status@modswform_engine.py:30 Get engine status succeeded.

2020-09-24    11:29:46,328    -    FORM_MAKER    -    INFO    :    MainThread(10616)
get_engine_status@modswform_engine.py:30 Get engine status succeeded.

2020-09-24    11:29:46,433    -    FORM_MAKER    -    INFO    :    MainThread(10616)
update_to_engine@modswform_engine.py:210 Save data to engine succeeded.

2020-09-24    11:30:18,562    -    FORM_MAKER    -    INFO    :    MainThread(10616)
generate_to_xlsx@modswform_xlsx.py:326 Save XLSX succeeded

.....
```

수동 업데이트 및 테스트

엑셀 파일과 사이냅 OCR 엔진에 포함된 스크립트를 이용하여 수동으로 업데이트 및 테스트를 할 수 있으며, 엔진 로컬 장비에서 수행할 수 있습니다.

윈도우 환경에서는 배치 파일(update_custom_template_win.bat), 리눅스 환경에서는 Shell 스크립트(update_custom_template.sh)를 이용하여 수행합니다.

스크립트를 이용하여 수동 업데이트 및 테스트, 서식 복구, 서식 목록 조회, 서식 정보 조회, 특정 서식 삭제를 수행할 수 있습니다.

서식 업데이트 및 삭제 시 사이냅 OCR 엔진에 즉시 반영되므로 사용 시 주의가 필요합니다.

서식 업데이트 스크립트 파라미터

- -h: 도움말
 - 실행 명령과 사용 예제를 출력합니다.
- -x: 서식 정보 엑셀 파일 경로
 - 문자열 타입 값을 지정할 수 있습니다.
 - 서식 업데이트 시 사용할 서식 정보 엑셀 파일 경로를 지정합니다.
 - -s 파라미터와 함께 사용시 엔진에 등록된 서식을 삭제하고 엑셀 파일의 서식 정보로 업데이트를 수행합니다.
 - 값을 지정하지 않으면 {사이냅 OCR 엔진 경로}/data/form_data.xlsx 파일을 사용합니다.
- -s: 업데이트 시 엔진에 등록된 서식 삭제
 - 엔진에 등록된 서식을 삭제하면서 서식 업데이트를 수행합니다.
 - -x 파라미터 또는 -r 파라미터와 함께 사용해야합니다.
- -p: 포트 번호 지정
 - 문자열 타입 값을 지정할 수 있습니다.
 - 서식 정보를 저장할 엔진의 포트번호 지정할 때 사용합니다.
 - 기본값은 62975입니다.
- -r: 기본 서식으로 복구
 - 엔진 설치 시 포함되어있는 초기 서식 정보로 복구합니다.
 - -s 파라미터와 함께 사용시 엔진에 등록된 서식을 삭제하고 복구합니다.
- -l: 서식 목록 출력
 - 엔진에 등록된 서식 목록을 출력합니다.
- -i: 서식 아이디 지정 및 서식 정보 조회
 - 문자열 타입 값을 지정할 수 있습니다.

- 단독 사용 시 서식 정보를 출력합니다.
- -d 파라미터와 함께 사용 시 삭제할 서식 아이디 지정에 사용합니다.
- -t 파라미터와 함께 사용 시 테스트할 서식 아이디 지정에 사용합니다.
- -d: 서식 삭제
 - 엔진에 등록된 서식 중 지정한 서식 아이디와 일치하는 서식을 삭제합니다.
 - -i 파라미터와 함께 사용해야 합니다.
- -t: 서식 테스트 파일 경로 지정
 - 문자열 타입 값을 지정할 수 있습니다.
 - 서식 테스트 파일 경로를 지정하면 서식 테스트가 수행됩니다.
 - 테스트 파일을 엔진에 등록된 서식 중 지정한 서식 아이디와 일치하는 서식 정보로 테스트합니다.
 - -i 파라미터와 함께 사용해야 합니다.
 - -o 파라미터, -a 파라미터, -q 파라미터를 선택적으로 지정할 수 있습니다.
- -o: 서식 테스트 결과 경로 지정
 - 문자열 타입 값을 지정할 수 있습니다.
 - 서식 테스트 결과를 저장할 경로를 지정합니다.
 - -i 파라미터, -t 파라미터와 함께 사용해야 합니다.
- -a: 페이지 번호 지정
 - 정수 타입 값을 지정할 수 있습니다.
 - 멀티 페이지 이미지 요청 시 페이지 번호 지정에 사용합니다.
 - -i 파라미터, -t 파라미터와 함께 사용해야 합니다.
 - 기본값은 0입니다.
- -q: 회전 보정 지정
 - none, image, words 문자열 값을 입력할 수 있습니다.
 - 서식 테스트 시 회전 보정 기능 사용 유무 지정, -t 옵션과 함께 사용
 - -i 파라미터, -t 파라미터와 함께 사용해야 합니다.
 - 기본값은 image 입니다.

서식 업데이트

서식 업데이트 시 기본적으로 "{SYNAP_OCR}/data/form_data.xlsx" 파일을 사용하며, "-x" 옵션을 통해 엑셀 파일을 지정할 수 있습니다.

