

AI로 복잡한 문서 이해하기: OCR, Parsing 등 문서 이해 관련 기술 소개

김동현 / NAVER Clova Visual AI OCR



발표자 소개



김동현

컴퓨터 공학 학사 | 2013.03 - 2019.08 | 인하대학교

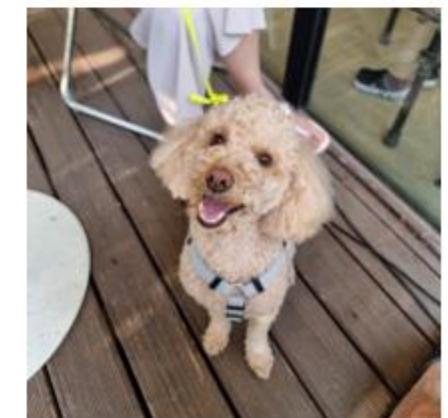
CLOVA OCR 팀 인턴 | 2018.07 ~ 2018.12

CLOVA OCR 팀 | RESEARCH ENGINEER | 2019.09 ~ 현재

KAIST AI 대학원 | 석사과정(PART TIME) | 2022.09 시작 예정

관심 분야

- BACKBONE
- LARGE MODELS
- MULTI MODAL



Members



이바도 (Leader)



김윤식



김성현



김동현



김대희



서석민



임유민



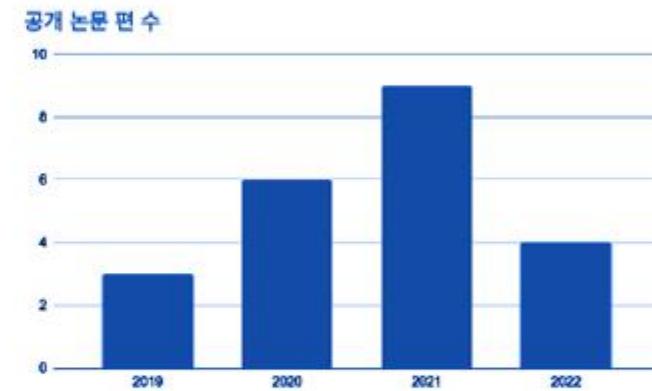
길태호

Papers

논문	학회	연도
CRAFT	CVPR	2019
WIW	ICCV	2019
TedEval	ICDARW	2019
CLEval	CVPRW	2020
SATRN	CVPRW	2020
DM-Font	ECCV	2020
DM-Font	CVPRW	2020
CRAFTS	ECCV	2020
GroupTransformer	COLING	2020

논문	학회	연도
LF-Font	AAAI	2021
SAD	AAAI	2021
DocMix	AAAIW	2021
LF-Font	CVPRW	2021
MX-Font	ICCV	2021
SynthTIGER	ICDAR	2021
HoughCL	ICMLW	2021
SelfReg	ICCV	2021
SWAD	NeurIPS	2021

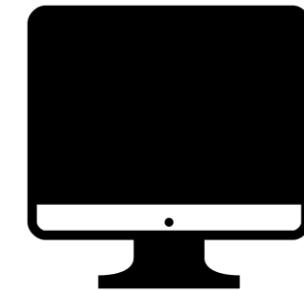
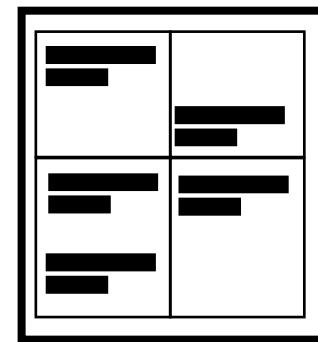
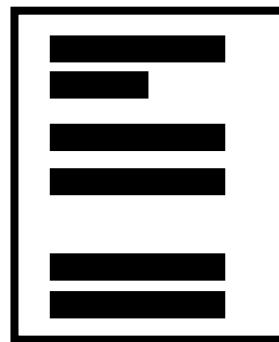
BROS	AAAI	2022
DEER	arXiv	2022
RewriteNet	CVPRW	2022
MATRN	ECCV Submitted	2022



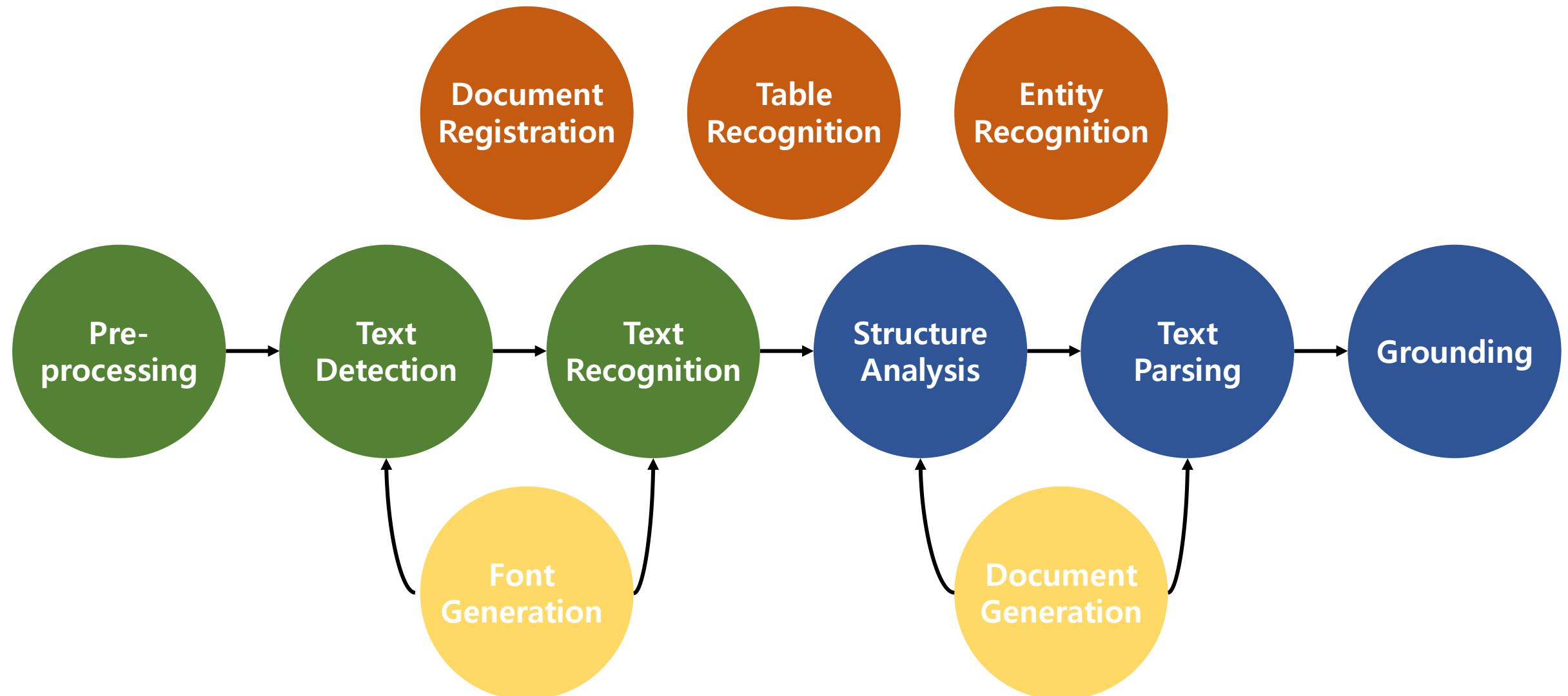
들어가기

오늘의 발표 : 문서

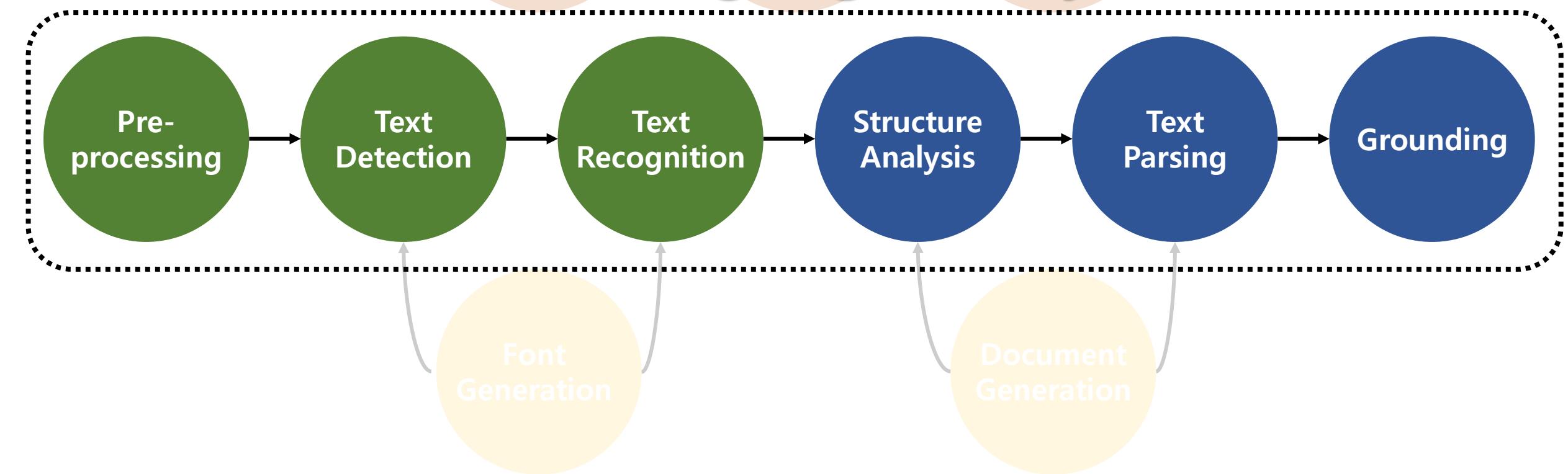
Document Understanding

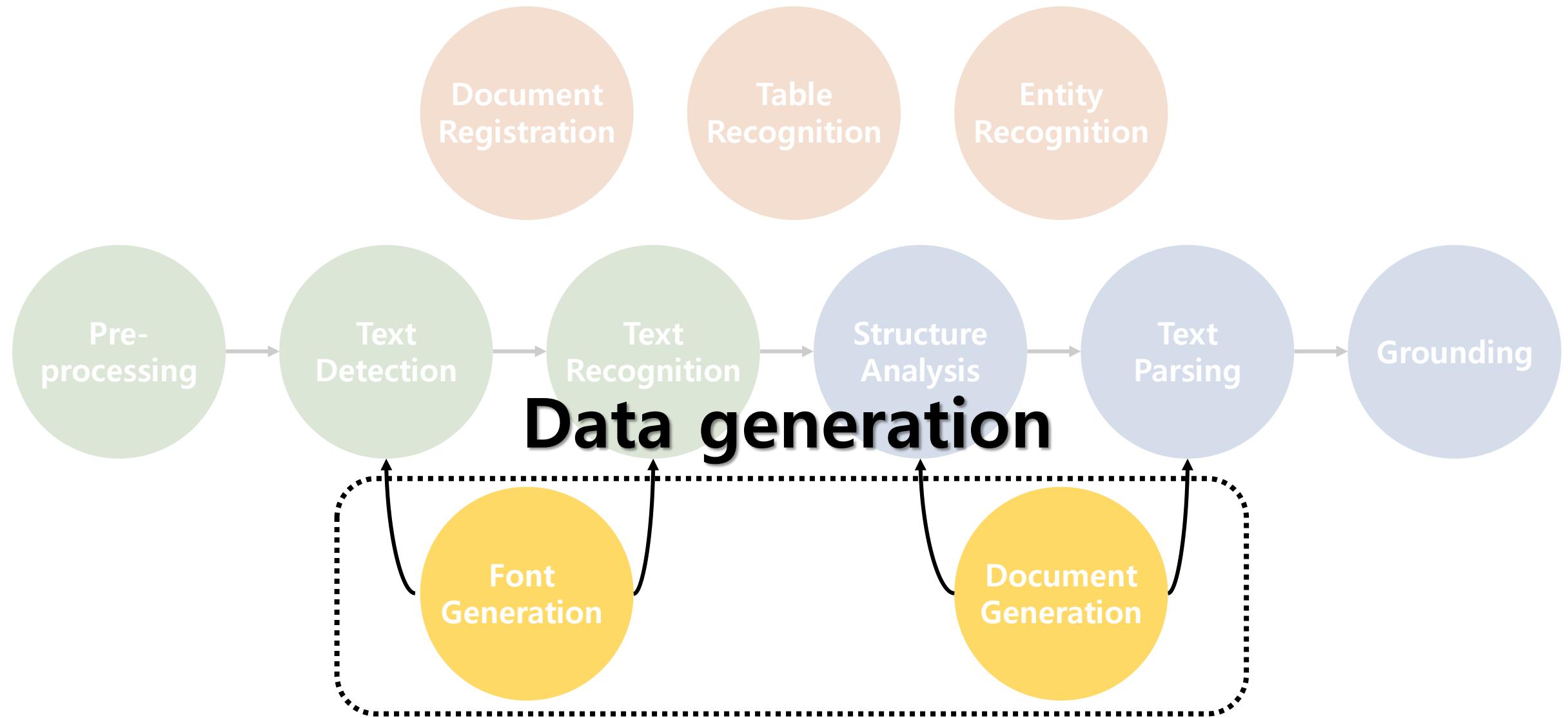


Document Understanding

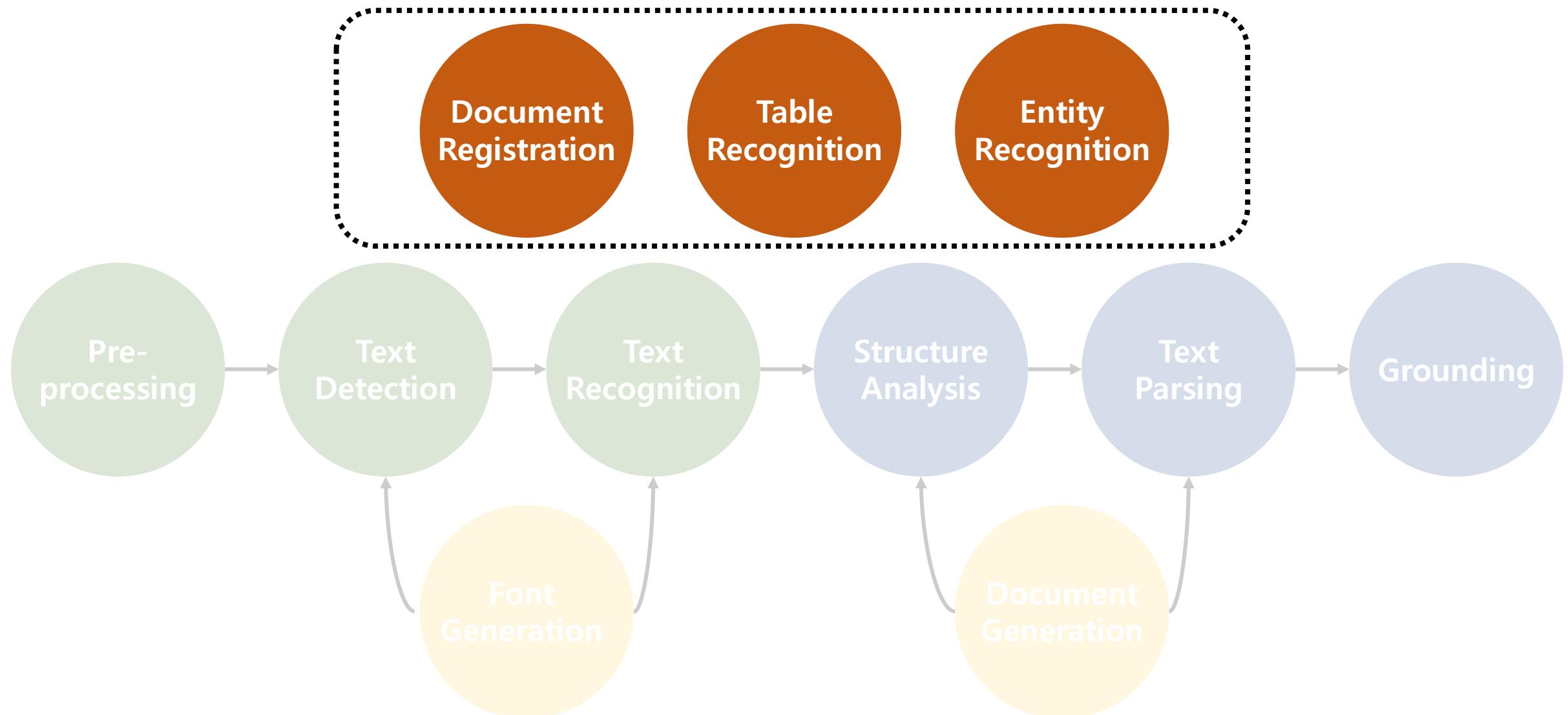


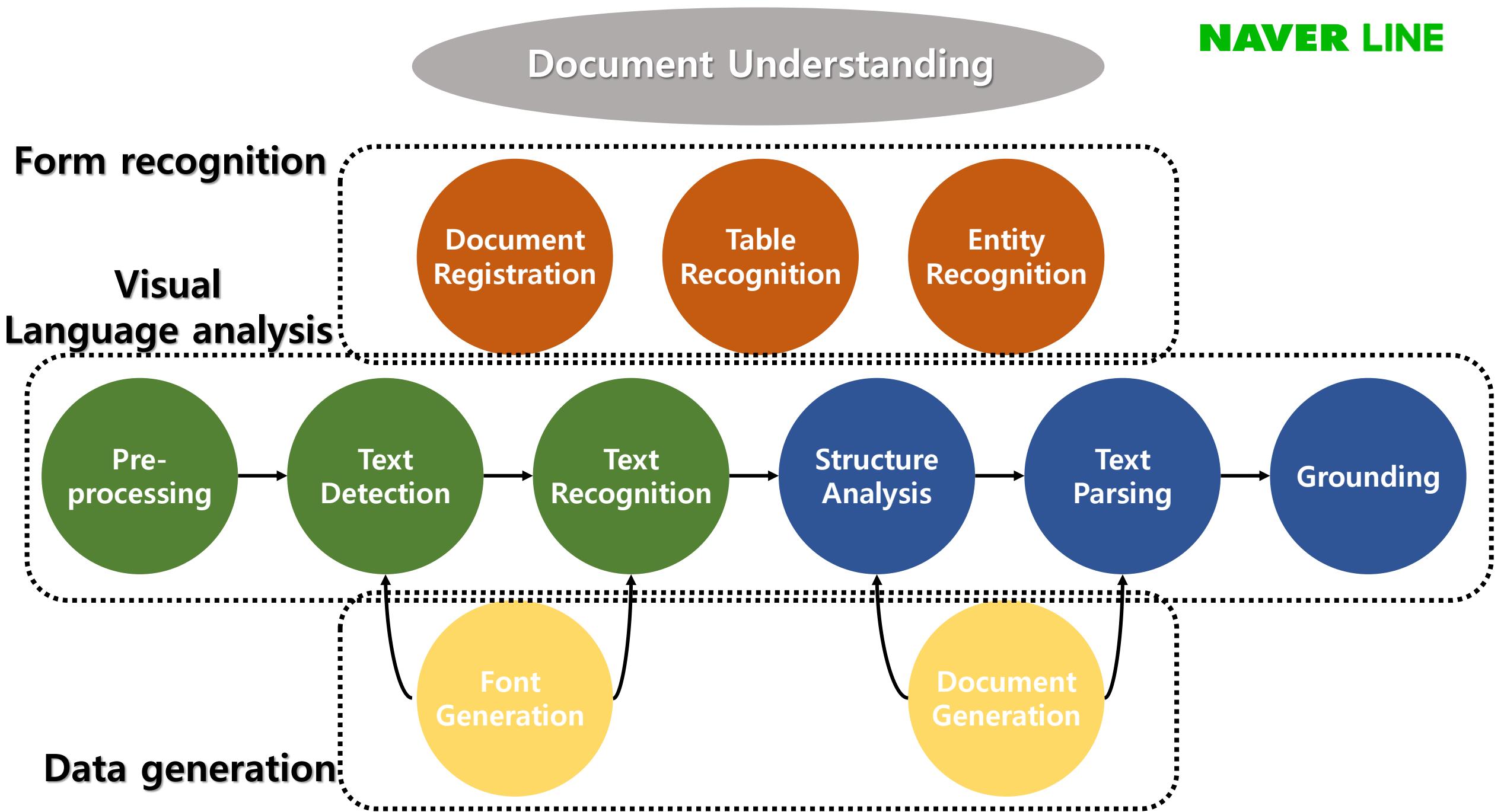
Visual language analysis

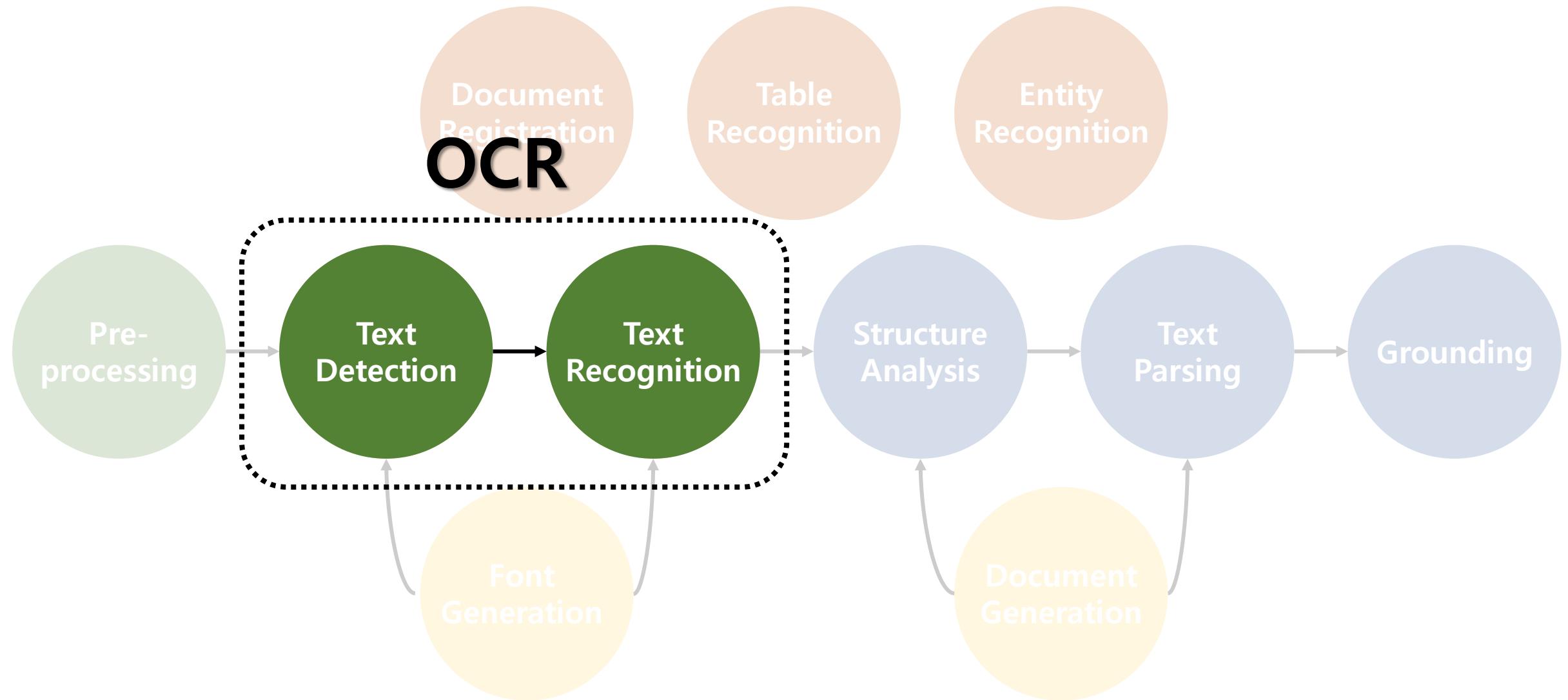


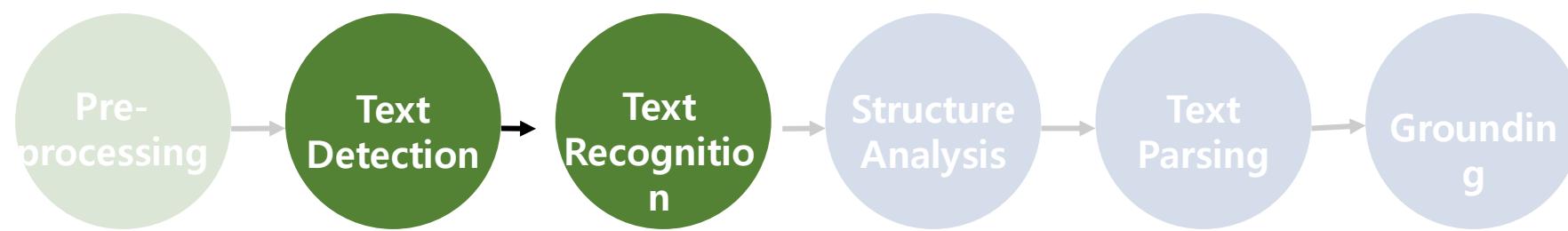


Form recognition



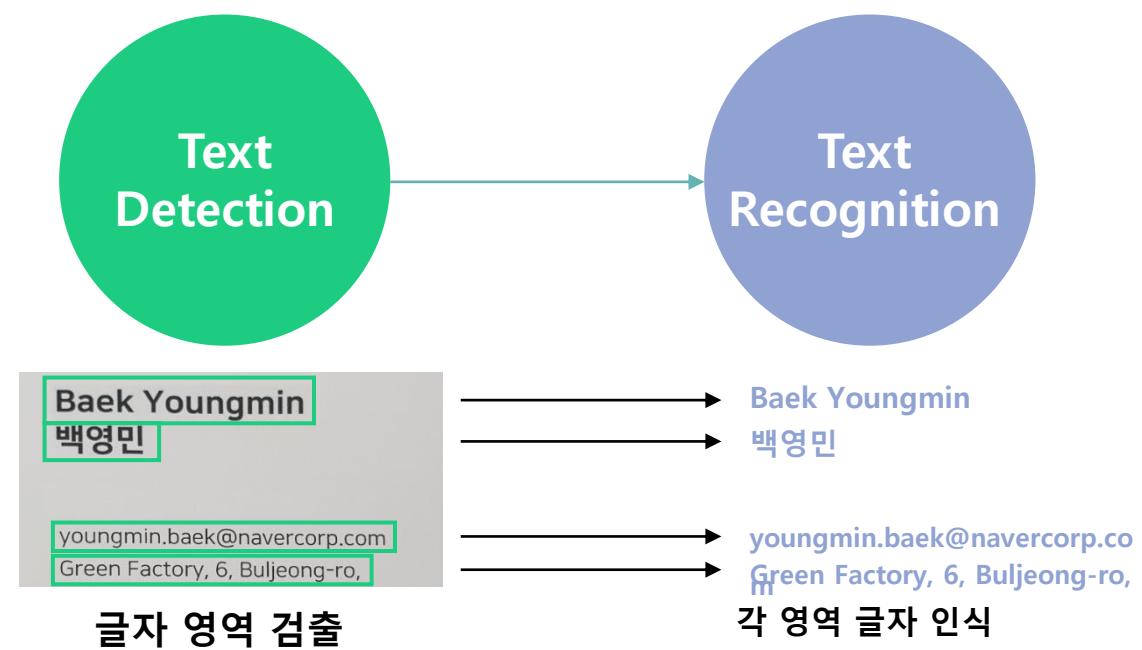


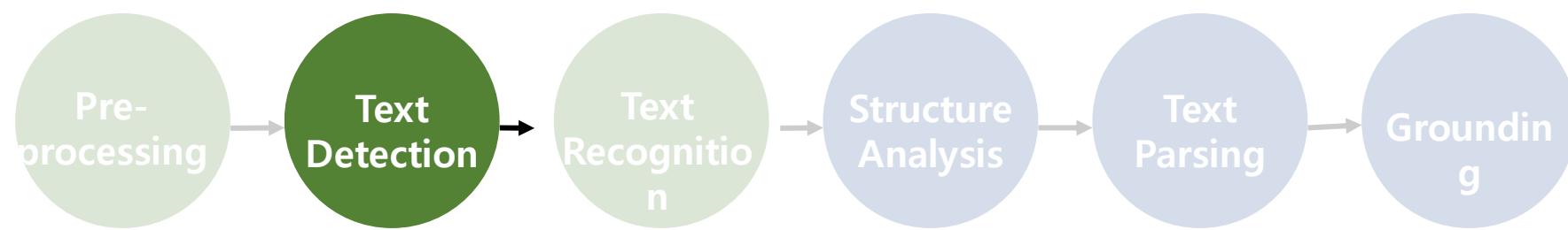




OCR (Optical Character Recognition)