```
// 리눅스
#> ./update_custom_template.sh
#> ./update_custom_template.sh -x {엑셀_파일_경로}
```

```
// 윈도우
#> update_custom_template_win.bat
#> update_custom_template_win.bat -x {엑셀_파일_경로}

// 사용 예제
#> ./update_custom_template.sh
Verify succeeded. All form data from xlsx is valid.
KOR_IDCARD_1999 conversion succeeded.
KOR_DRIVE_LICENSE_2014 conversion succeeded.
KOR_PASSPORT conversion succeeded.
KOR_FAMILY_REGISTER_COPY_ONLINE_START conversion succeeded.
KOR_FAMILY_REGISTER_COPY_ONLINE_BACK conversion succeeded.
KOR_FAMILY_REGISTER_COPY_EXTRACT_START conversion succeeded.
KOR_FAMILY_REGISTER_COPY_EXTRACT_BACK conversion succeeded.
Update form data is successful.

Test recognize form...
ID_CARD : Success.
PASSPORT : Success.
FAMILY_REGISTER : Success.
All test images were successfully form recognized.
```

서식 복구

"-r" 옵션을 통해 최초 예제 항목으로 서식을 복구합니다.

```
// 리눅스
#> ./update_custom_template.sh -r
// 윈도우
#> update_custom_template_win.bat -r
```

서식 목록 조회

엔진에 등록되어있는 서식 목록을 조회할 수 있습니다.

해당 작업 수행 결과를 서식 정보 조회, 서식 삭제, 서식 테스트에서 활용할 수 있습니다.

```
// 리눅스
#> ./update_custom_template.sh -l
// 윈도우
#> update_custom_template_win.bat -l
```

```
// 사용 예제
#> ./update_custom_template.sh -l
KOR_BUSINESS_LICENSE: 한국_사업자등록증
KOR_BUSINESS_LICENSE_CORP: 한국_사업자등록증_법인사업자
KOR_DRIVE_LICENSE_2014: 한국_운전면허증_2014년
KOR_DRIVE_LICENSE_SAMPLE: 한국_운전면허증_샘플
KOR_FAMILY_REGISTER_COPY_ONLI_1: 한국_등본_2019년_온라인발급
KOR_FAMILY_REGISTER_EXTRACT_O_1: 한국_초본_2019년_온라인발급
KOR_FAMILY_REGISTER_EXTRACT_O_2: 한국_초본_2019년_온라인발급
KOR_FAM_REG_COPY_2021_ONLINE_1: 한국_등본_2021년_온라인발급
KOR_FAM_REG_COPY_2021_ONLINE_2: 한국_등본_2021년_온라인발급
KOR_IDCARD_1999: 한국_주민등록증_1999년
KOR_PASSPORT: 한국_여권_2008년
```

서식 정보 조회

서식 아이디를 이용하여 엔진에 등록된 해당 서식의 정보를 조회할 수 있습니다.