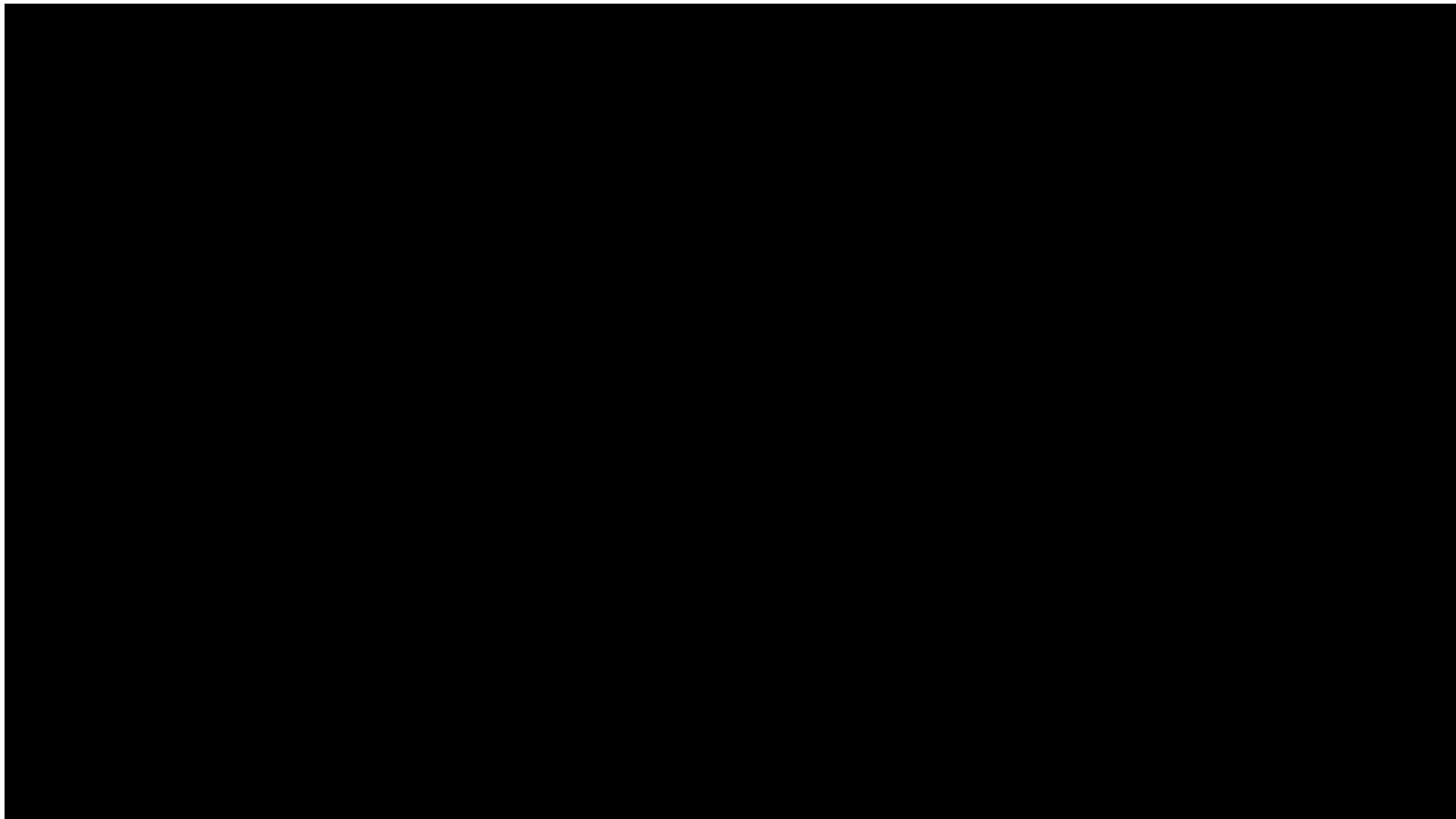
글자 영역 검출 + 각 영역 글자 인식

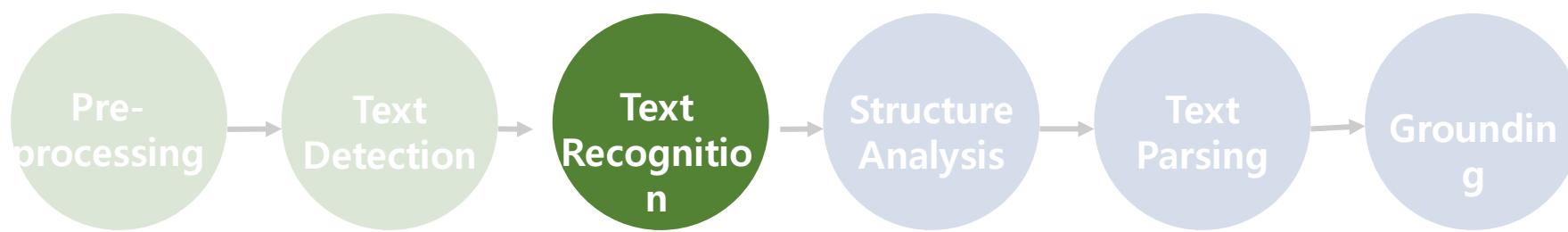




Text Detection

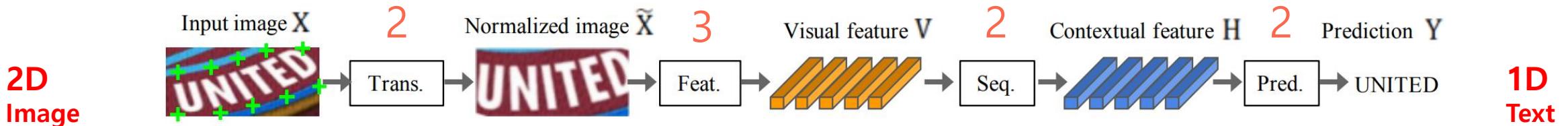
CRAFT (Character Region Awareness for Text Detection, CVPR 19)





Text Recognition

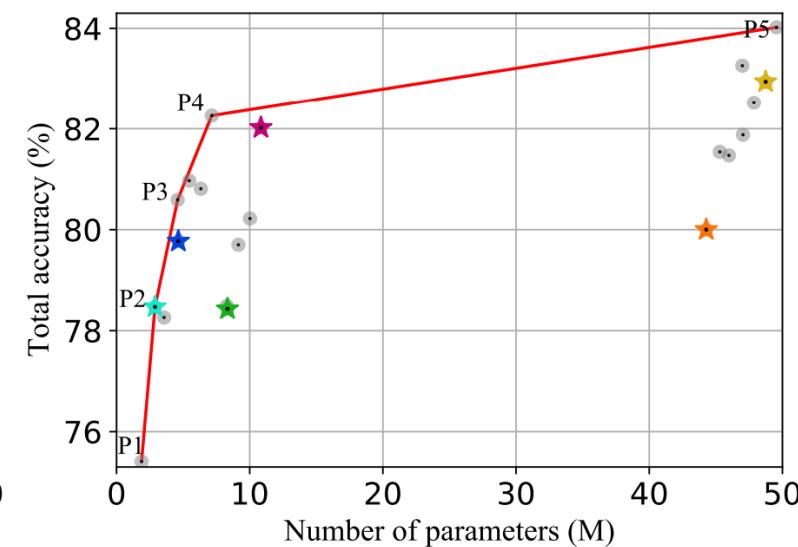
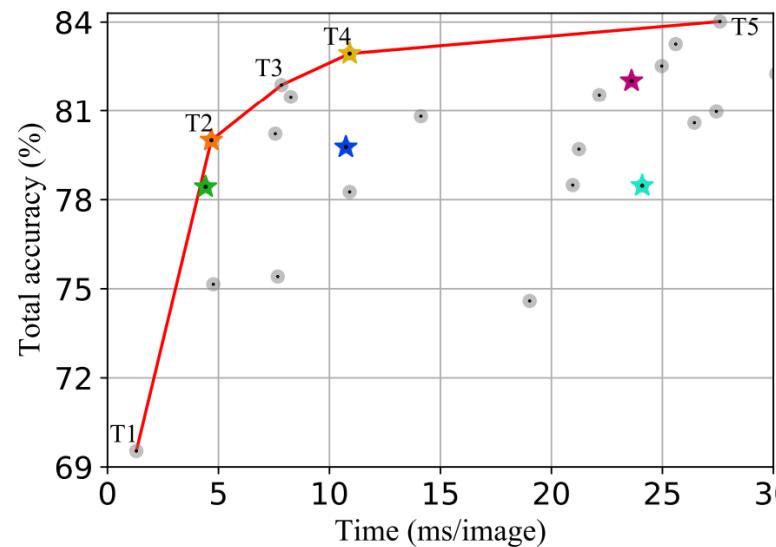
What is wrong with scene text recognition model comparisons? (ICCV19)

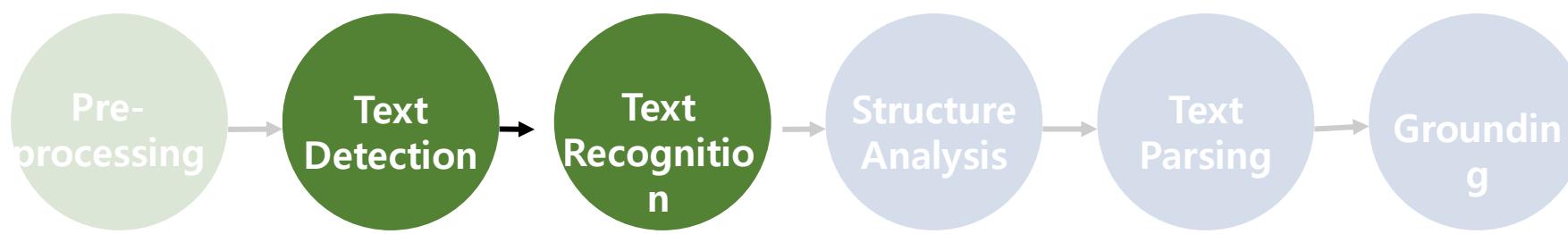


Previously proposed combinations

- | | | |
|-----------------------------|-------------------------------|----------------------------------|
| ★ CRNN: None-VGG-BiLSTM-CTC | ★ R2AM: None-RCNN-None-Attn | ★ Rosetta: None-ResNet-None-CTC |
| ★ RARE: TPS-VGG-BiLSTM-Attn | ★ GRCNN: None-RCNN-BiLSTM-CTC | ★ STAR-Net:TPS-ResNet-BiLSTM-CTC |

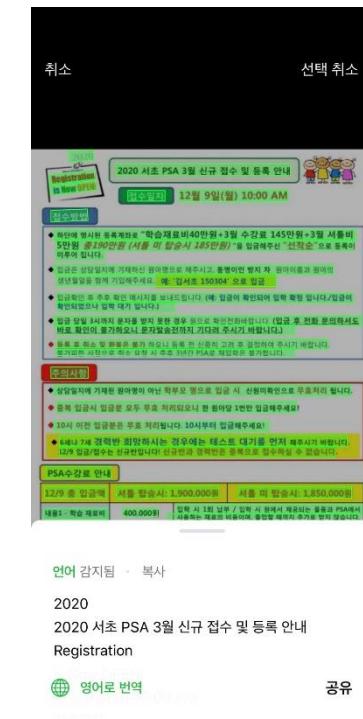
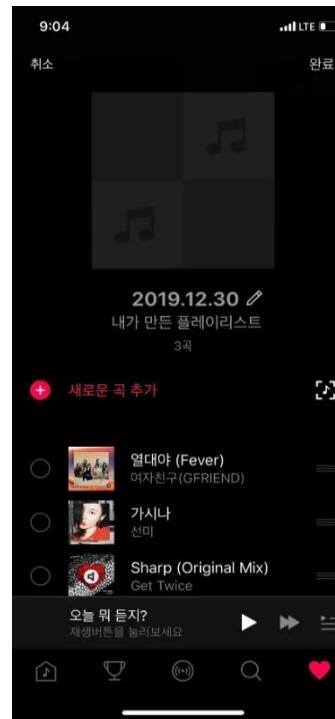
24
combinations





OCR Applications

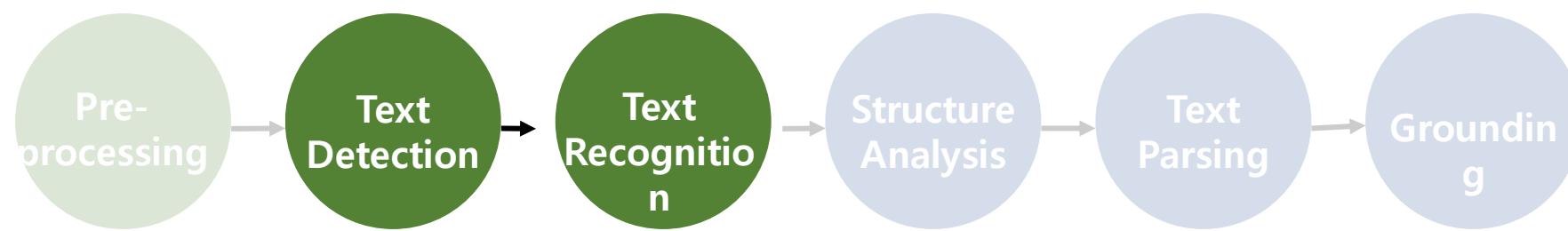
신용카드, 지로영수증, 음악 리스트 등을 인식하여 서비스 사용성 개선



OCR 스캔하기

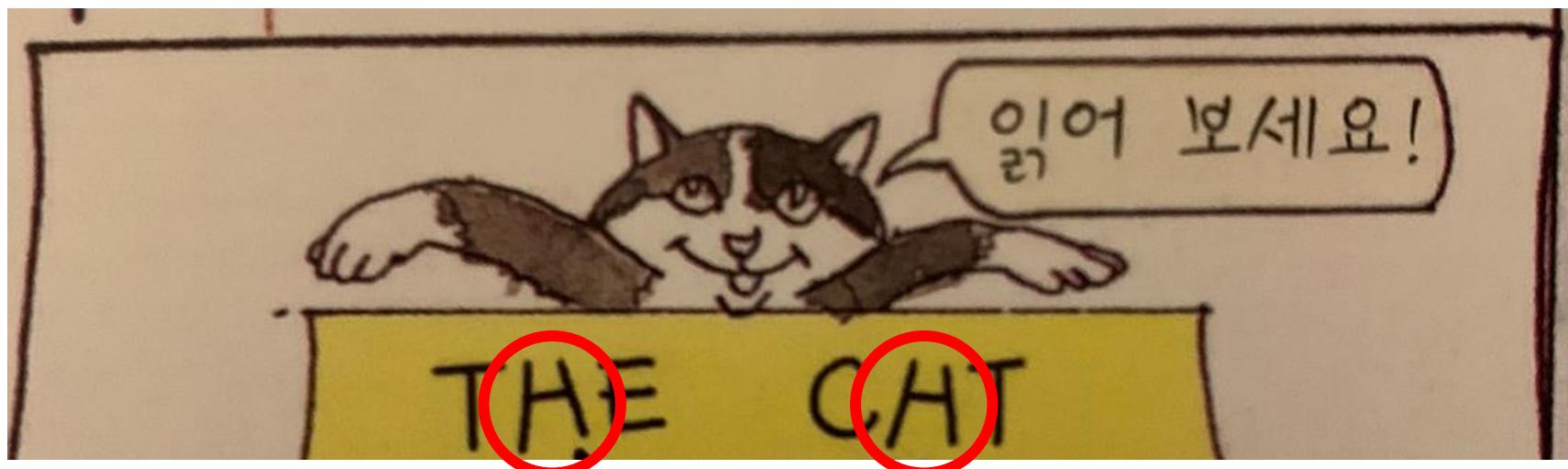
알아두세요

- OCR스캔은 지방세 3종(재산세, 자동차세, 주민세)과 지로용지에 사용이 가능합니다.
- 고객님께서 활용된 고지서는 네이버로 전달되어 납부에 필요한 정보를 추출합니다.
- 추출된 결과를 반드시 확인하시고, 인식이 정확하지 않은

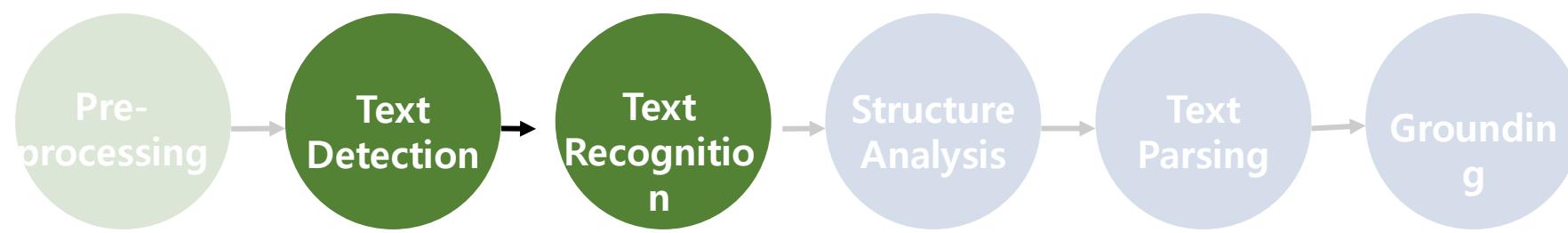


OCR의 한계점

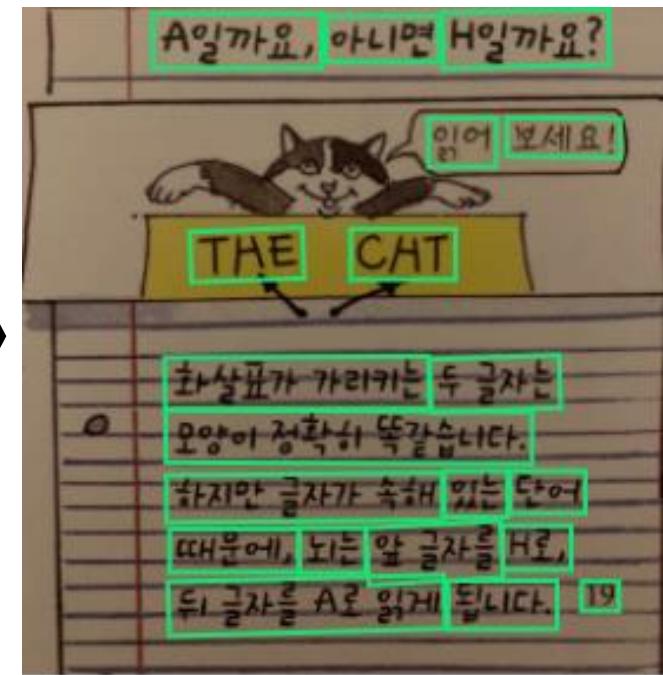
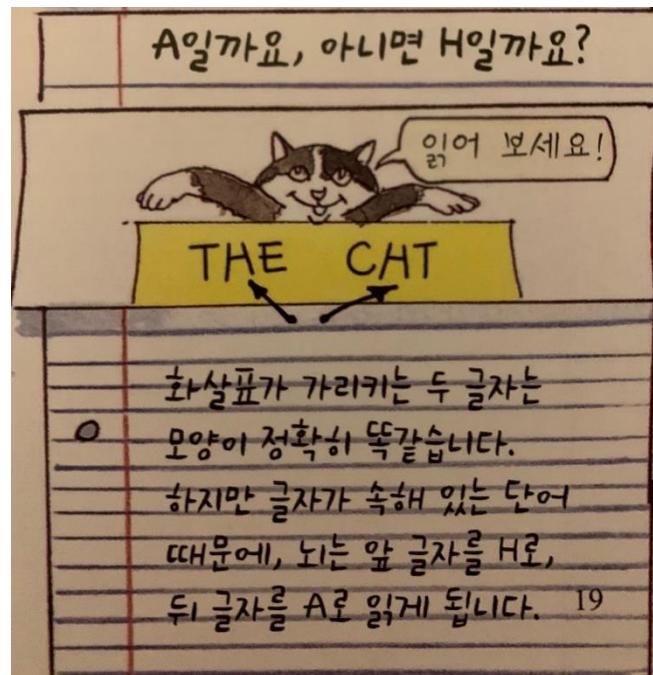
여기서 잠깐 퀴즈!!!



→ THE CAT



OCR의 한계점



A일까요, 아니면 H일까요?

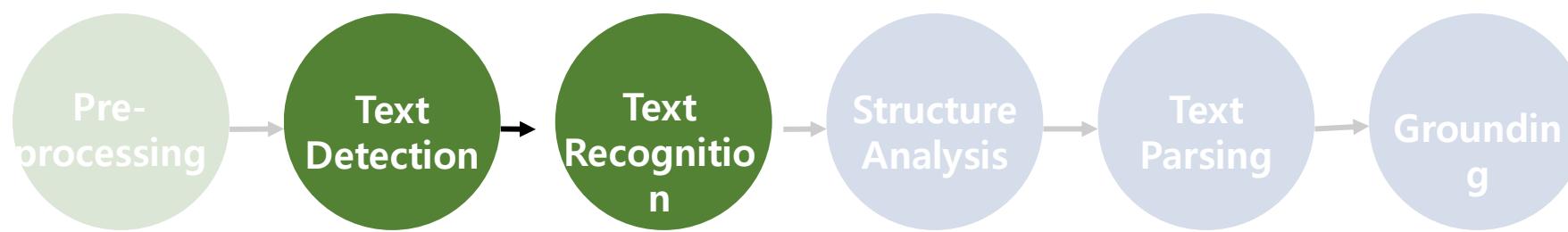
읽어 보세요!

THE CHT

화살표가 가리키는 두 글자는
모양이 정확히 똑같습니다.

하지만 글자가 속해 있는 단어
때문에, 뒤는 앞 글자를 H로,
뒤 글자를 A로 읽게 됩니다. 19

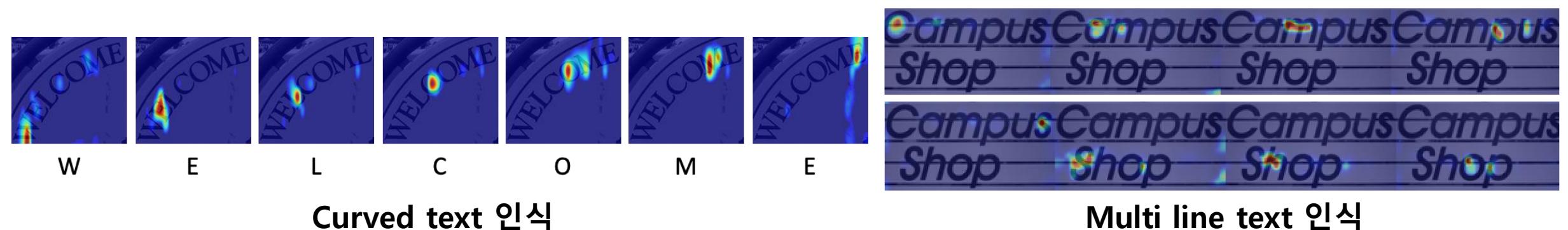
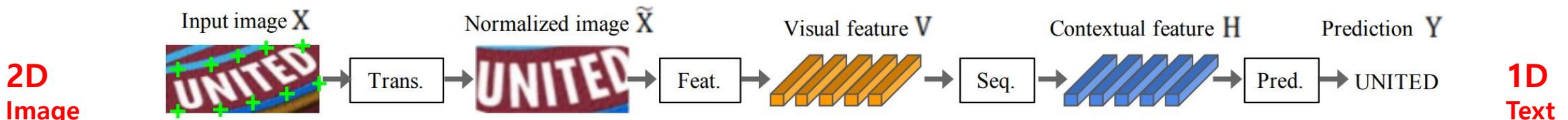
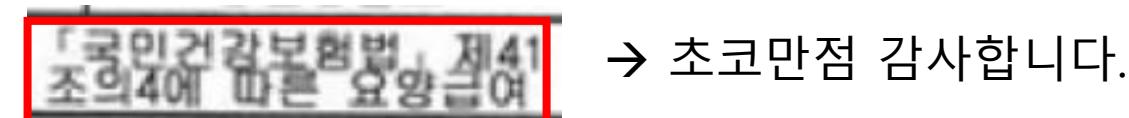
→ OCR은 정직합니다.



OCR의 한계점

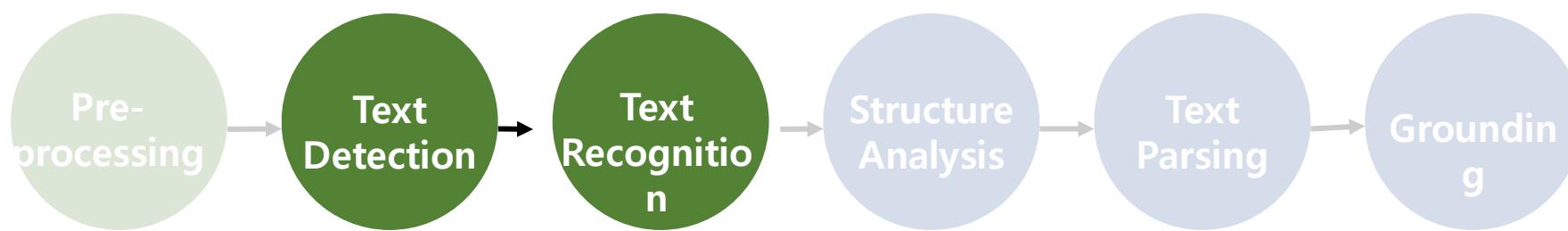
Multi line detection

- 텍스트간 간격이 적은 경우 multi line text가 single box 검출



"On Recognizing Texts of Arbitrary Shapes with 2D Self-Attention"

Paper: <https://arxiv.org/abs/1910.04396>, GitHub: <https://github.com/clovaai/SATRN>



E2E OCR: DEER(Detection-agnostic End-to-End Recognizer for Scene Text Spotting)

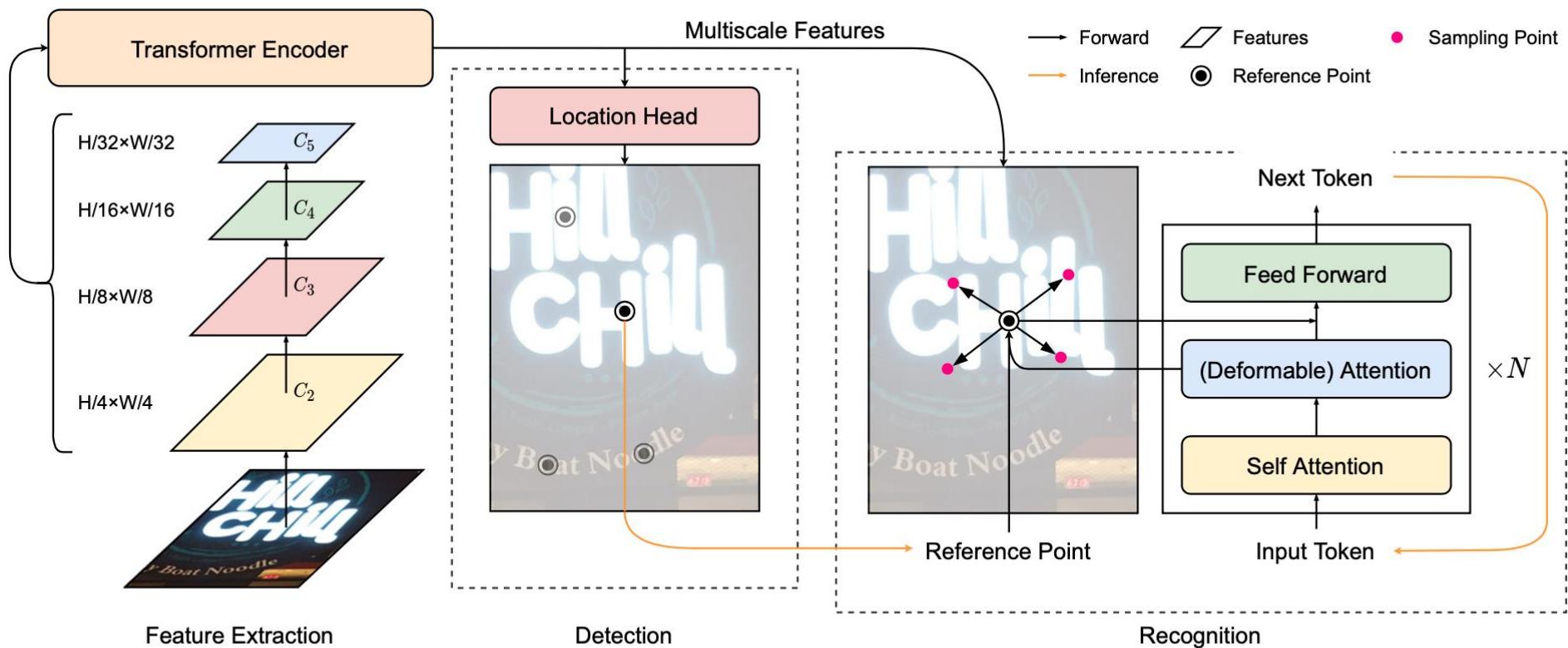
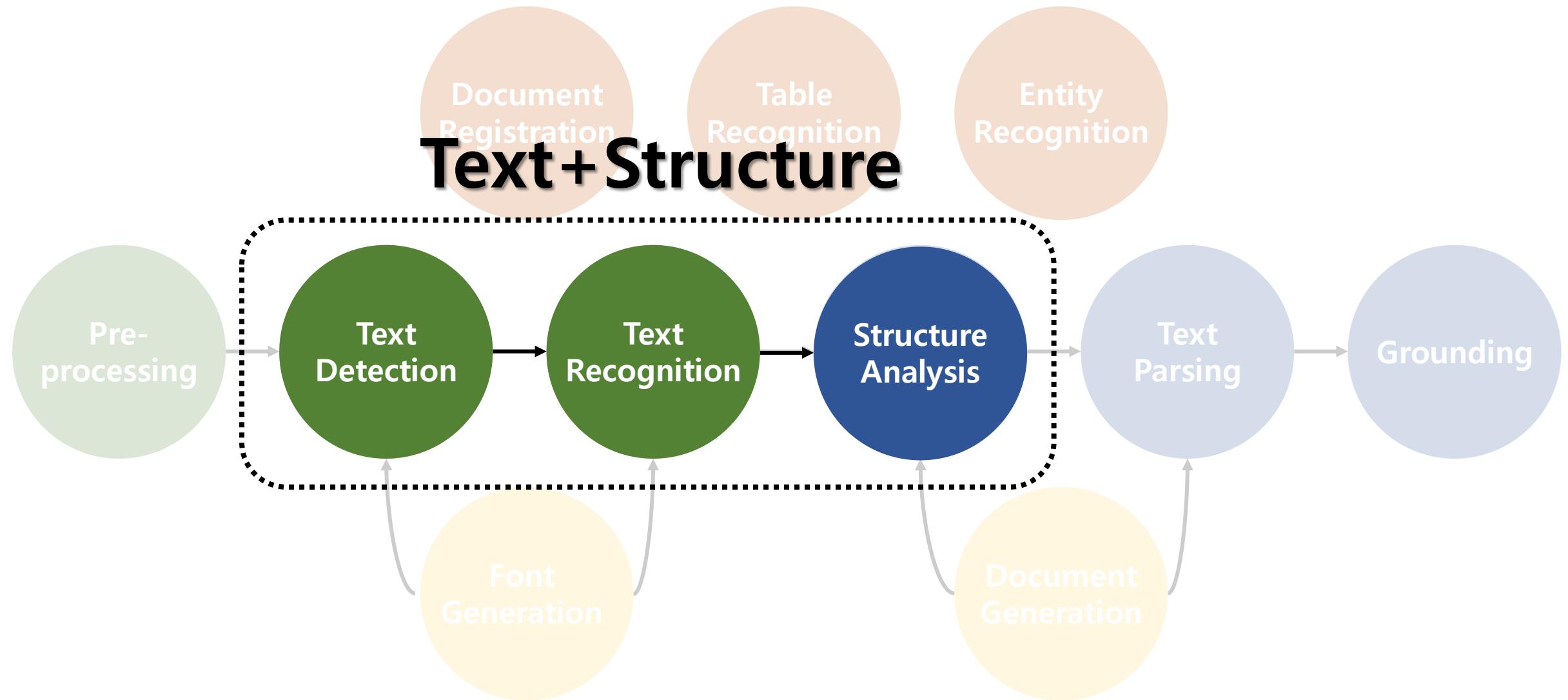
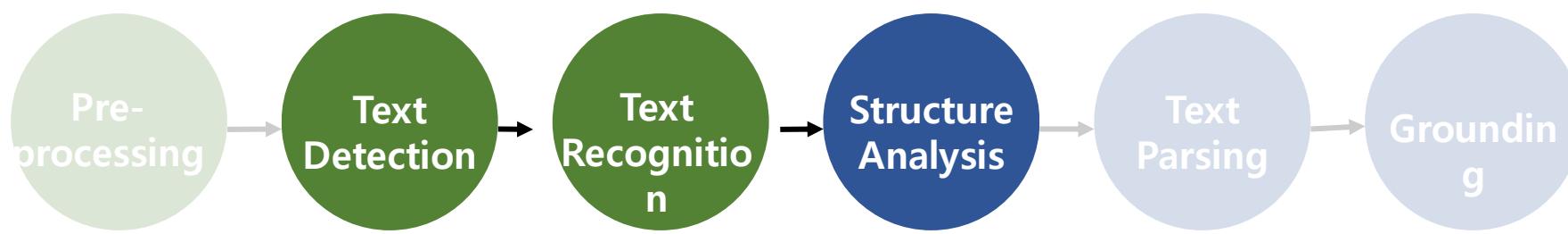


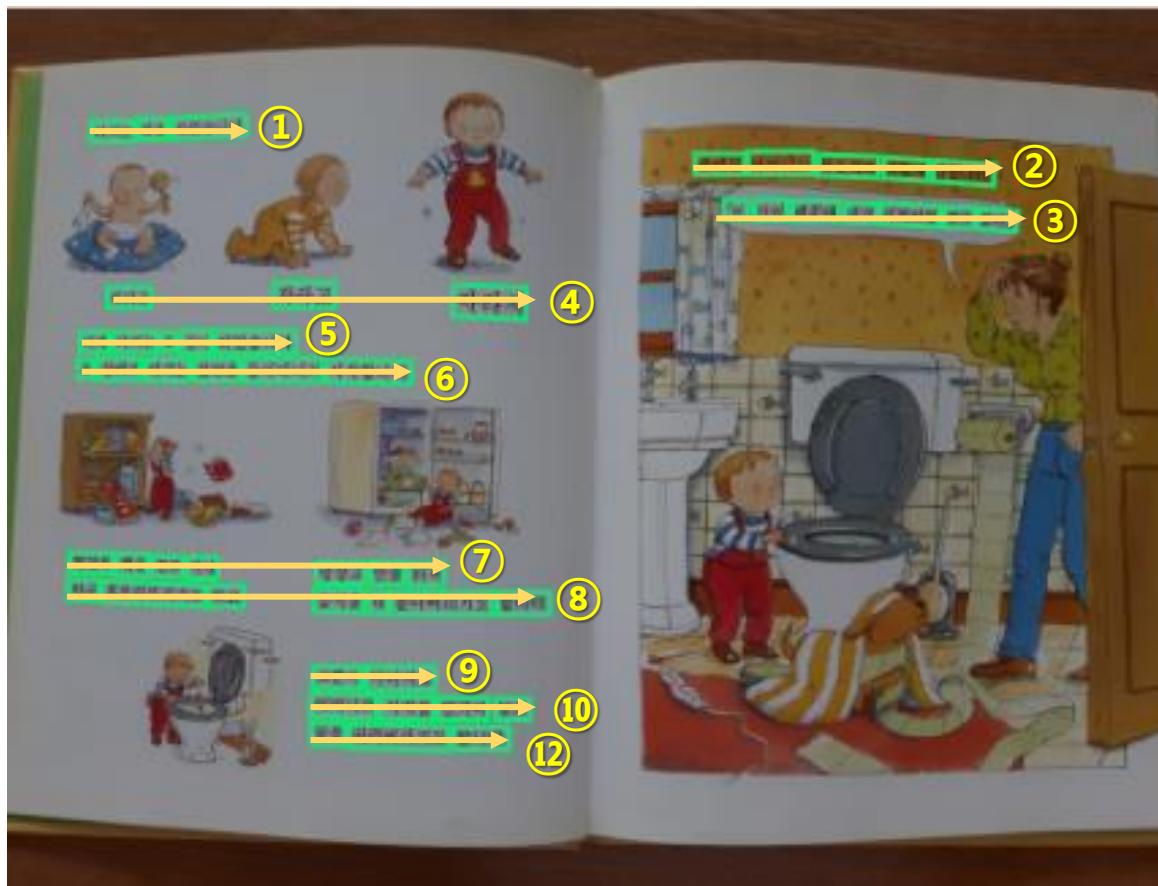
Figure 3. Overview of the proposed method. Refined feature tokens from the encoder is used as an input to the location head and the text decoder. During inference, center points from the location head is used as reference points.





Text sorting

왜 필요한가요???



Text json

아기는 점점 자랐습니다.

때때로 어머니는 한숨지며 이렇게 말합니다.

「이 아이 때문에 내가 미쳐버릴 것만 같아.」

자라고 자라고 자라서

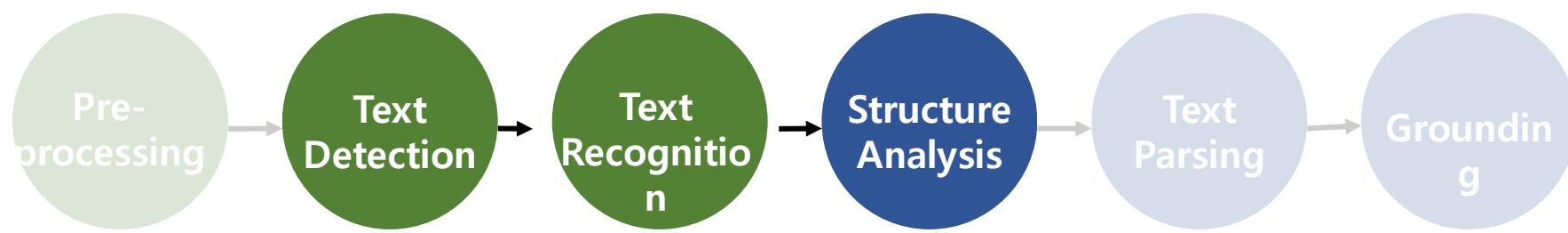
이제 아기는 두 살이 되었습니다.

두 살배기 아기는 집안을 돌아다니기 시작합니다.

책장의 책을 전부 꺼내 냉장고 안을 뒤져

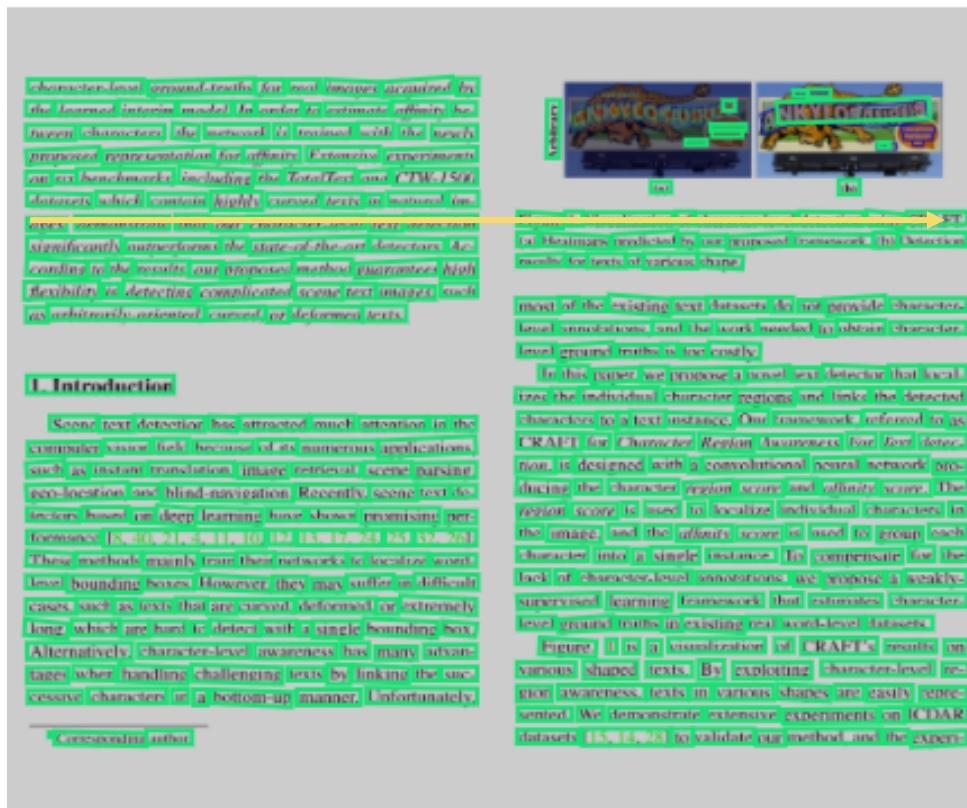
마구 흐트러뜨리기도 하고, 음식을 다 쓸아버리기도 합니다.

그뿐이 아니지요



Text structure

Text sorting만으로 충분할까요???

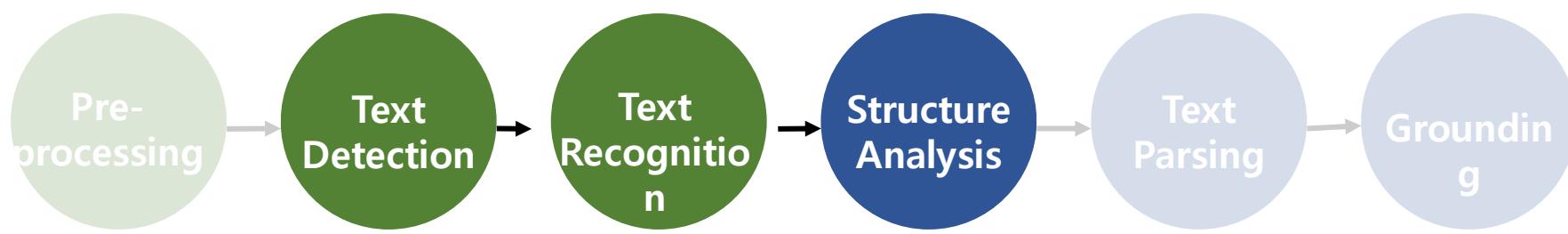


Text json

character into a single instance. To compensate for the level bounding boxes, However, they may suffer in difficult lack of character-level annotations, we propose a weakly-supervised learning framework that estimates character-level ground truths in existing real word-level datasets. Figure 1 is a visualization of CRAFT's results on various shaped texts. By exploiting character-level region awareness, texts in various shapes are easily represented. We demonstrate extensive experiments on ICDAR datasets [15, 14, 28] to validate our method, and the ex-

long, which are hard to detect with a single bounding box, level ground truths in existing real word-level datasets. Alternatively, character-level awareness has many advantages when handling challenging texts by linking the suc-

Figure 1 is a visualization of CRAFT's results on various shaped texts. By exploiting character-level region awareness, texts in various shapes are easily represented. We demonstrate extensive experiments on ICDAR datasets [15, 14, 28] to validate our method, and the ex-



Text structure

경쟁사는 어떤가요???

Objects	Labels	Logos	Text	Properties	Safe Search
	바나나우유 대한민국 대표하는 No.1 가공우유		+Block 8 +Paragraph 1 1974년에 태어난 빙그레 바나나 맛 우유는 그 달콤한 맛과 독특한 용기 모양으로 전 국민의 사랑을 받고 있는 가공우유 No.1 제품입니다. 단지 우유, 할아리 우유, 뚱 바라고도 불리는 빙그레 바나나 맛 우유와 함께 마음 까지 가득 채워세요~!		
	대한민국 대표 가공우유 빙그레 바나나우유 1974년에 태어난 빙그레 바나나우유는 그 달콤한 맛과 특별한 디자인으로 전 국민의 사랑을 받고 있는 가공우유 No.1 제품입니다. 단지 우유, 할아리 우유, 뚱 바라고도 불리는 빙그레 바나나 맛 우유와 함께 마음 까지 가득 채워주세요~! * 빙그레 디자인 우유, 할아리 우유, 뚱바라는 빙그레의 등록상표입니다.	 240ml 208kcal 1974.06	+Block 9 +Paragraph 1 * 빙그레 단지 우유, 할아리 우유, 뚱바는 빙그레의 등록상표입니다.		

banana_milk.png

→ 단순 text order이 아닌, text structure가 필요

LayoutLM: Pre-training of Text and Layout for Document Image Understanding

Yiheng Xu*, charlesyihenxu@gmail.com Harbin Institute of Technology Minghao Li* liminghao1630@buaa.edu.cn Beihang University Lei Cui lecui@microsoft.com Microsoft Research Asia Shaobin Huang sbhuang@microsoft.com Microsoft Research Asia Furu Wei furiwei@microsoft.com Microsoft Research Asia Ming Zhou mingzhou@microsoft.com Microsoft Research Asia

1. INTRODUCTION

Pre-training techniques have been verified successfully in a variety of NLP tasks in recent years. Despite the widespread use of pre-training models for NLP applications, they almost exclusively focus on text-level manipulation, while neglecting layout and style information that is vital for document image understanding. In this paper, we propose the LayoutLM to jointly model interactions between text and layout information across scanned document images. We first introduce the pre-training framework of LayoutLM, which aims to learn document image understanding tasks such as information extraction from scanned documents. Furthermore, we also leverage image features to incorporate words' visual information into LayoutLM. To the best of our knowledge, this is the first time that text and layout are jointly learned in a single framework for document-level pre-training. It achieves state-of-the-art results in several downstream tasks for document understanding (from 70.72 to 79.27, recent understandings from 94.02 to 95.33) and document image classification (from 93.07 to 94.42). The code and pre-trained LayoutLM models are publicly available at <https://aka.ms/layoutlm>.

CCS CONCEPTS

- Information systems • Business intelligence • Computing methodologies • Information extraction: Transfer learning • Applied computing • Document analysis

KEYWORDS

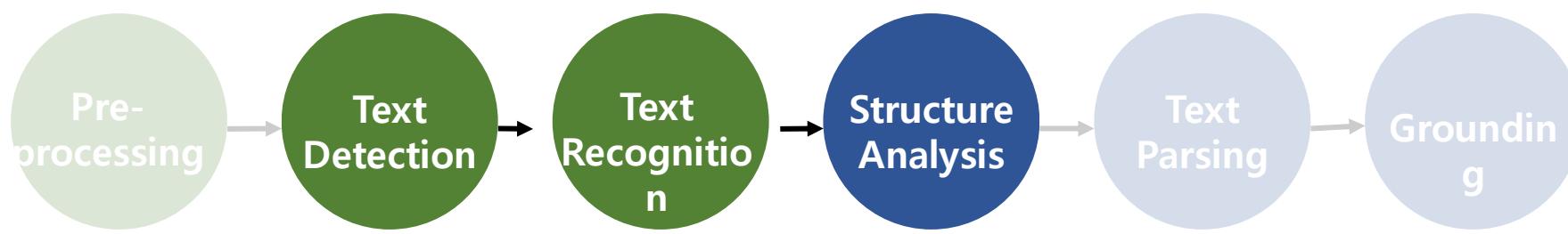
LayoutLM, pre-training, document image understanding, ACM Reference Format:

This work is licensed under a Creative Commons License by Lei Cui, Shaobin Huang, Furu Wei, and Ming Zhou - 2020, LayoutLM: Pre-training of Text and Layout for Document Image Understanding, In Proceedings of the 36th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '20), August 23–27, 2020, Virtual Event, CA, USA. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3394486.3401177>