```
// 리눅스
#> ./update_custom_template.sh -i {서식아이디}

// 윈도우
#> update_custom_template_win.bat -i {서식아이디}

// 사용 예제
#> ./update_custom_template.sh -i KOR_IDCARD_1999
id: KOR_IDCARD_1999
name: 한국_주민등록증_1999년
std_height: 53.98
std_width: 85.6
type: ID_CARD
```

서식 삭제

서식 아이디를 이용하여 해당 서식을 엔진에서 삭제할 수 있습니다.

```
// 리눅스
#> ./update_custom_template.sh -d {서식아이디}
```

```
// 윈도우
#> update_custom_template_win.bat -d {서식아이디}

// 사용 예제
#> ./update_custom_template.sh -d KOR_IDCARD_1999
Delete template is successful.

----- LIST OF REMAINED TEMPLATE -----
KOR_BUSINESS_LICENSE: 한국_사업자등록증
KOR_BUSINESS_LICENSE_CORP: 한국_사업자등록증_법인사업자
KOR_DRIVE_LICENSE_2014: 한국_운전면허증_2014년
KOR_DRIVE_LICENSE_SAMPLE: 한국_운전면허증_샘플
KOR_FAMILY_REGISTER_COPY_ONLI_1: 한국_등본_2019년_온라인발급
KOR_FAMILY_REGISTER_EXTRACT_O_1: 한국_초본_2019년_온라인발급
KOR_FAMILY_REGISTER_EXTRACT_O_2: 한국_초본_2019년_온라인발급
KOR_FAM_REG_COPY_2021_ONLINE_1: 한국_등본_2021년_온라인발급
KOR_FAM_REG_COPY_2021_ONLINE_2: 한국_등본_2021년_온라인발급
KOR_PASSPORT: 한국_여권_2008년
```

서식 테스트

샘플 이미지와 스크립트를 이용하여 현재 설정되어 있는 서식을 테스트할 수 있습니다.

해당 작업 수행 시 매칭 여부, 인식된 영역 위치 등 서식 설정에 도움이 되는 정보를 확인할 수 있습니다.

- -o 옵션을 함께 이용하여 테스트 결과를 저장할 경로를 지정할 수 있습니다.
- -a 옵션을 함께 이용하여 멀티 페이지 tiff 또는 pdf 일 경우 테스트할 페이지 인덱스를 지정할 수 있습니다.
- -q 옵션을 함께 이용하여 회전 보정 옵션(none/image/words)을 지정 할 수 있습니다.

```
// 리눅스
#> ./update_custom_template.sh -t {이미지경로} -i {테스트할 서식 아이디} -o {테스트 결과를 저장할 경로} -a {페이지인덱스} -q {ON/OFF}

// 윈도우
#> update_custom_template.bat -t {이미지경로} -i {테스트할 서식 아이디} -o {테스트 결과를 저장할 경로} -a {페이지인덱스} -q {ON/OFF}
```

```
// 사용 예제
#> ./update_custom_template.sh -t /SynapOCR/data/sample/idcard.png -i KOR_IDCARD_1999
boxes
[177, 74, 457, 86, 1.0, '주민등록증']
[171, 221, 429, 74, 0.9996, '박문수( )']
[138, 329, 507, 47, 1.0, '950717-9120274']
[265, 426, 145, 56, 1.0, '포천시']
[117, 431, 139, 49, 1.0, '경기도']
[422, 431, 137, 48, 1.0, '신북면']
[581, 431, 137, 48, 1.0, '중앙로']
[113, 482, 177, 48, 0.9999, '207번길']
[306, 483, 80, 51, 1.0, '54']
[919, 599, 147, 77, 0.3873, '풀재(']
[594, 627, 120, 42, 1.0, '2012.']
[736, 627, 35, 40, 1.0, '5']
[795, 627, 55, 38, 1.0, '17']
[915, 670, 153, 83, 0.6023, '행의']
[661, 685, 214, 57, 1.0, '포천시장']
[479, 687, 155, 55, 1.0, '경기도']