Panel contributions during training at Microsoft Research Asia

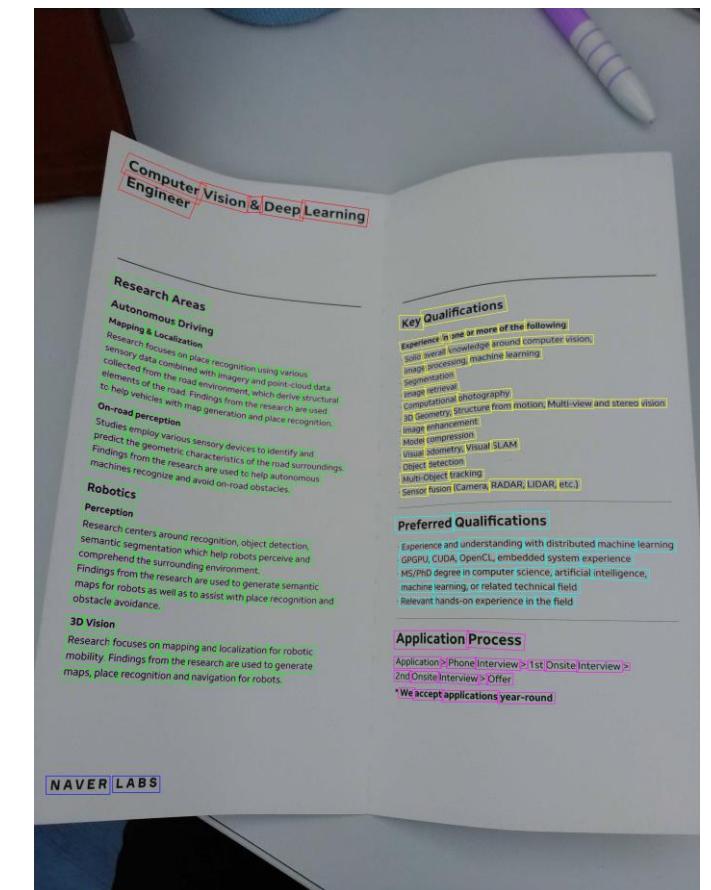
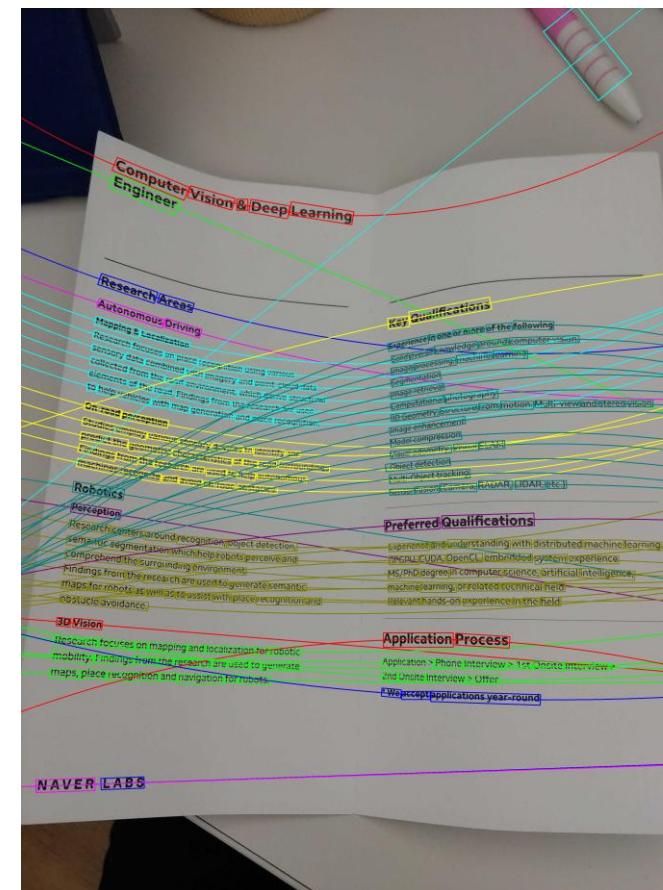
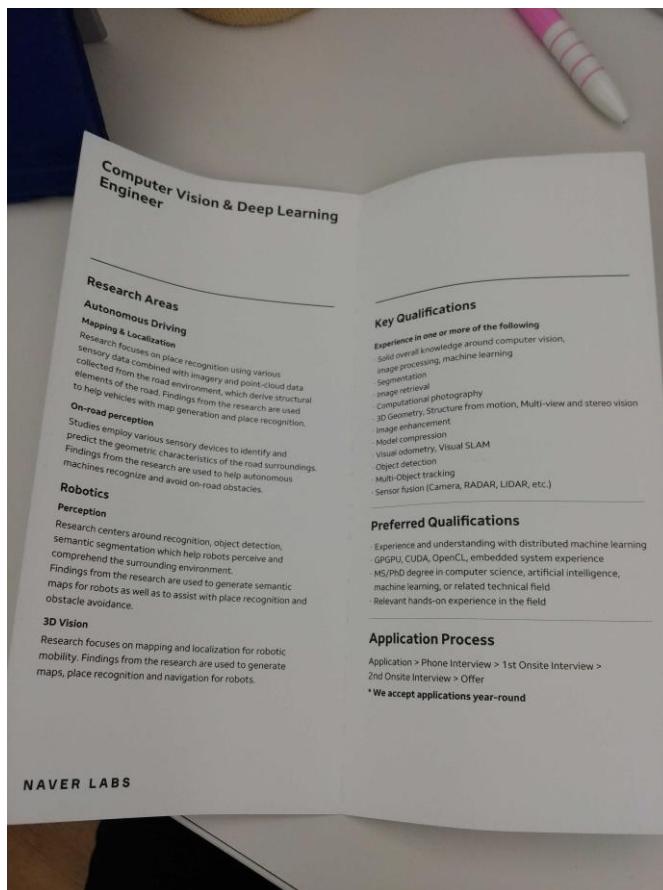
Permissions to make digital or hard copies of all or part of this work for personal or internal use, or the use of the work for educational purposes and/or research, or to post on servers or to redistribute to lists requires prior permission or a fee. Request permission to post a copy or permission to make multiple copies of a document, such as to list in a series, 1) is to request a table reading method for PDF documents based on Convolutional Neural Networks (CNN). After that, [21, 24, 29] also leveraged more advanced Faster R-CNN model [19] or Mask R-CNN model [6] to further improve the accuracy of document layout analysis. In addition, [28] presented an end-to-end multimodal, fully convolutional network for extracting semantic structures from document images, taking advantage of text embeddings from pre-trained NLP models. More recently, [11] introduced a Graph Convolutional Network (GCN) based model to combine textual and visual information to

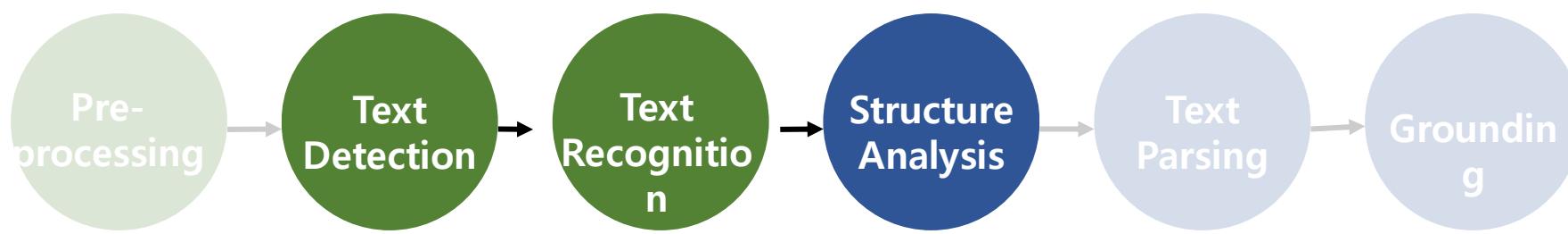
+Page 1
+Block 1
+Paragraph 1
LayoutLM : Pre - training of Text and Layout for Document Image Underst anding
+Paragraph 2
Document Image Underst anding
+Block 2
+Paragraph 1
Yiheng Xu * charlesyihenxu@gmail.com Harbin Institute of Technology
+Block 3
Minghao Li * liminghao1630@buaa.edu.cn
+Paragraph 2
Beihang University



책읽기 구조화 방법 제안

Line regression sorting



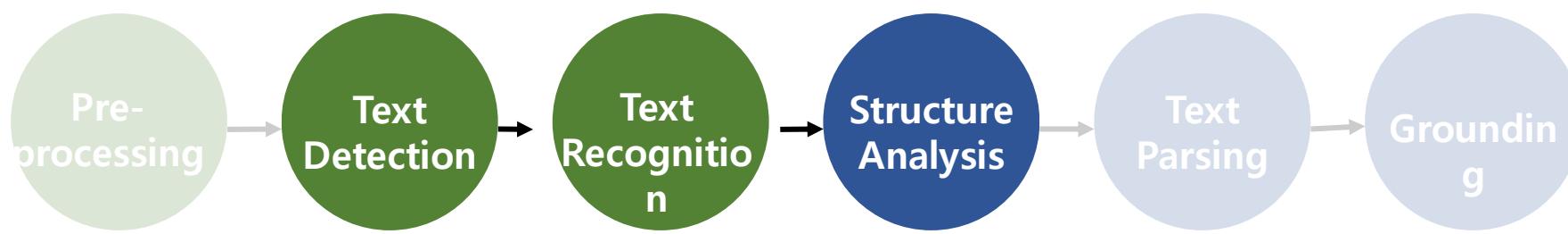


책읽기 구조화 방법 결과

Clustering + Line regression + Page split



아기는 점점 자랐습니다.
자라고 자라고 자라서
이제 아기는 두 살이 되었습니다.
두 살배기 아기는 집안을 돌아다니기 시작합니다.
책장의 책을 전부 머금고 냉장고 안을 뒤져
책장을 전부 머금고 냉장고 안을 뒤져 음식을 다 쏟아버리기도 합니다.
그뿐이 아니지요.
어머니의 시계를 변기에 넣고 물을 내려버리기도 합니다.
때때로 어머니는 한숨지며 이렇게 말합니다.
「이 아이 때문에 내가 미쳐버릴 것만 같아.」



OCR+ sorting application

OCR + TTS를 결합한 책읽기 기능



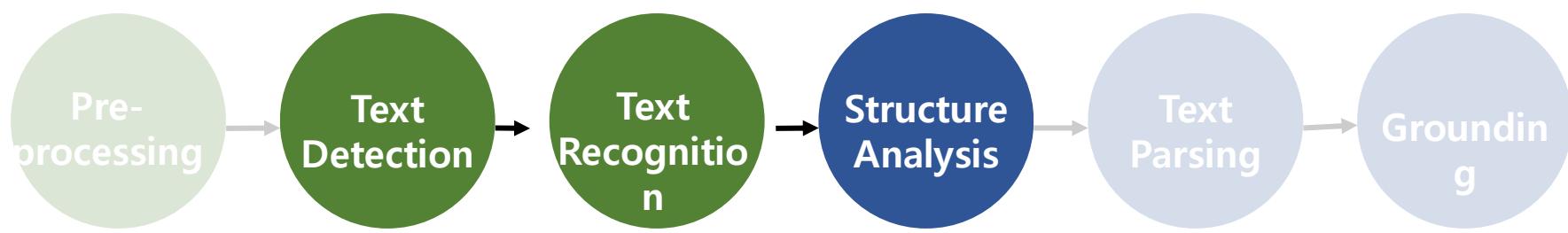
구매하기 장바구니 담기 선물하기

상세정보 리뷰 664 Q&A 308

★★★★★ 4
park**** · 20.10.29. · 신고

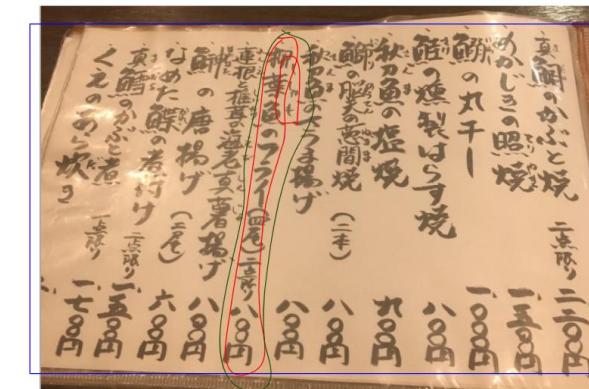
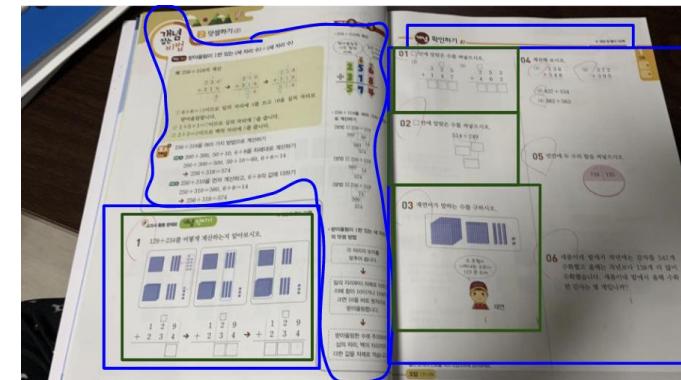
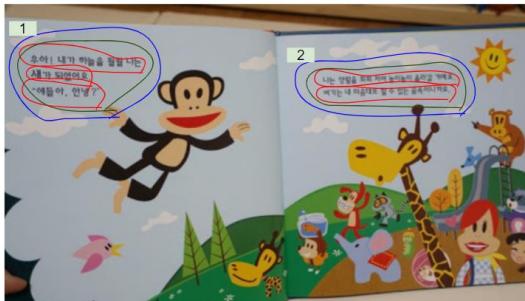
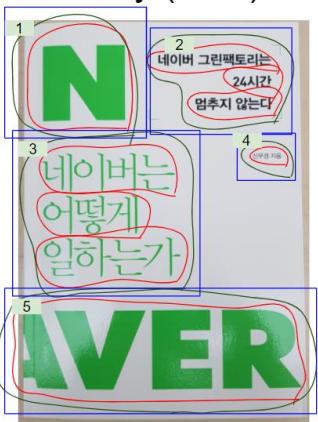
본인은 클로바램프가 출시되기 전부터 이미 클로바 인공지능 스피커를 사용하고 있었습니다. 사실 클로바램프에서 제공하는 조명 기능과 독서 기능을 제외하면 기존의 클로바 스피커의 기능과 거의 다를 것이 없는 것 같았기 때문에 가성비가 조금 높지 않은가 하는 생각이었습니다.
그러나 시각장애를 가진 본인에게는 독서 기능은 꼭 사용하지 않으면 안 될 중요한 기능이었기 때문에 큰마음 먹고 클로바램프를 구입하기로 결심하고 주문을 했습니다.
사실 그동안 국내에서 돌아다니는 여러 가지 OCR 기능을 사용해 보았지만 독서 기능을 만족시켜 줄 이הלך다할 제품은 발견하지 못했던 터이기에, 클로바램프를 구매하면서도 또 후회하는 일을 하는 것은 아닐까? 하고 반신반의하며 제품이 도착하기를 기다렸습니다.
주문 후 약 3~4일이 지나서 제품이 도착을 했고, 저는 기대 반 의심 반으로 제품의 포장을 풀고 설정을 완료한 다음, 가장 먼저 독서 기능을 실행시켜 보았습니다.
앗! 그런데 클로바램프는 이제까지의 OCR 또는 AI 기능과는 비교할 수 없이 좋다는 것을 단번에 알 수 있었습니다.
목소리부터 시작해서 OCR의 정확도가 이제까지 그 어느 기기보다 훨씬 좋다는 것을 금방 느낄 수 있었기 때문입니다.
기존에 클로바램프와 약 2,000 여권의 책이 제휴되어 있어서, 제휴된 서적들은 음원으로 읽어준다는 말을 듣고, 음원 이야 미리 만들었을 수 있는 것인가 하고 애시당초 기대를 하지 않았습니다. 그래서 음원에 없는 도서를 테스트해 봤는데 결과는 대만족이었습니다.
한 가지 아쉬운 점이 있다면, 클로바램프의 암이 고정되어 있어서 A3 사이즈의 큰 책은 범위 밖의 문자를 읽어주지 않습니다. 램프의 암이 기존의 스텐드처럼 아래로 향하는 것입니다.

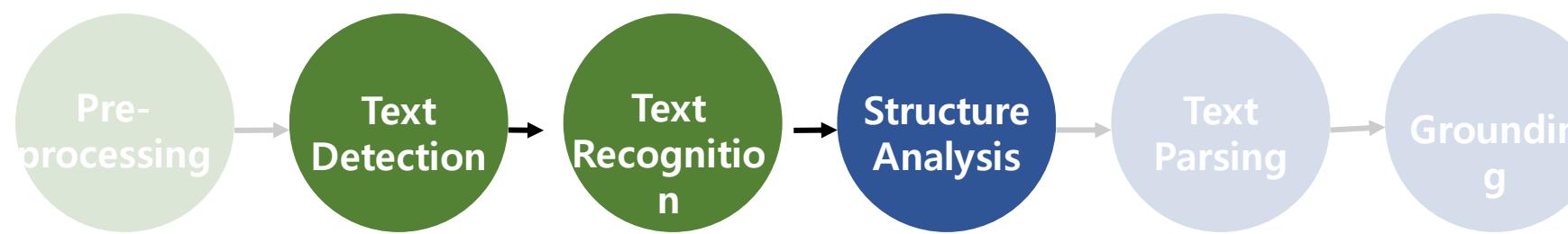
램프 후기 모음



Text structure analysis (21.1Q)

- 고차원 문서 이해를 위해서는 단순 Sorting이 아닌 구조화가 필요
- 보다 나은 구조화를 위해서 **Language + Layout** 모델 학습
- 다양한 domain에 대해서 case study

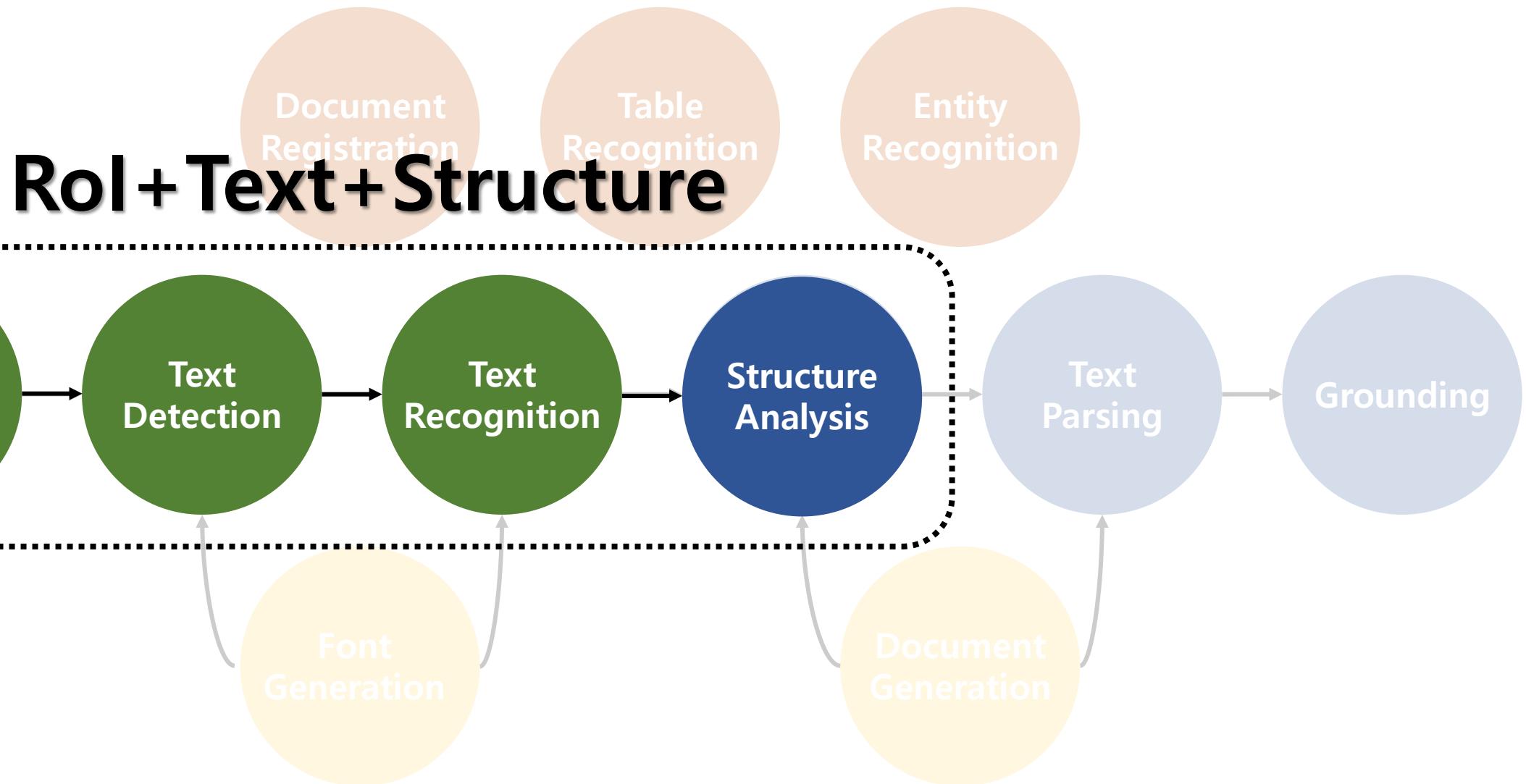


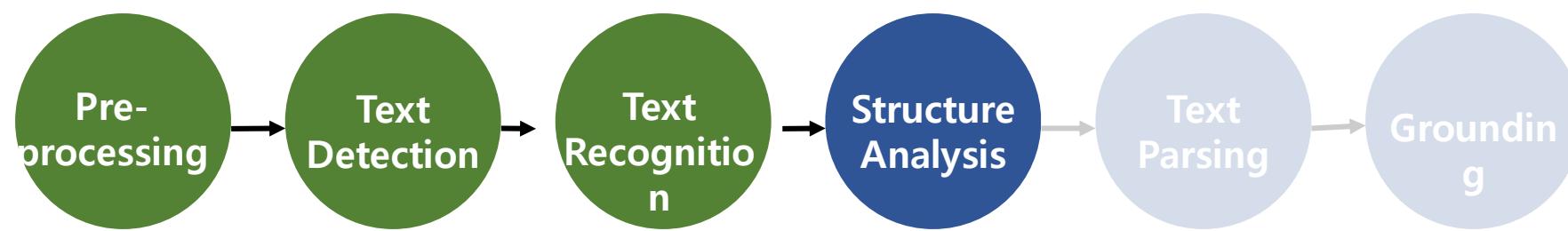


Hierarchical Text Structuring (22.1Q)

- Paragraph, Line 정보까지 heatmap 으로 추출
- 기존 deer pipeline 을 이용한 e2e 형태 구축

Type of Precaution	Definition	Example
Personal Precautions	Actions that individuals can take to reduce the risk of spreading disease	<p>Wash your hands frequently and thoroughly. See "Handwashing" in this chapter.</p> <p>Treat all blood and other body fluids as infectious materials</p> <p>Cover your mouth and nose when you sneeze</p> <p>Eat well and get enough exercise and sleep.</p>

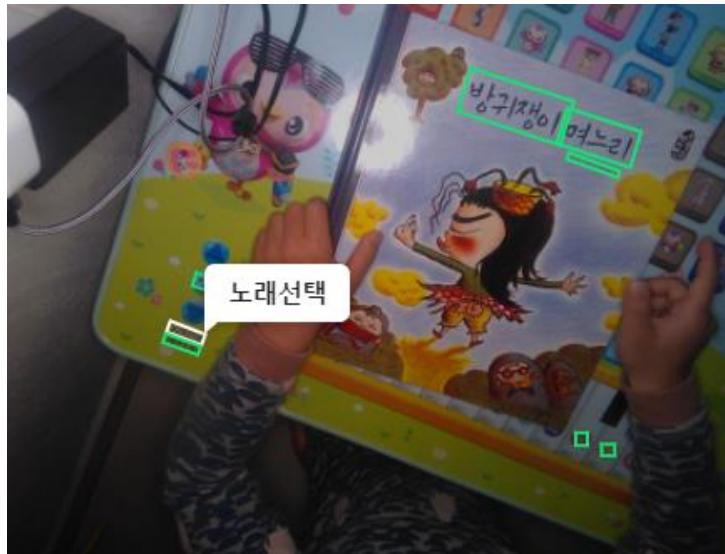


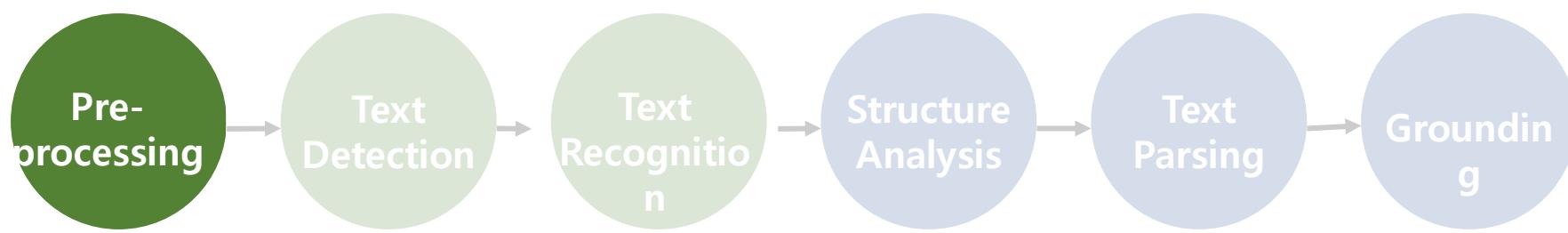


Pre-processing의 필요성

- OCR의 목적 : 이미지 내에 모든 글자 인식
- OCR을 사용하는 이용자의 목적 : 관심 영역 내에 글자 인식

관심 영역(Roi) → Region of Interest (O), Return on Investment (X)





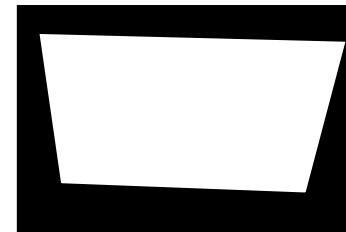
ROI 검출기

- 관심 대상체 (책 혹은 문서)의 코너를 검출



입력 이미지

ROI 검출기

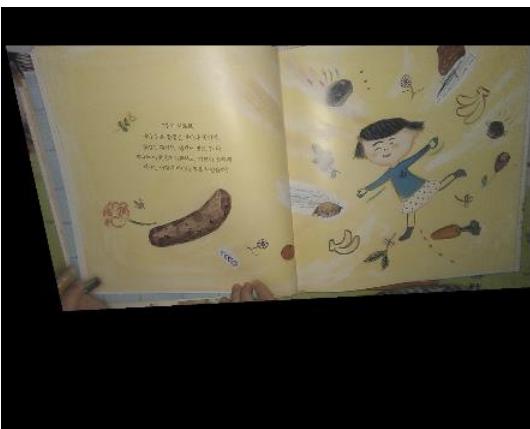


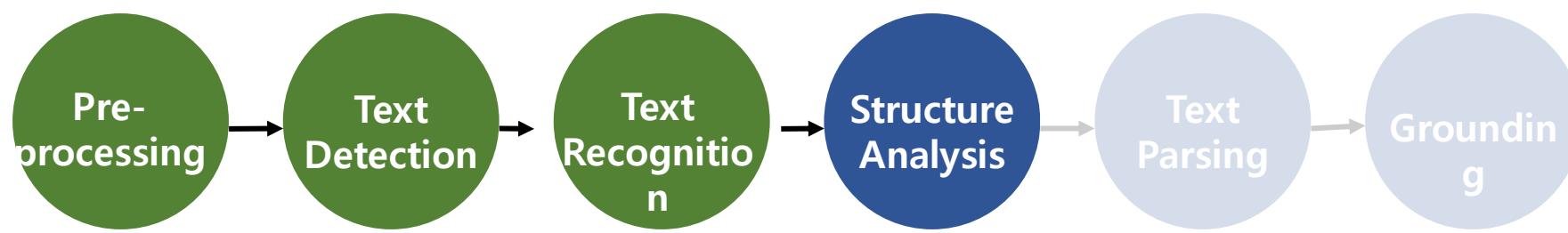
이미지 마스킹



Text
Detection

Text
Recognitio
n

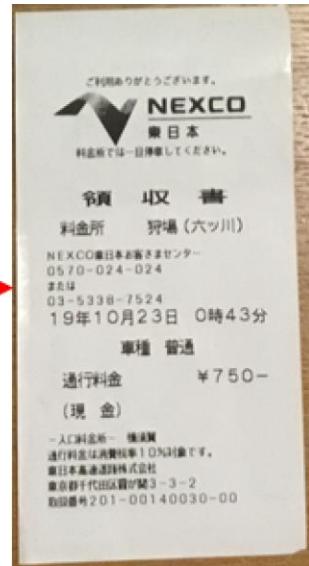




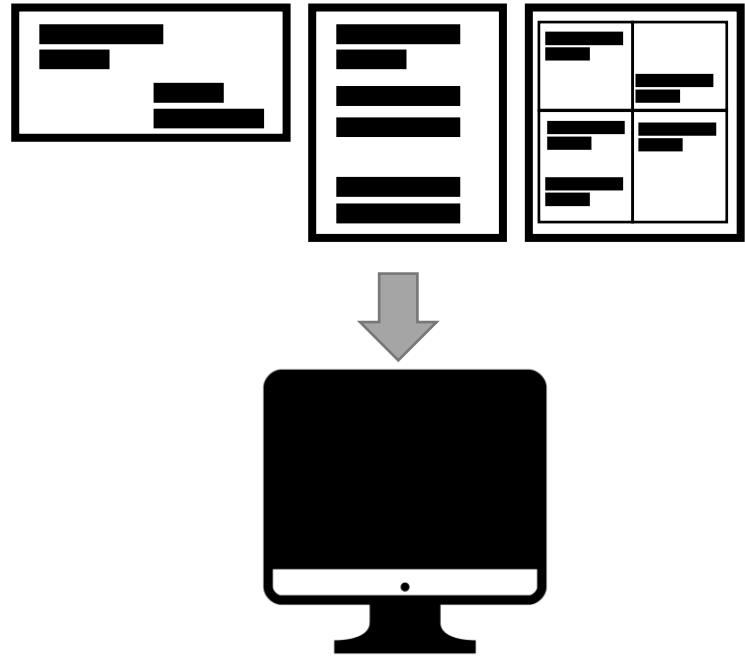
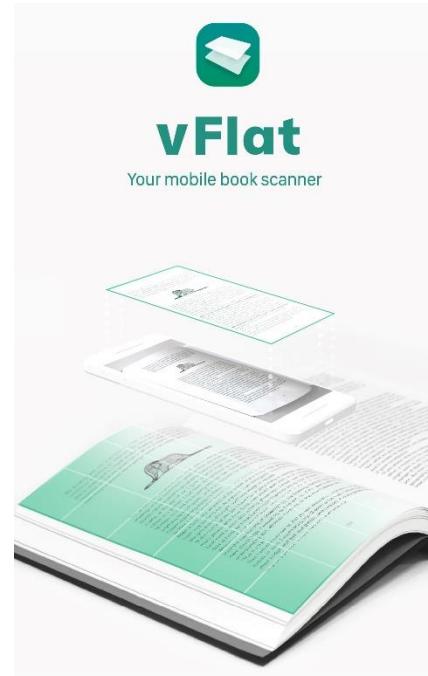
Future works of pre-processing



Multiple ROI Detection
→ Mobile SDK화
(21.1Q)

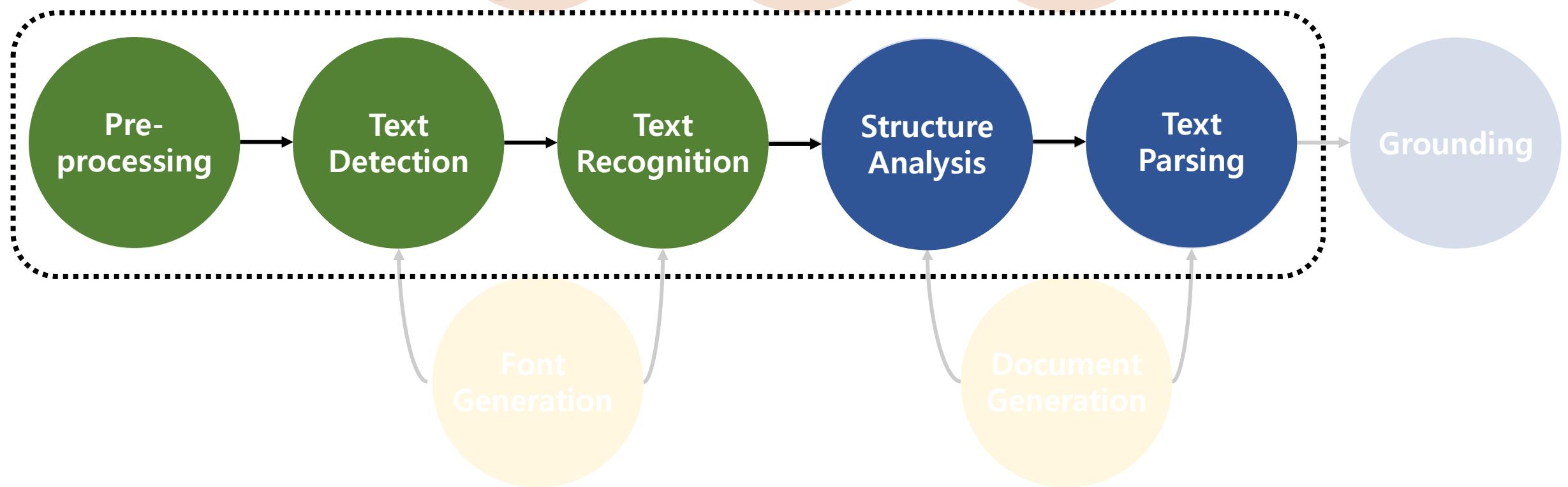


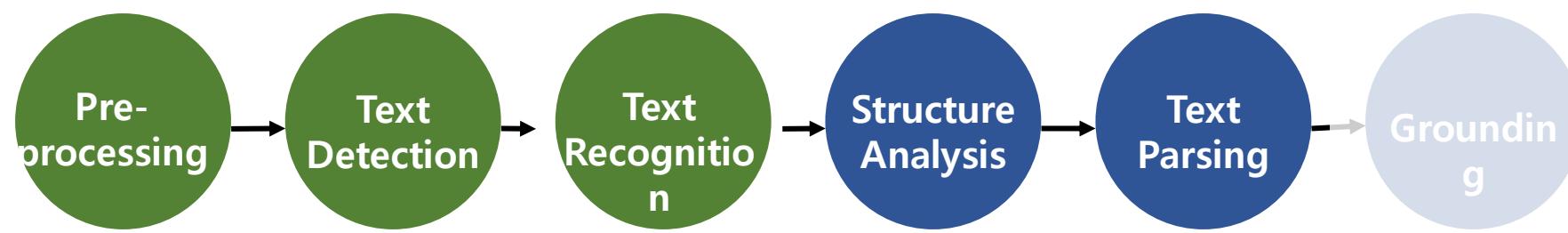
Document image rectification
→ 스캔 기능 구현
(21.2Q)



Document classification
→ 자동으로 적합한 모델로 인식
(21.3Q)

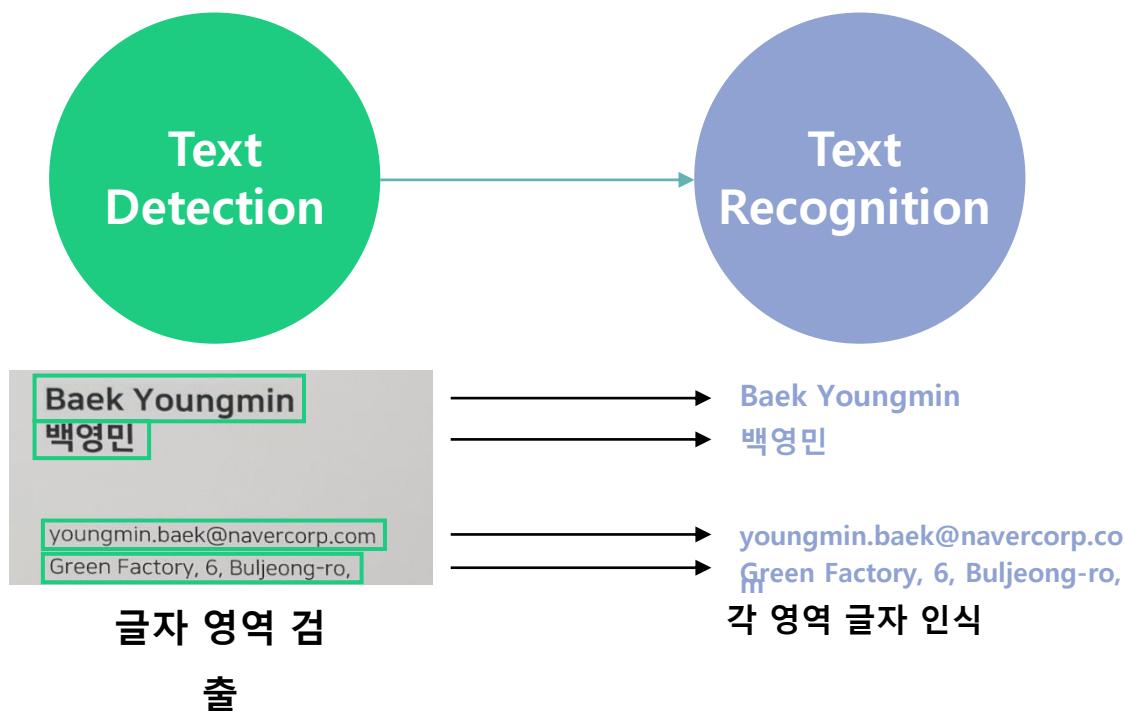
OCR specialized model

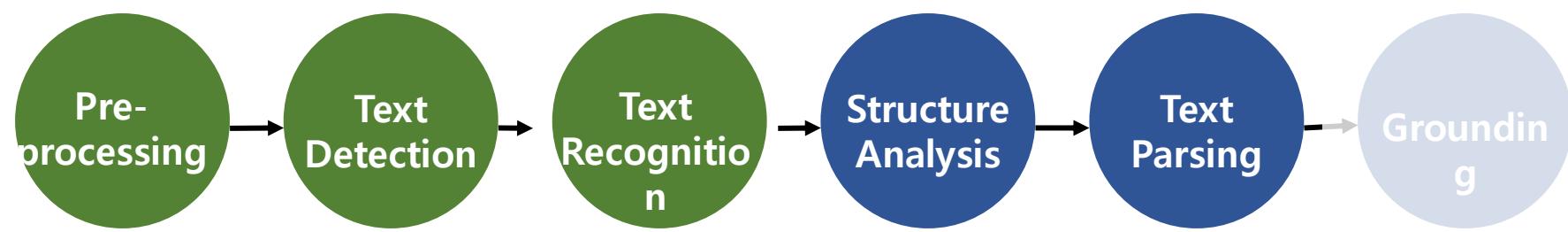




OCR (Optical Character Recognition)

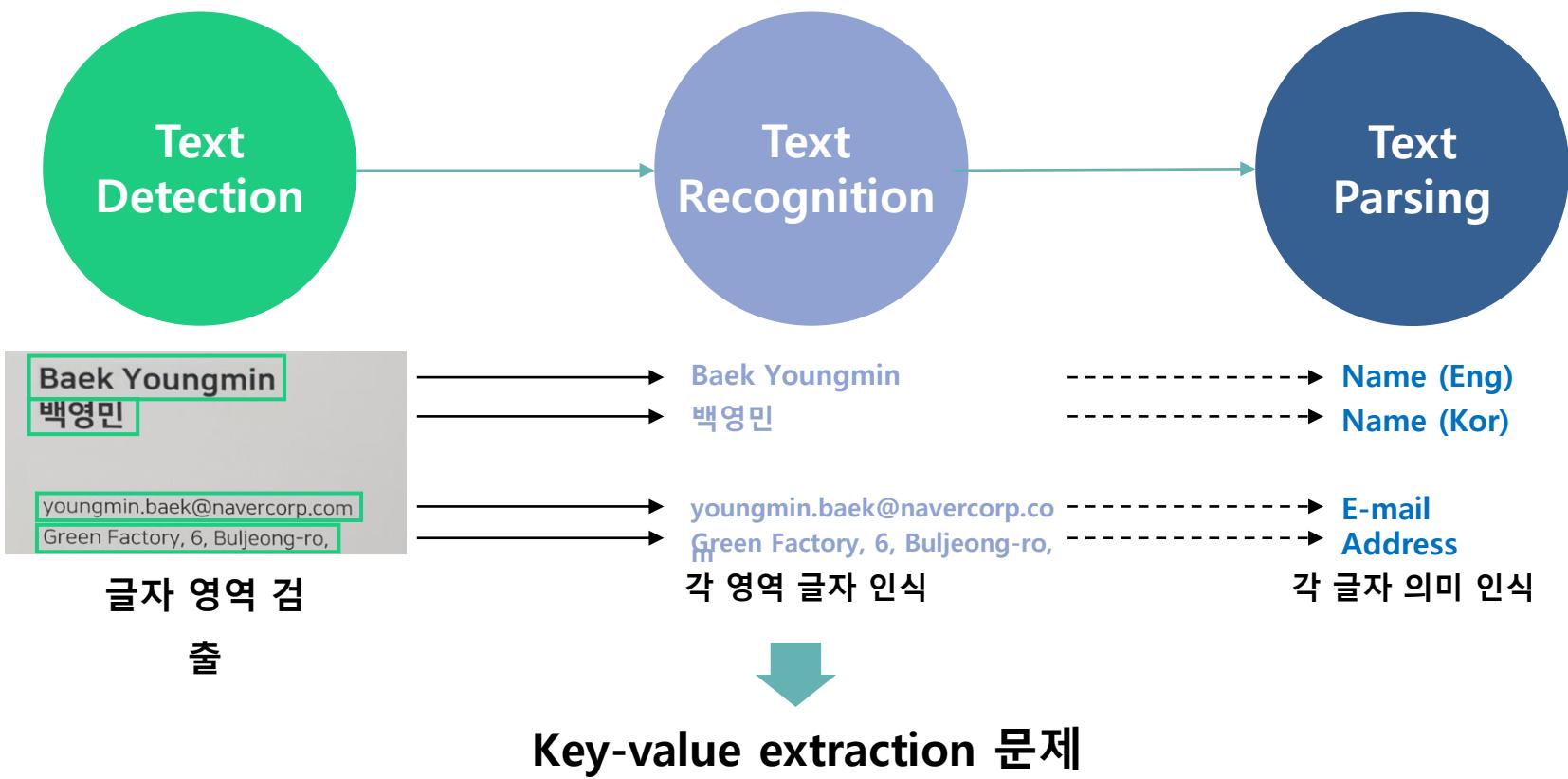
글자 영역 검출 + 각 영역 글자 인식

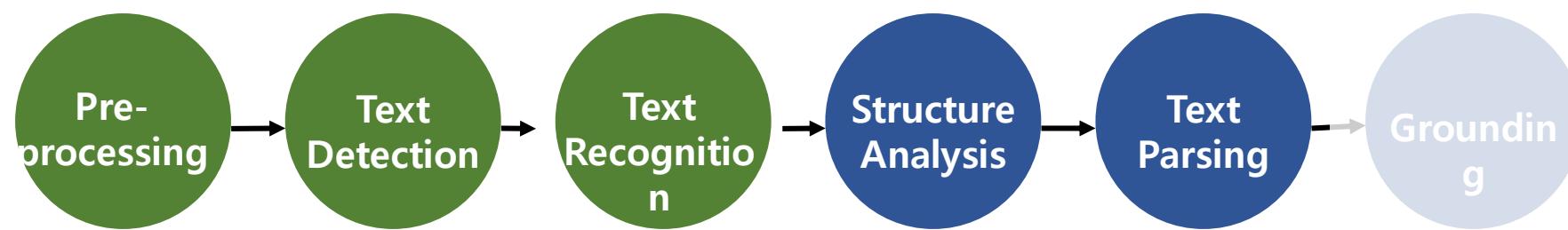




특화 모델

OCR + Parsing

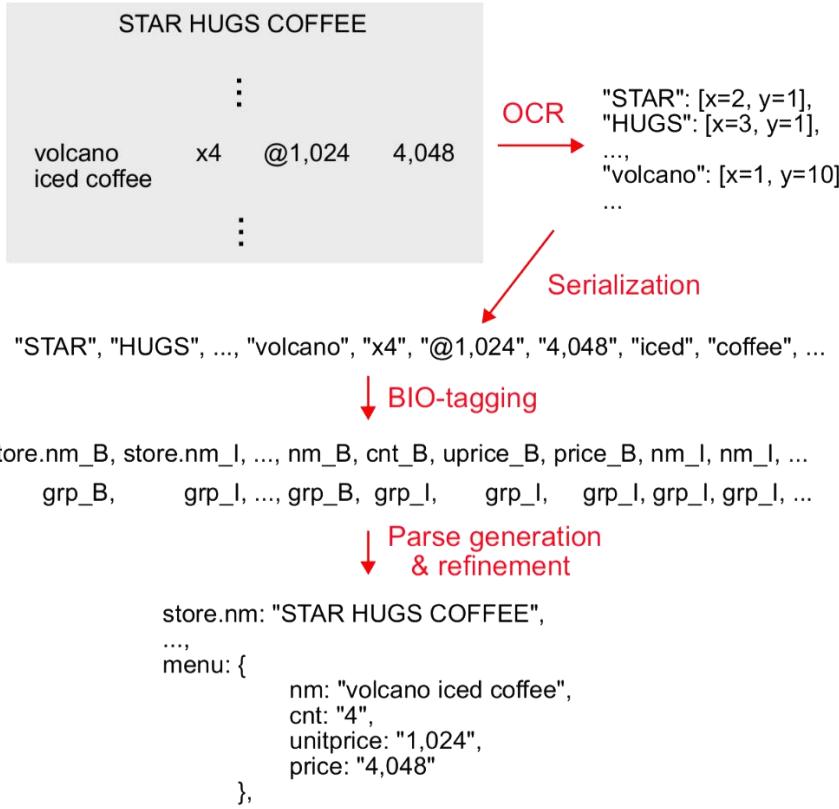




Post-OCR Parsing

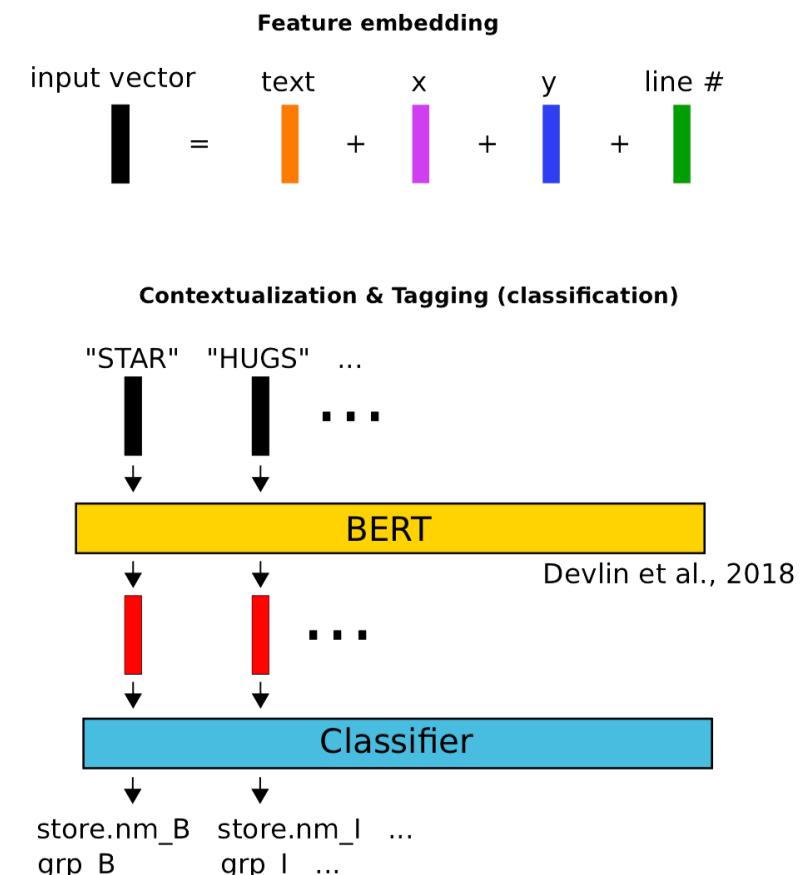
언어 모델을 통한 글자 의미 분류

Receipt image

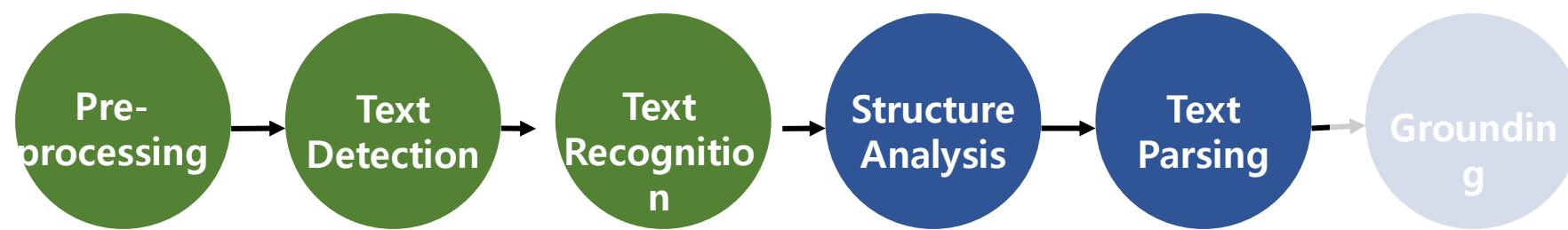


Ontology 기반 BIO tagging

<https://openreview.net/forum?id=SJgjf695UB>



BERT 기반 Text classification

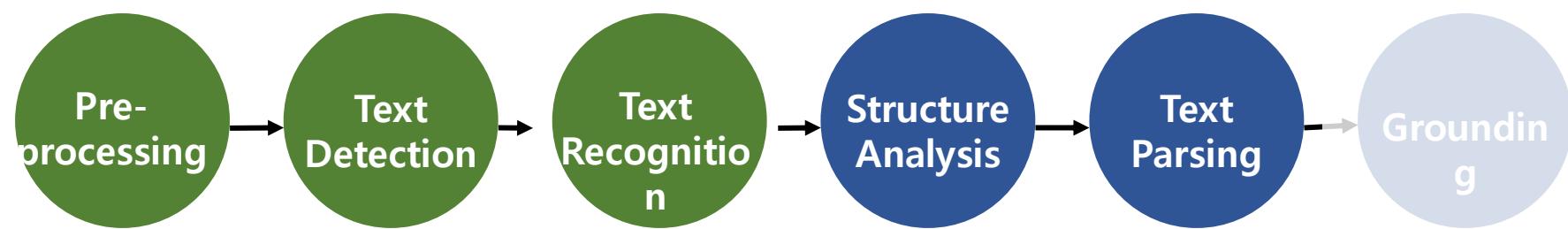


Text parsing applications

Name Card Recognition

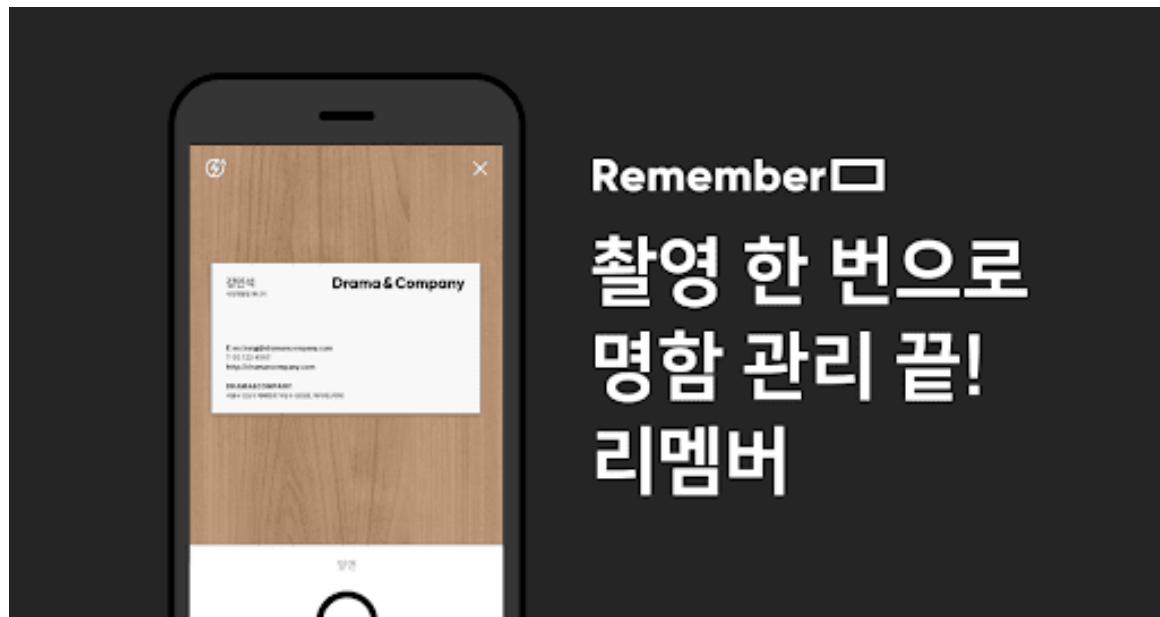


```
"result": {  
    "address": [  
        "〒101-0051 東京都千代田区神田神保町3-5  
        住友不動産九段下ビル3F"  
    ],  
    "department": [  
        "日本創発グループ",  
        "CREATIVE PLATFORM GROUP",  
        "九段センター",  
        "営業部"  
    ],  
    "email": ["t.tokyo@lithmatic.co.jp"],  
    "fax": ["0332654589"],  
    "fullName": ["東京 太郎", "Taro Tokyo"],  
    "homepage": ["http://www.lithmatic.net/"],  
    "mobile": ["09000000000"],  
    "tel": ["0332658081"]  
}
```