-----
matched_forms
id: KOR_IDCARD_1999
name: 한국_주민등록증_1999년
type: ID_CARD
boundary: [-2, 3, 1285, 810]
confidence: 0.975
['title', '제목', '주민등록증', [177, 74, 457, 86], "", 1.0, [177, 74, 457, 86]]
['name', '이름', '박문수( )', [171, 221, 434, 74], "", 0.9996, [171, 221, 429, 74]]
['idnum', '주민등록번호', '950717-9120274', [138, 329, 507, 47], 'mosaic_right', 1.0, [138, 329, 507, 47]]
['addr', '주소', '경기도 포천시 신북면 중앙로 207번길 54', [113, 426, 613, 108], 'mosaic_right', 1.0, [113, 426, 613, 108]]
['issuedate', '발급일자', '2012. 5 17', [594, 627, 274, 42], 'mosaic_right', 1.0, [594, 627, 256, 42]]
['authority', '발급기관', '경기도 포천시장', [479, 685, 404, 57], "", 1.0, [479, 685, 396, 57]]
['picture', '사진', "", [786, 48, 450, 525], 'mosaic_full', 1, [786, 48, 450, 525]]

-----
csv_file_name: /SynapOCR/data/sample/idcard.csv
```


form_boxes_image: /SynapOCR/data/sample/idcard_form.png
masked_image: /SynapOCR/data/sample/idcard_mask.png
matched_results_image: /SynapOCR/data/sample/idcard_result.png
overlay_image: /SynapOCR/data/sample/idcard_overlay.png
Test images were successfully form recognized.

각 결과 파일은 아래와 같습니다.

- {입력_이미지_파일이름}.csv: 서식 인식 결과 CSV 파일
- {입력_이미지_파일이름}_form.png: 추정된 템플릿 박스를 그린 이미지



- {입력_이미지_파일이름}_mask.png: 설정한 옵션에 따라 마스킹한 결과 이미지









































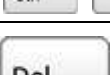


- {입력_이미지_파일이름}_result.png: 서식 인식 결과 박스를 그린 이미지



- {입력_이미지_파일이름}_overlay.png: OCR 인식 결과 박스(파란색)와 추정한 템플릿 박스(빨간색)를 그린 이미지



단축키 가이드

구분	단축키	설명
파일 메뉴	 	새 서식 만들기
	 	서식 엑셀 파일 열기
	 	저장하기
	  	엑셀 파일로 저장하기
	  	엔진으로 저장하기
	 	종료하기
편집 메뉴	 	되돌리기 (Undo)
	  	다시실행 (Redo)
	  	다시 불러오기 (Reload)
엔진 메뉴	 	엔진에서 서식 데이터 불러오기
	  	엔진으로 수정사항 저장하기
서식 아이템 선택 시	   	아이템 위치 조정
	    	아이템 미세 위치 조정
	 	아이템 복사
		아이템 삭제
서식 화면 조정	 + 캔버스 영역 드래그	패닝
	 + 마우스 휠	확대 / 축소

제품 공급자 정보

(주)사이냅소프트

회사 주소	(07805) 서울특별시 강서구 마곡중앙4로 62		
웹사이트	https://support.synapsoft.co.kr , https://www.synap.co.kr/ocr		
대표 전화	(02) 2039-3780	팩스 번호	02-6918-0950
제품 문의	(02) 2039-3781 sales@synapsoft.co.kr	기술 지원	(02) 2039-3782 support@synapsoft.co.kr