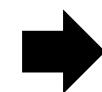


Text parsing applications

Name Card Recognition

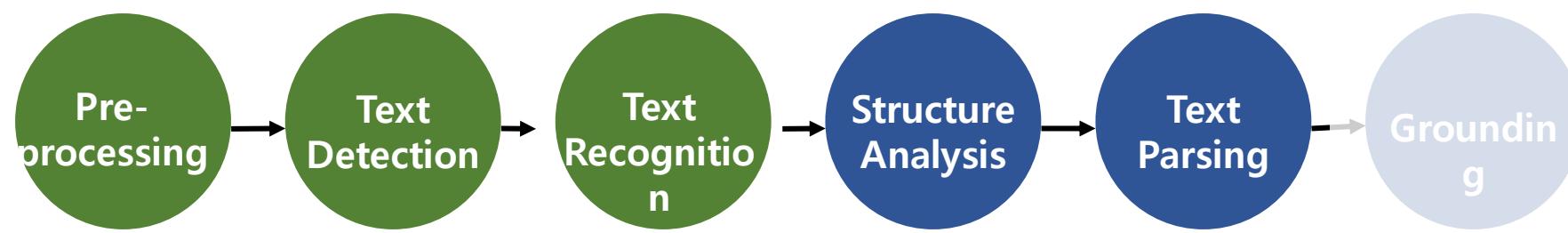


Remember / myBridge 명함 자동 인식 기능
제공

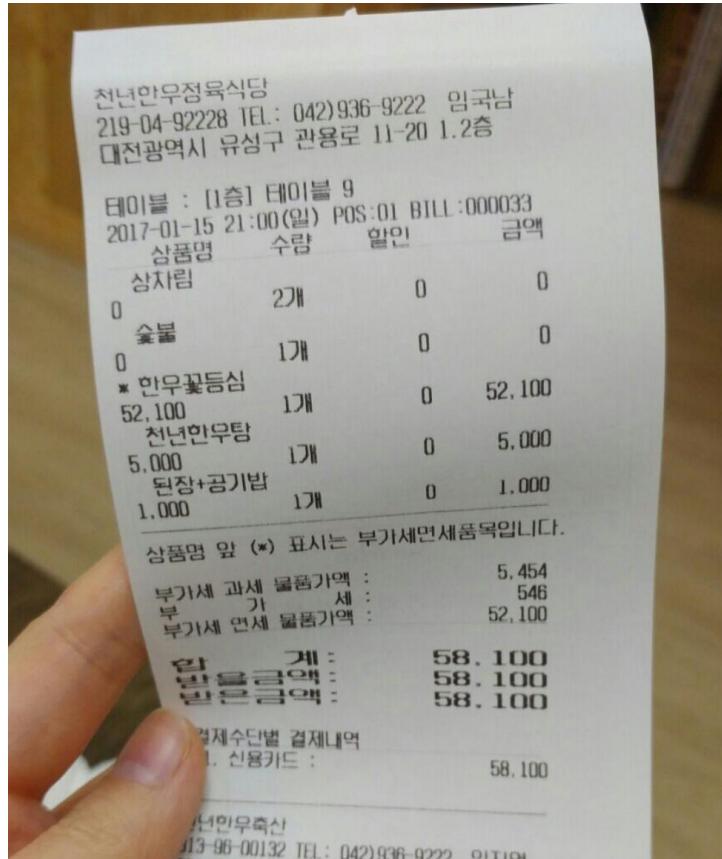


NAVER WORKS
LINE WORKS

20.4Q

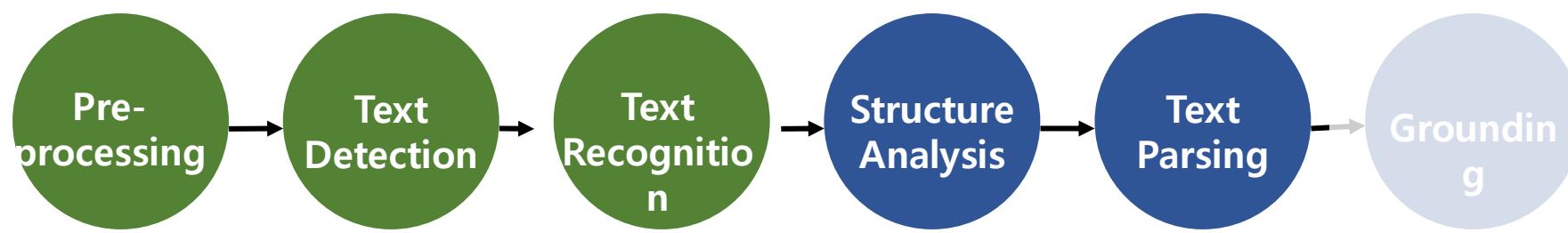


Text parsing applications



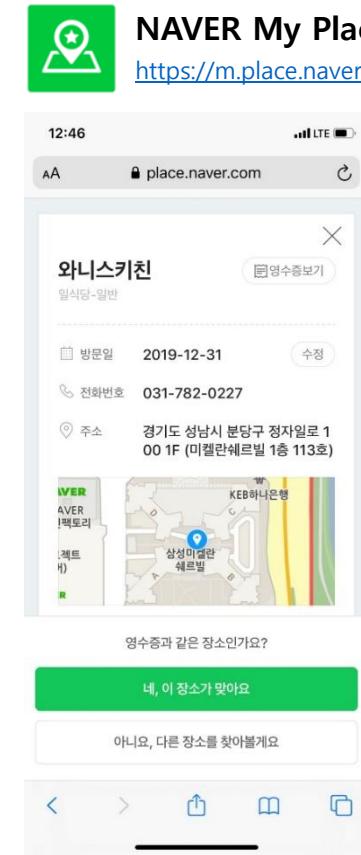
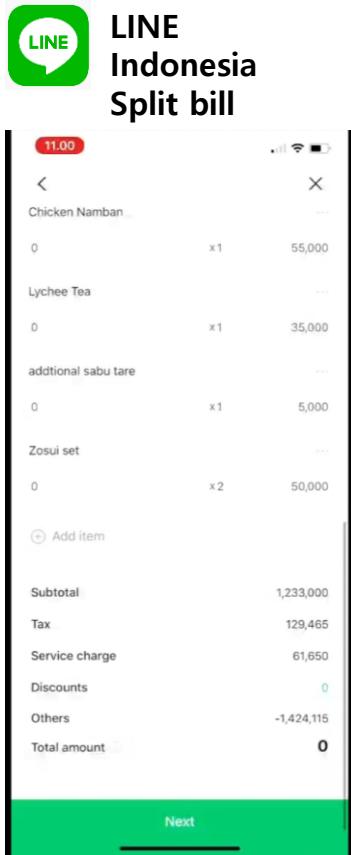
Receipt Recognition

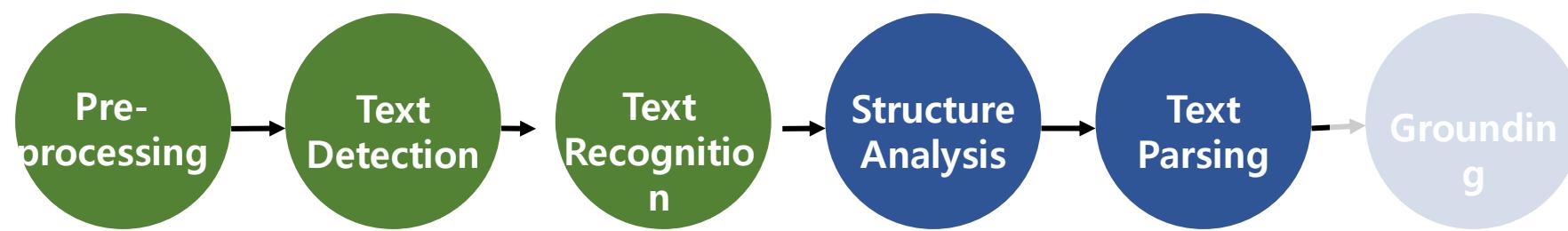
Type	Name	Value
Info	Store Address	대전광역시 유성구 관용로 11-20 1.2층
	Store Business Number	219-04-92228
	Store Name	천년한우정육식당
	Store Phone Number	042)936-9222
	Payment Date	2017-01-15
	Payment Time	21:00:00
Items	상차림	2
	숯불	1
	* 한우꽃등심	1
	천년한우탕	1
	된장 + 공기밥	1



Text parsing applications

Receipt Recognition





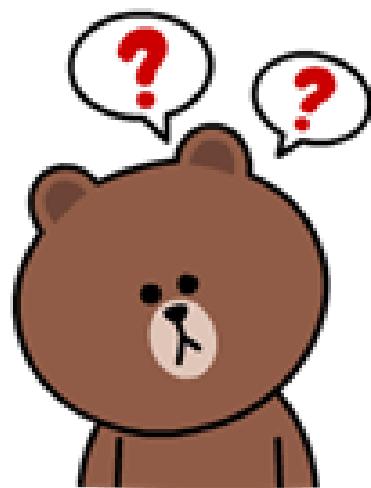
Text parsing의 한계 및 연구방향 (1/2)

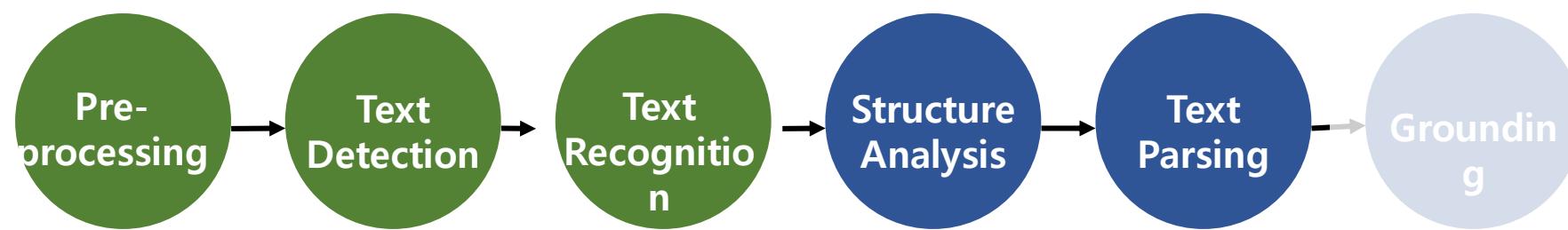
- 2D의 이미지 **layout 정보를 배제**하고 text를 1D sequence로 변환

COMMERCIAL INVOICE

Exporter The Electric Shop 1 Plate Street Brisbane, QLD, 4300 Australia	IncoDocs Number INCO-3764-CIRA-5789	Pages 1 of 1		
Consignee The SwitchGear CO. LTD Xiao Dong Shan Xiamen Fujian, 361006 China	Buyer (If not Consignee)			
Method of Dispatch Sea	Type of Shipment FCL	Country of Origin of Goods Australia Country of Final Destination China		
Vessel / Aircraft Seamax Stanford	Voyage No 638E	Terms / Method of Payment 50 percent deposit, 50 percent before shipment		
Port of Loading Brisbane	Date of Departure 22 JUN 2017			
Port of Discharge Shanghai	Final Destination Shanghai	Marine Cover Policy No Letter of Credit No LC140050		
Product Code	Description of Goods	Unit Quantity	Price	Amount
2216	Lightbulb 2 x 37 watt	1000	4.00	4000.00
2218	Elecwire 20m	450	28.00	12600.00
2226	Cable wire 40m	70	482.00	33740.00

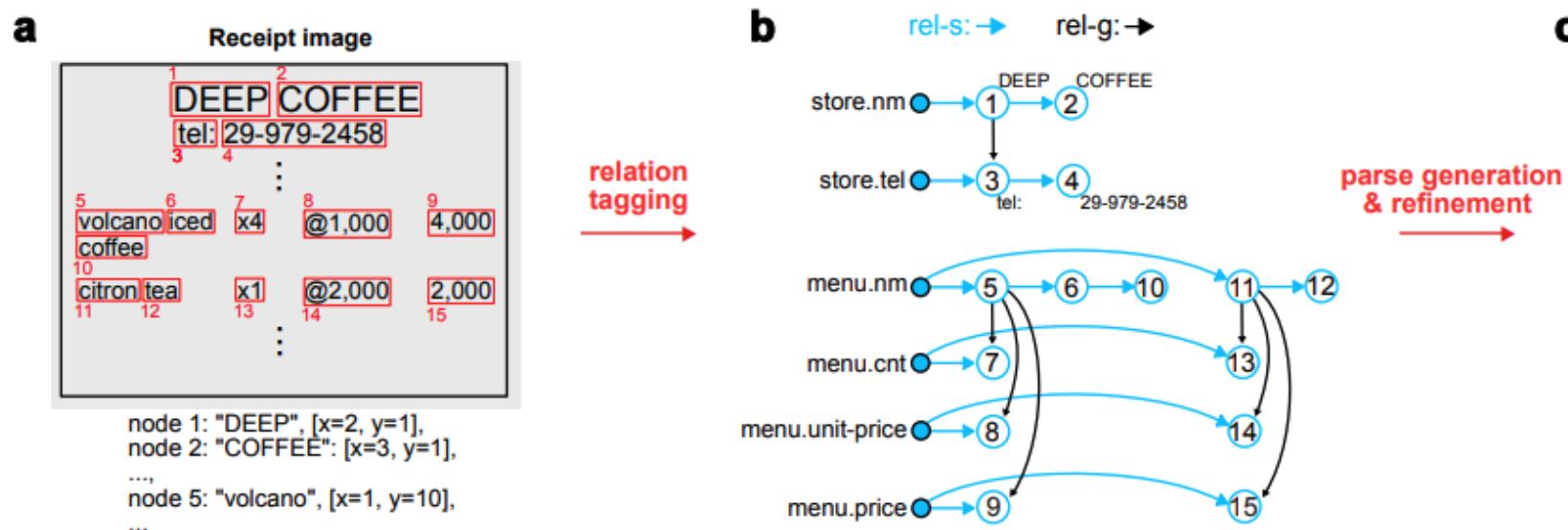
- serialization 종속적
- 복합 테이블 존재



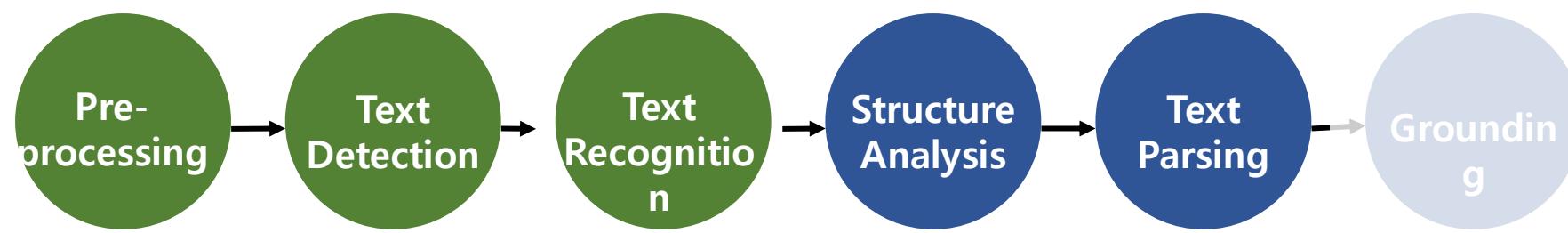


Text parsing의 한계 및 연구방향 (1/2)

- 2D의 이미지 **layout 정보를 배제**하고 text를 1D sequence로 변환 → **Spatial Dependency Parsing**

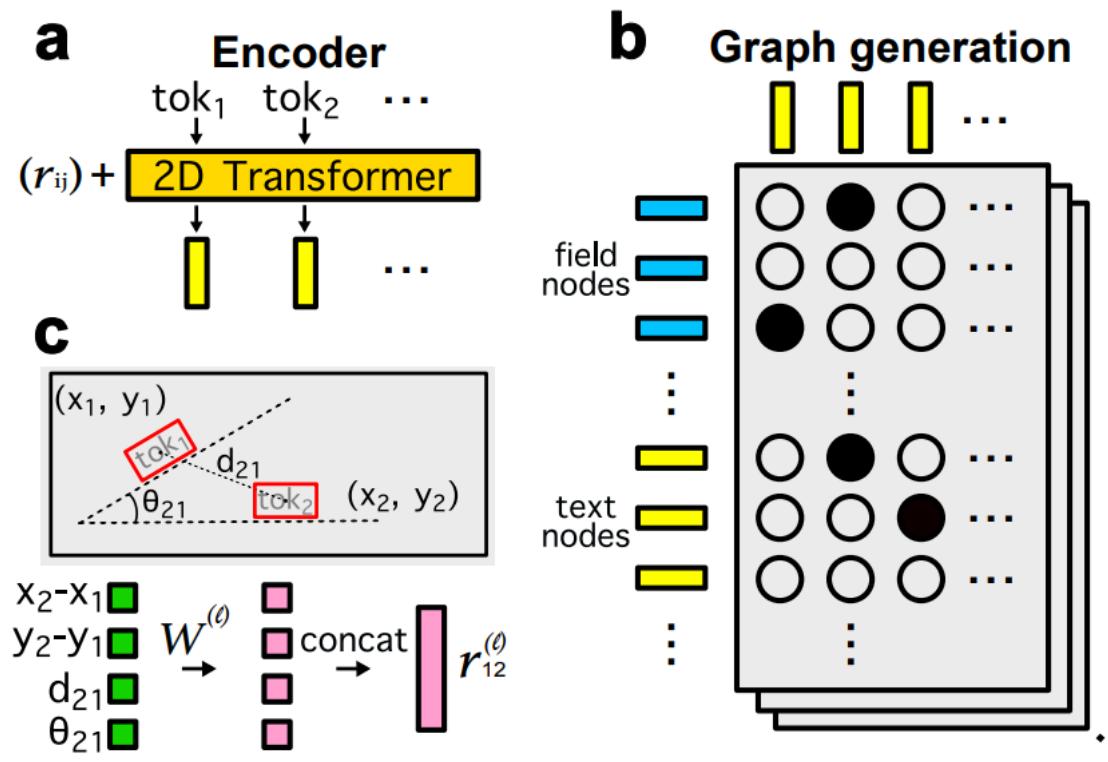


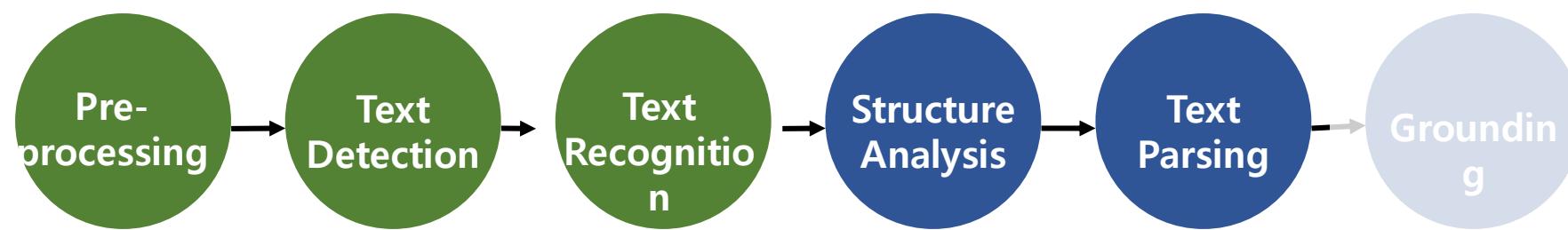
영수증 2D Relational Parsing 예제



Text parsing의 한계 및 연구방향 (1/2)

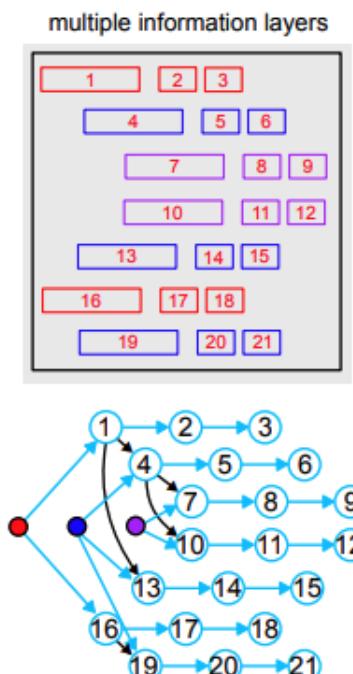
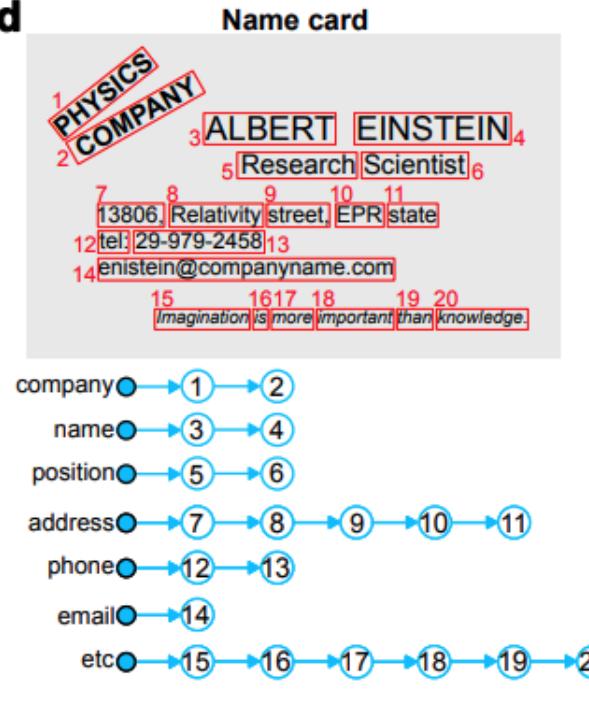
- 2D의 이미지 **layout 정보를 배제**하고 text를 1D sequence로 변환 → **Spatial Dependency Parsing**



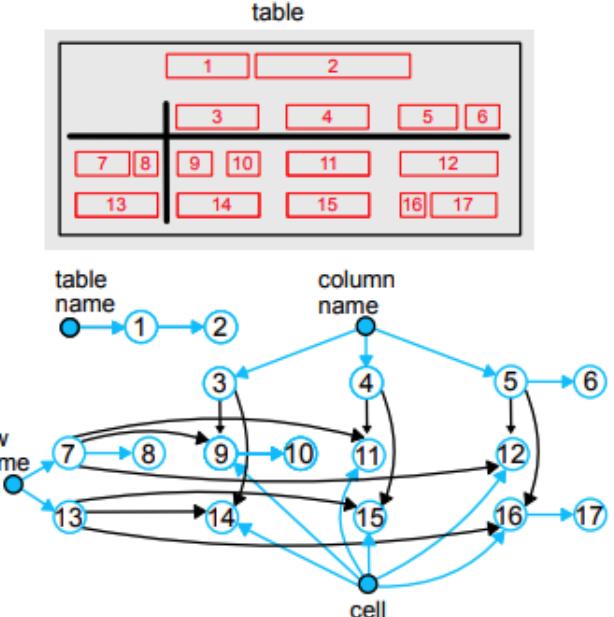
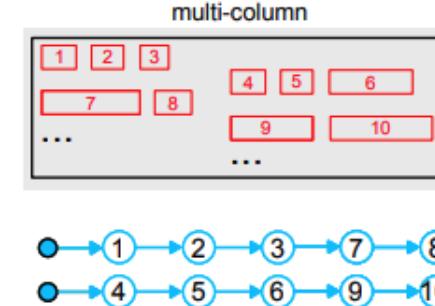


Text parsing의 한계 및 연구방향 (1/2)

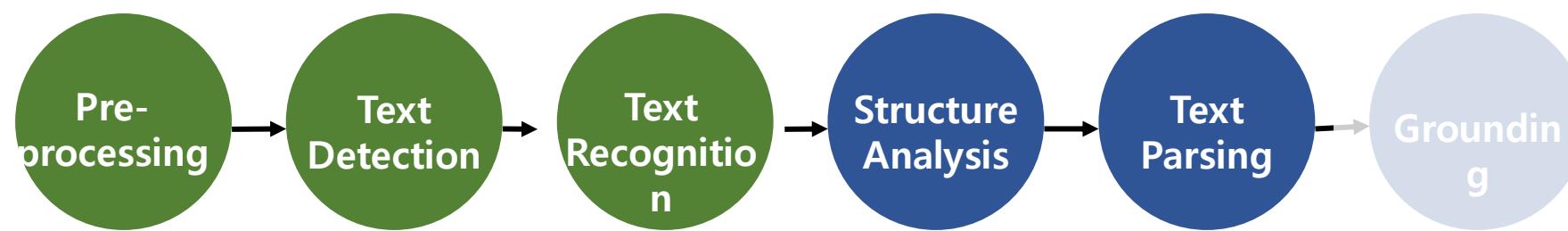
- 2D의 이미지 **layout 정보를 배제**하고 text를 1D sequence로 변환 → **Spatial Dependency Parsing**



e other conceptual examples

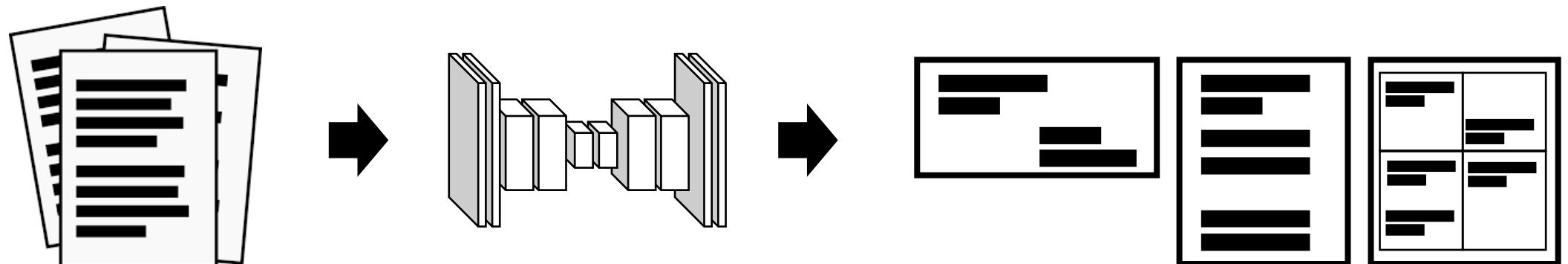


복잡한 Layout/Table 인식 문제



Text parsing의 한계 및 연구방향 (2/2)

- 언어 모델 학습을 위해 각 task마다 **많은 데이터가 필요** → **많은 양의 문서 데이터를 사전 학습**

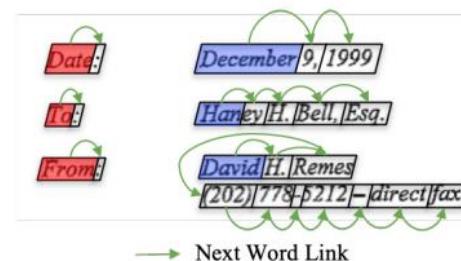


Downstream task

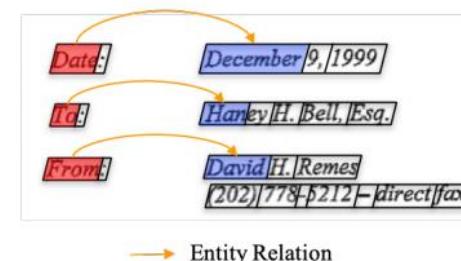
<i>Date</i>	<i>December/9/1999</i>
<i>To:</i>	<i>Harley H. Bell, Esq.</i>
<i>From:</i>	<i>David H. Remes</i> <i>(202) 778-5212 - direct/fax</i>

■ Question ■ Answer

(a) Starting token classification

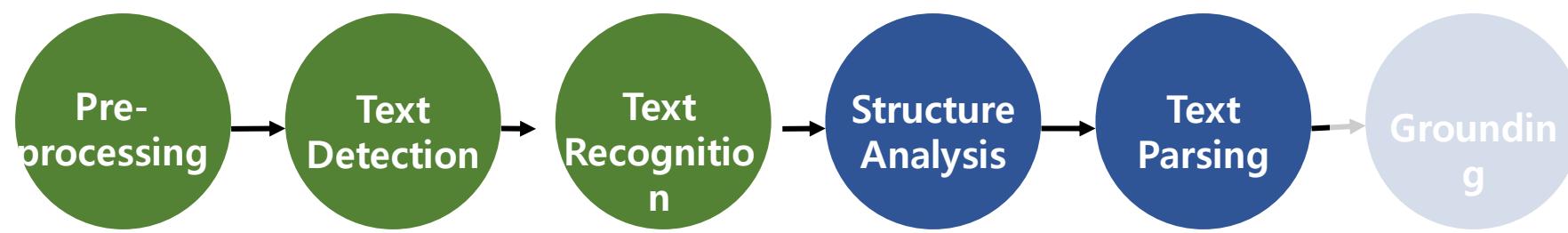


(b) Next token classification



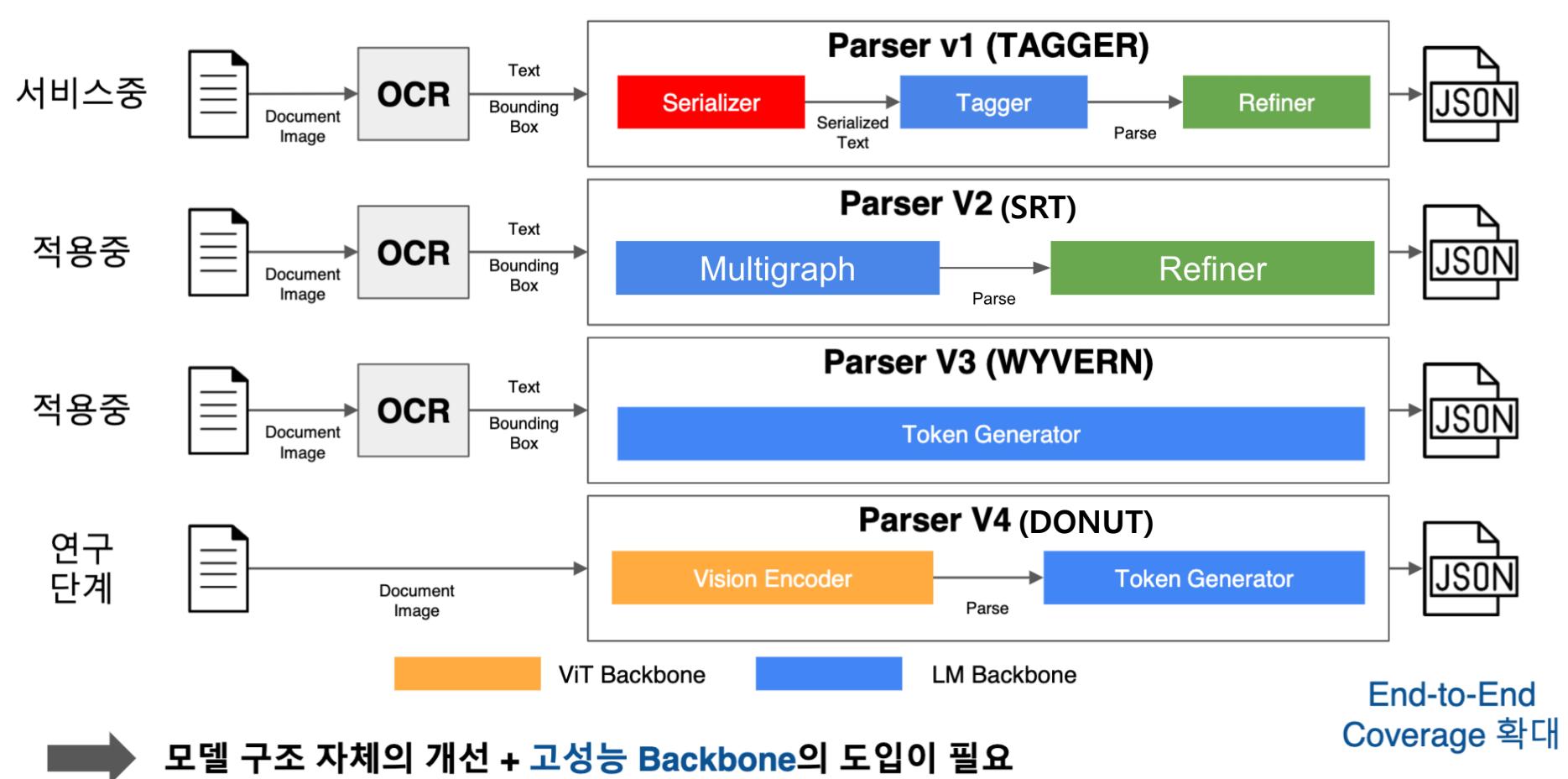
(c) Entity linking (EL) task

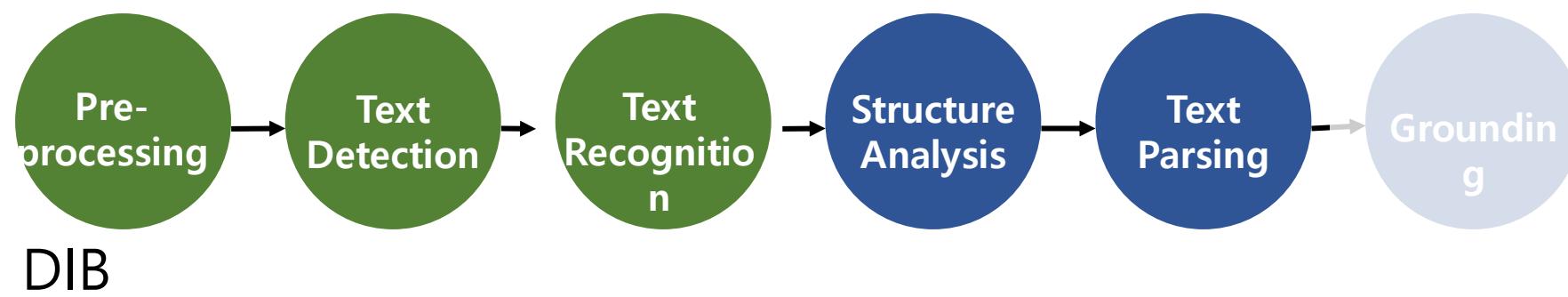
21.1Q : 명함, 영수증
21.3Q : Invoice 등 복잡한 문서



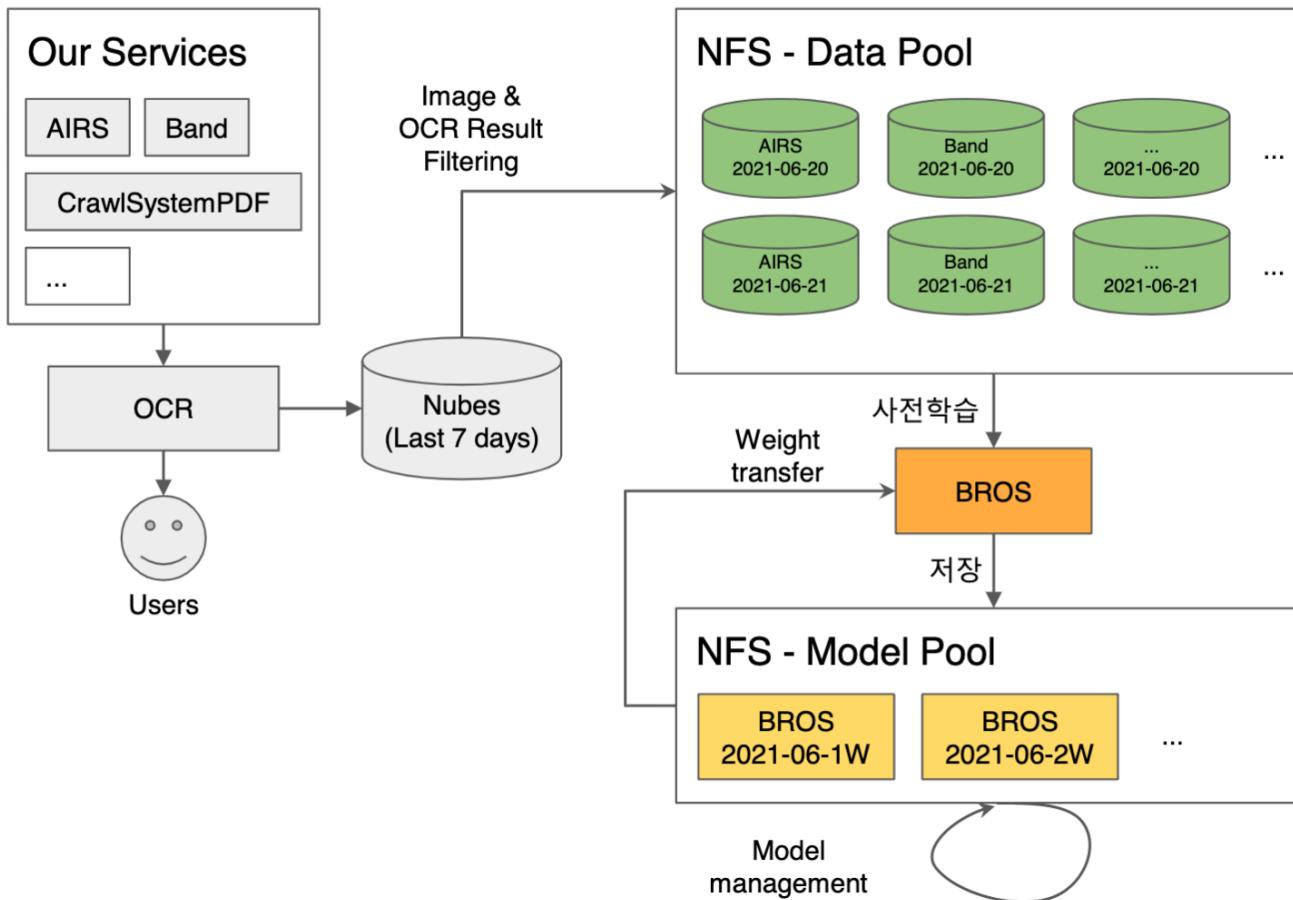
The new Text Parsing

Parser - Generation

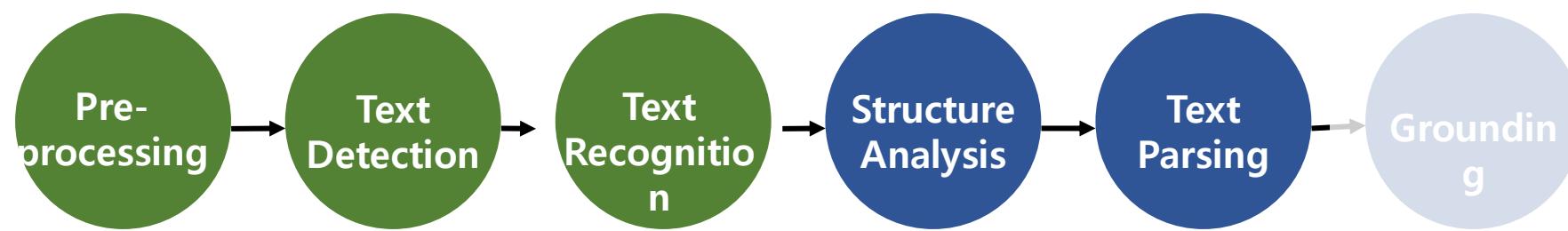




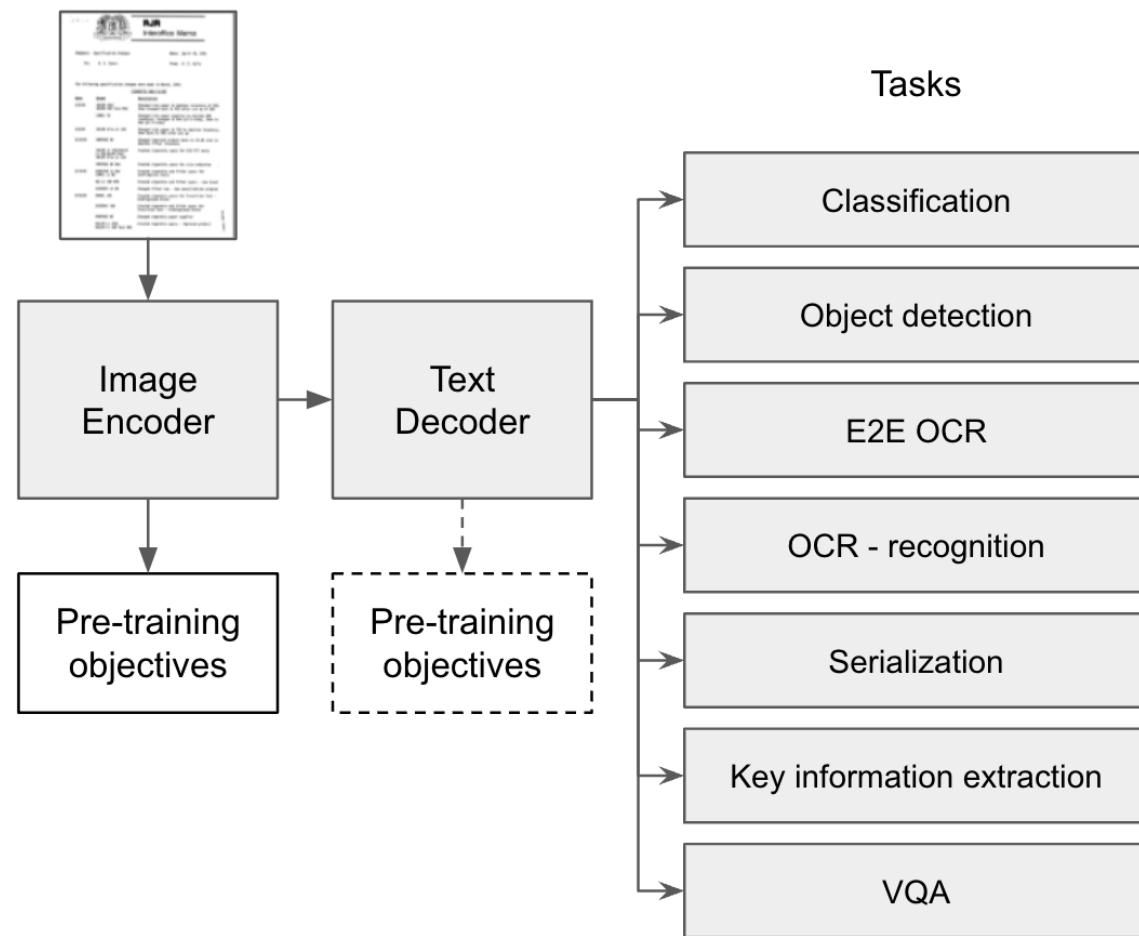
BROS Training Automation



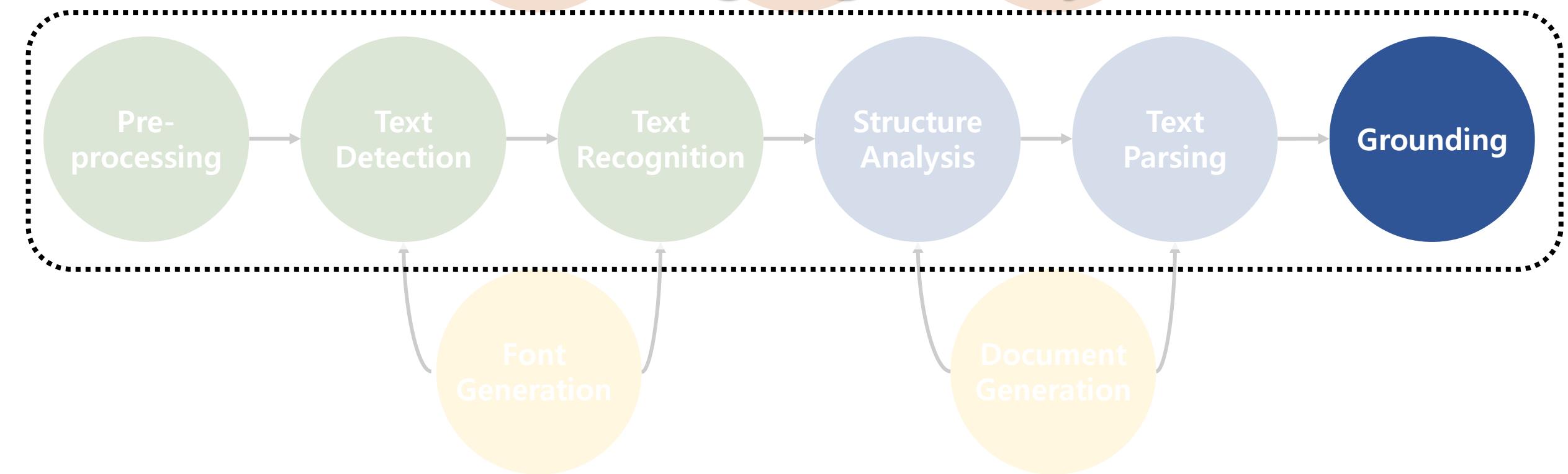
- Data Management
 - 사전학습용 General OCR 서비스 인입 데이터의 관리
 - Data Process
 - Sampling과 Filtering을 통한 Target Image 정제
 - Training Scheduler
 - BROS 사전 학습 주기 관리
 - Model Management
 - 학습된 BROS 모델의 저장 및 성능 관리
- 학습 자동화를 통한 고성능 Back-bone의 지속적인 생산

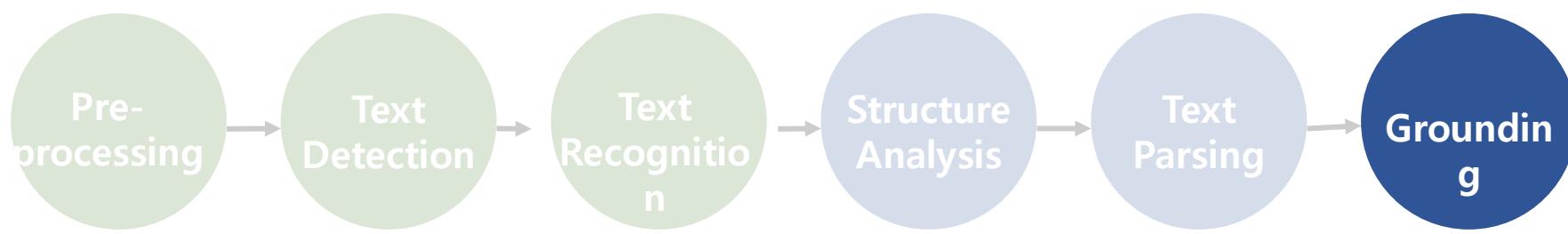


Double-U (W) Project



Visual language analysis





Grounding 이란?

- OCR을 통해 인식된 결과를 **사용자가 원하는 형태로 정규화** 하는 기술

주소 Grounding

- 신분증 인식 결과의 오류 보정
- zipcode 출력을 위한 주소 분할



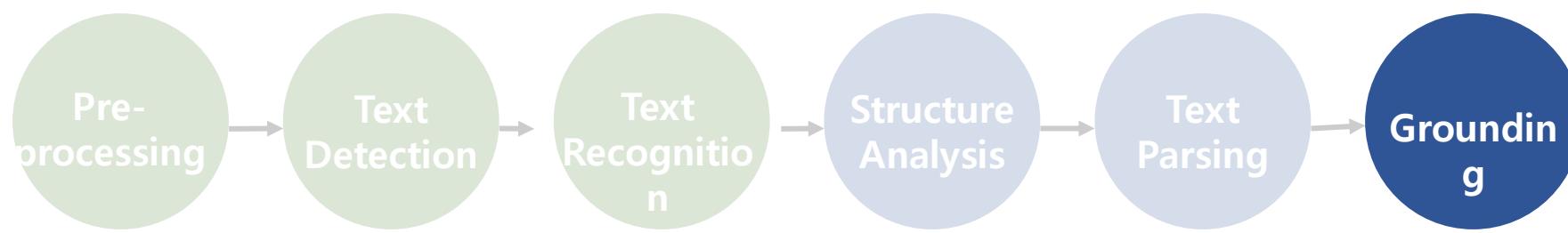
東京都于代田区霞が関2-1-2
(千 을 于로 오인식)

Region1 : 東京都
Region2 : 千代田区
Region3 : 霞が関
Region4 : 2-1-2
Post_Code : 1000013
Region1_Yomi : トウキョウト
Region2_Yomi : チヨダク
Region3_Yomi : カスミガセキ
(ツキノヒルヲノゾク)

메뉴 Grounding

- 메뉴명이 가게/영수증마다 상이
- 예) 아이스 아메리카노, 아메리,
아아, 아메리카S





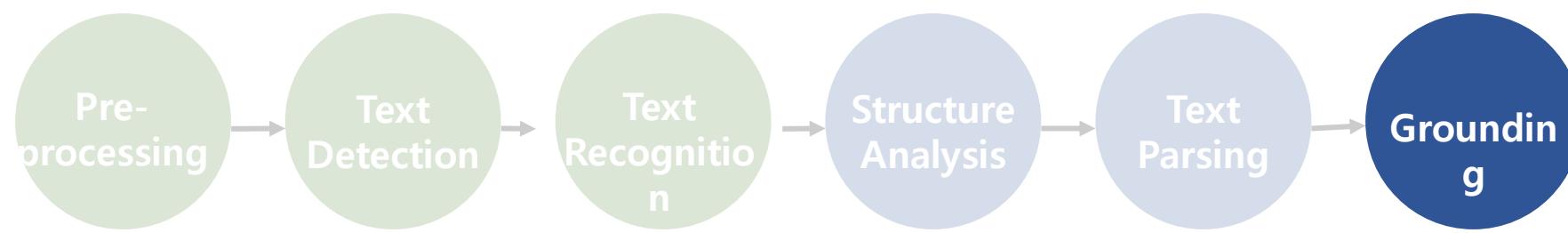
Grounding 활용 예제

- PoI(Point of Interest) Matching
 - 대상 필드 : 상호명, 전화번호, 주소
 - 목적 : (1) OCR 오인식 결과 보정, (2) 대상 자체 애매한 정보에 대한 보정



top_1_poi_entry

id	1539579425
tel	031-701-7233
address	경기도 성남시 분당구 분당로53번길 10 동호플라자 201호
jibun	경기도 성남시 분당구 서현동 268-6 동호플라자 201호
name	카쓰야 서현점
category	카쓰야
latitude	37.3835251
longitude	127.121781
kor_score	0.0020388525



Grounding 활용 예제

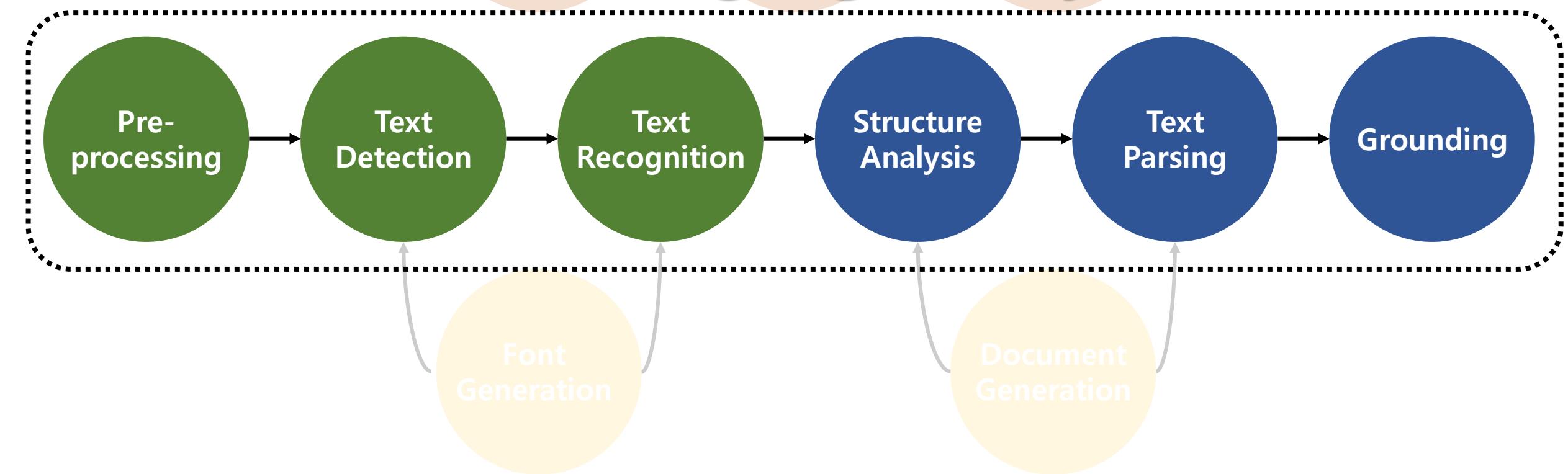
- PoI(Point of Interest) Matching
 - 대상 필드 : 상호명, 전화번호, 주소
 - 목적 : (1) OCR 오인식 결과 보정, (2) 대상 자체 애매한 정보에 대한 보정

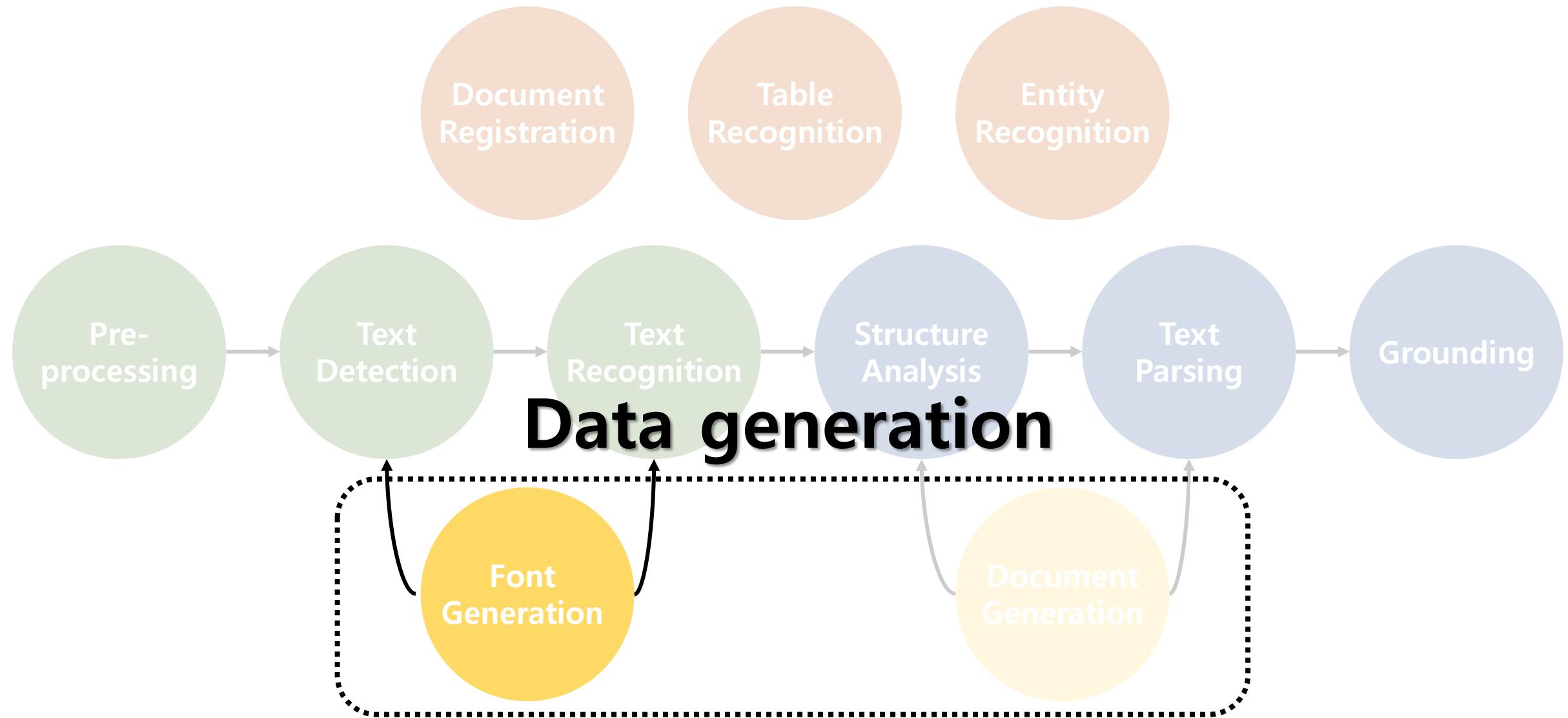


top_1_poi_entry

id	31383787
tel	031-726-0023
address	경기도 성남시 분당구 정자동 100
jibun	경기도 성남시 분당구 정자동 180
name	아티제 분당점
category	<ul style="list-style-type: none"> 음식점 카페, 디저트 카페 아티제
latitude	127.106200903863
longitude	37.3596259275659
kor_score	0.002057117

Visual language analysis





Font generation

손글씨 인식 성능 향상

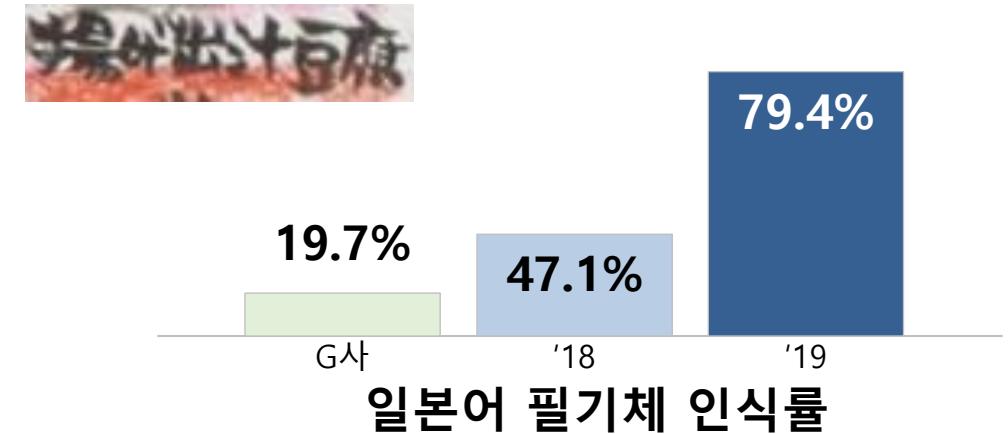
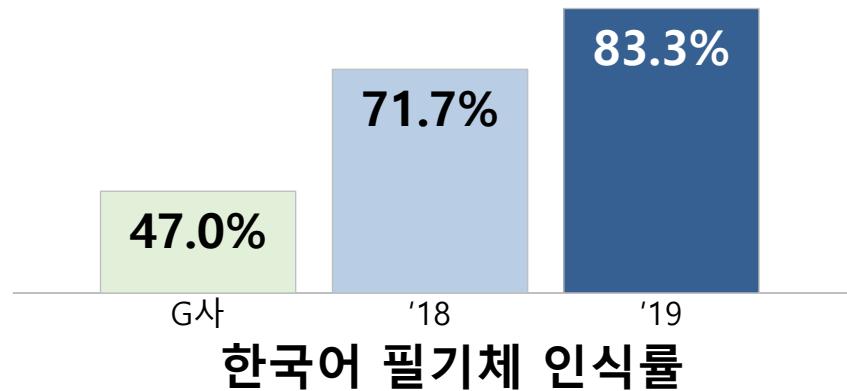
포함된 풀방법

이해력을 향상

잘고 있다니.

篆いていくうちに感動を吸収していきます

や感動を吸収していきます。



Font generation

손글씨 폰트 생성

2019 네이버 손글씨 공모전

- 256개의 손글씨 → 11,172 글자 생성



네이버 한글날 손글씨 공모전 - 앞으로 쓰여질 우리의 이야기를 응원합니다

▷ 694,949 | 등록 2019.10.07.

신고

tree****

2019.10.10. 20:02 · 신고

그동안 네이버가 한 일 중에서 가장 칭찬할 만한 프로젝트입니다. 이렇게만 하면 사랑받는 민족 기업됩니다^^~

제니

지금까지 네이버에서 기획한 이벤트 중에 가장 마음이 따뜻해지는 이벤트인 것 같아요! 내 손글씨를 폰트로 만들어주는 기능도 만들어지길 바라는 마음에서 설문조사도 참여했어요~^^

2019.10.09. 22:33:31 | 신고

Font generation

손글씨 폰트 생성

2019 네이버 손글씨 공모전

- 109개의 나눔 손글씨 공개



[글꼴 전체 보기](#)

[글꼴 소개글](#) | 기본 문구

나눔손글씨 가림연꽃

취업 준비입니다. 글씨에 성적이 담기다지만, 저는 이상을 담으려 해요. 불투명한 미래에도 희망이 있다는 믿음을 가지려 합니다.

나눔손글씨 갈맷글

저희 동네 예쁜 바닷길 이름이 길벗입니다. 그 이름에서 따왔습니다. 구불구불 편안한 길 같은 글로 쓰였으면 합니다.

글꼴명을 검색해보세요

나눔손글씨 강부장님체

오래전부터 글씨를 많이 쓰는 사무실에서 일해온 강 부장입니다. 사무실에서 흔히 볼 수 있는 이 글씨로 타이핑해보면 어떨까요?



유튜브 방송

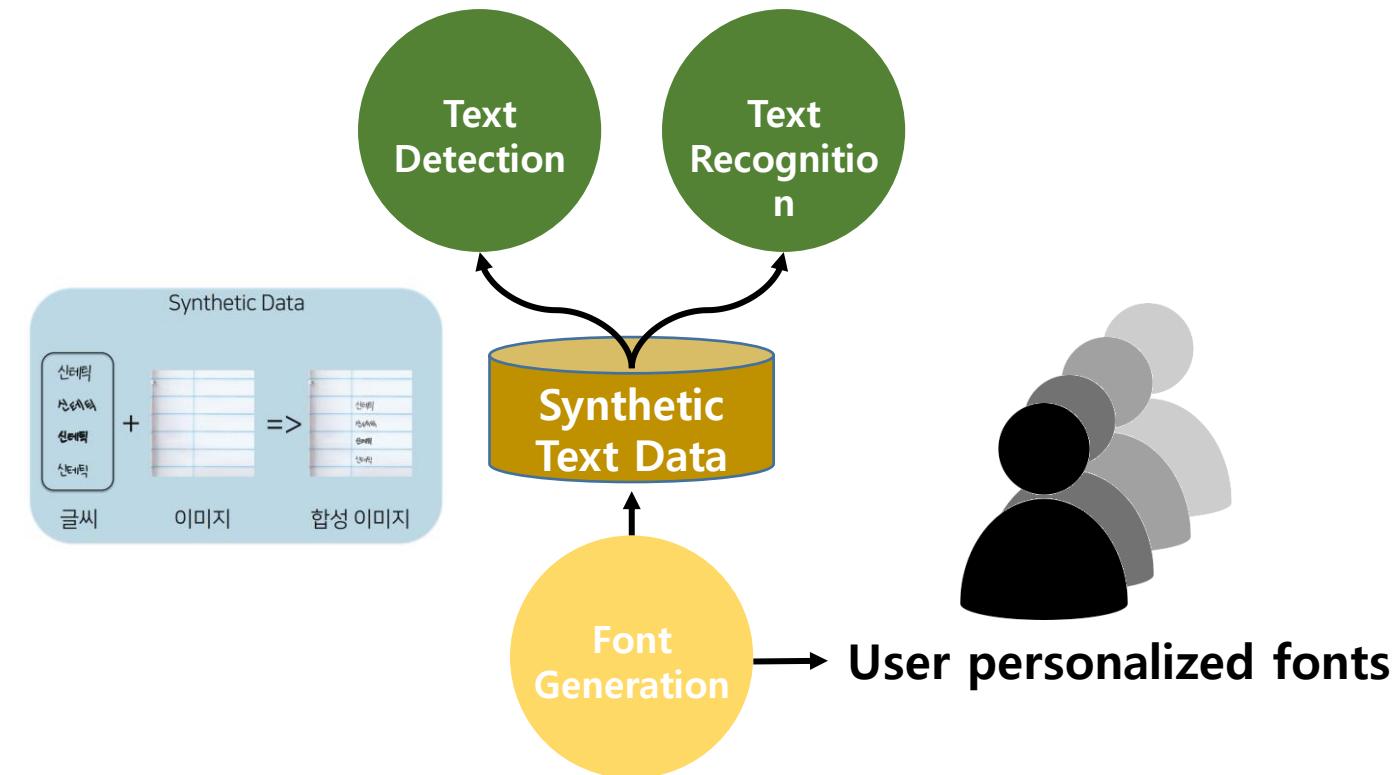
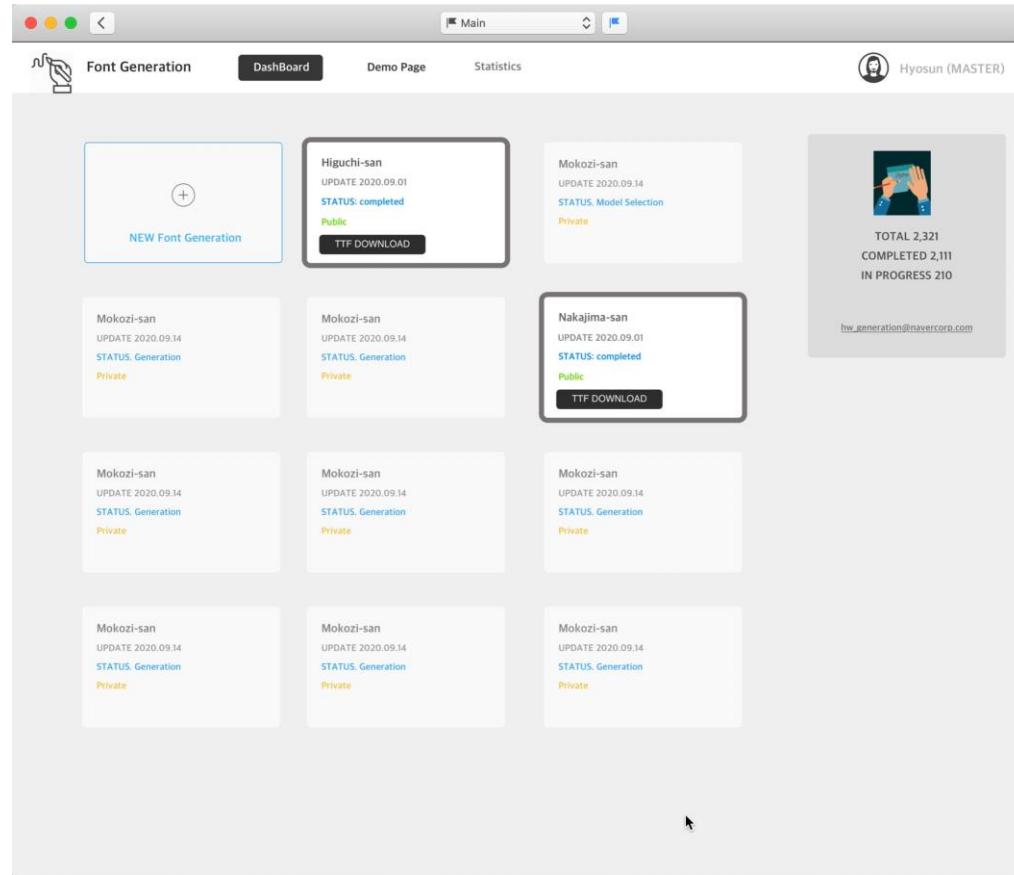


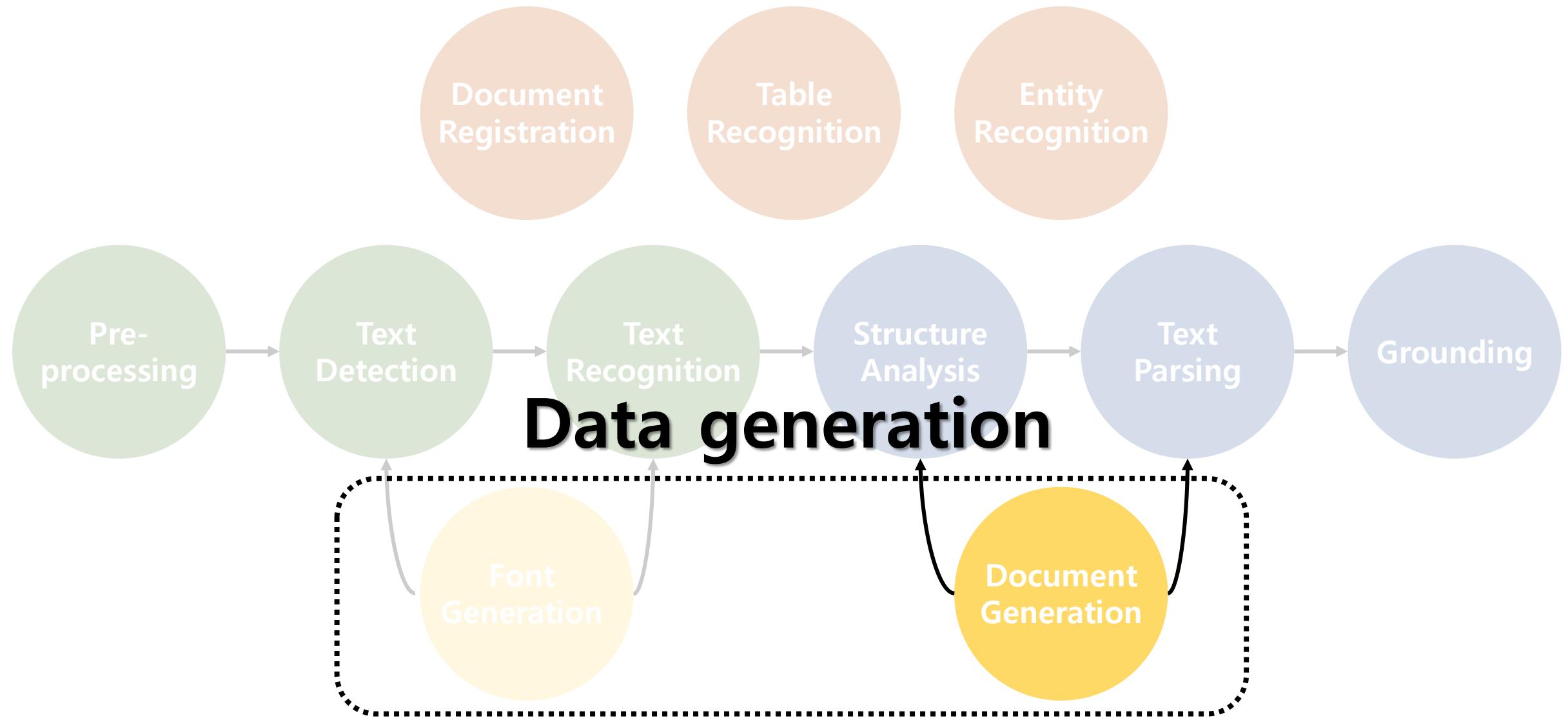
드라마 ‘스타트업’ 엔딩 크레딧

Font generation

폰트 생성의 Next step

개인화 폰트를 위한 admin tool 개발 중 (20.1Q)





Document generation

문서 이미지 생성의 목적

1. 수집할 수 없는 민감 정보 데이터 확보



신분증



신용카드

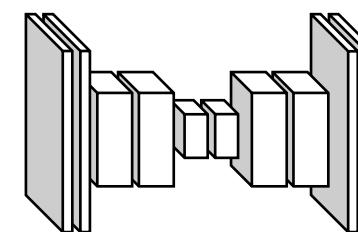
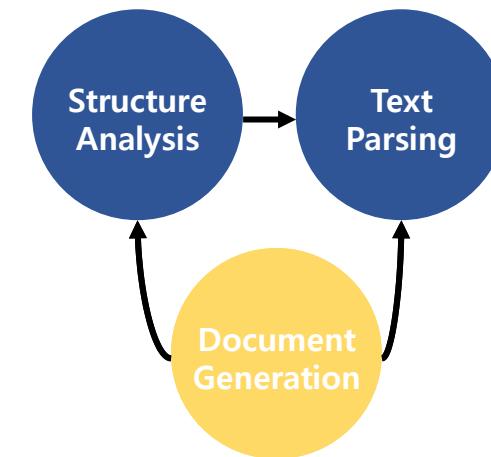


기밀문서

2. 다양한 언어/레이아웃을 학습하기 위한 데이터 확보



다양한 형태의 문서 이미지

사전 학습 문서 모델
(언어, 레이아웃)

Document generation

문서 생성 응용 분야 (CAPTCHA)

새로운 CAPTCHA를 제안

- 기존 CAPTCHA



글자 입력 방식의 CAPTCHA



General OCR 영수증 신용카드 사업자 등록증 고지서

한국어와 영어를 인식할 수 있는 기본 모델입니다.
왜곡이 있거나 복잡한 이미지에서도 정확하게 텍스트를 인식하며, 수기로 작성한 손글씨의 인식률도 뛰어납니다.



Text json

38
5PA s27

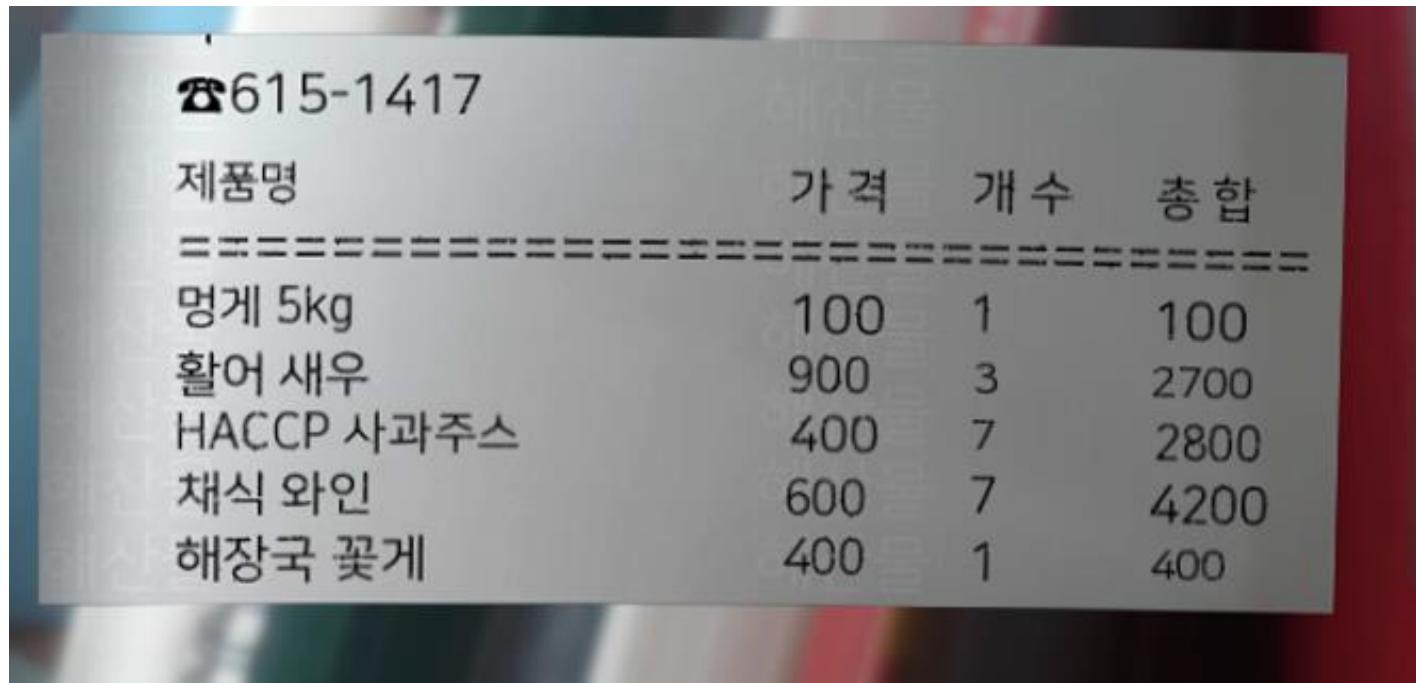
심지어 Clova OCR로는 쉽게 인식...



Document generation

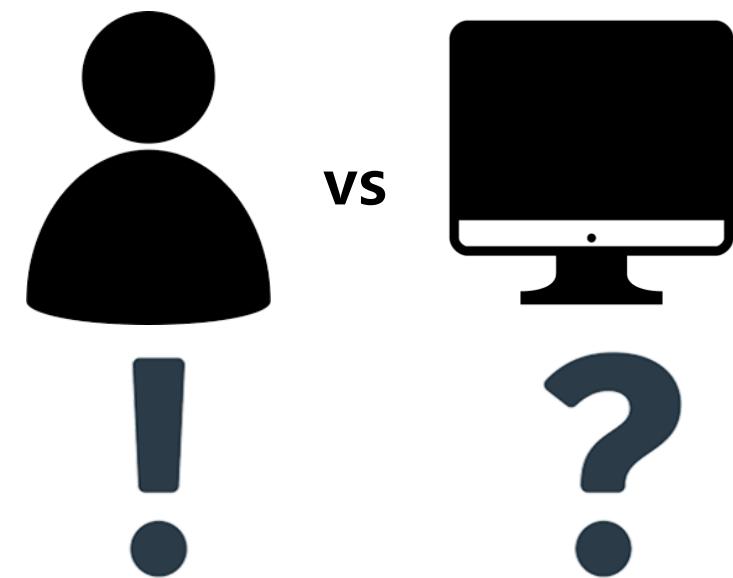
문서 생성 응용 분야 (CAPTCHA)

Semantic parsing이 필요한 영수증 인식을 문제로 내면...?



제품명	가격	개수	총합
멍게 5kg	100	1	100
활어 새우	900	3	2700
HACCP 사과주스	400	7	2800
채식 와인	600	7	4200
해장국 꽃게	400	1	400

Q1) 구입한 해산물 개수는 총 몇개인가?



Document generation

문서 이미지 생성 방법

CAPCHA용 영수증 생성 Process

제품명	가격	개수	총합
매운맛 고춧가루	800	2	1600
캔 콜라	100	1	100
과즙 망고	700	2	1400
세척 킹크랩	600	1	600
아이들간식 멸치	200	2	400

원활한 배송을 위해 배송 신청은 오후 2시까지 받고 있습니다.

빛나는 아울렛 탐라점

경산시 동대문구 월광동길 348

041-6328

제품명	가격	개수	총합
매운맛 고춧가루	800	2	1600
캔 콜라	100	1	100
과즙 망고	700	2	1400
세척 킹크랩	600	1	600
아이들간식 멸치	200	2	400

원활한 배송을 위해 배송 신청은 오후 2시까지 받고 있습니다.

빛나는 아울렛 탐라점

경산시 동대문구 월광동길 348

041-6328

제품명	가격	개수	총합
매운맛 고춧가루	800	2	1600
캔 콜라	100	1	100
과즙 망고	700	2	1400
세척 킹크랩	600	1	600
아이들간식 멸치	200	2	400

원활한 배송을 위해 배송 신청은 오후 2시까지 받고 있습니다.

빛나는 아울렛 탐라점

광동길 348

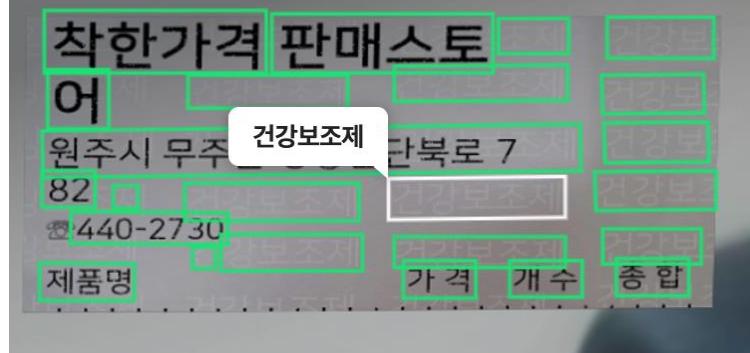
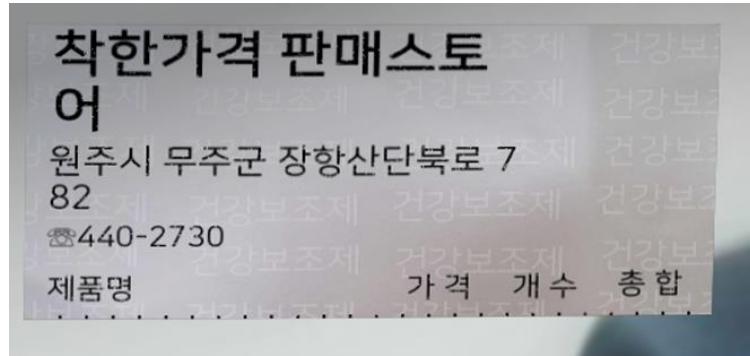
건강보조제

</div

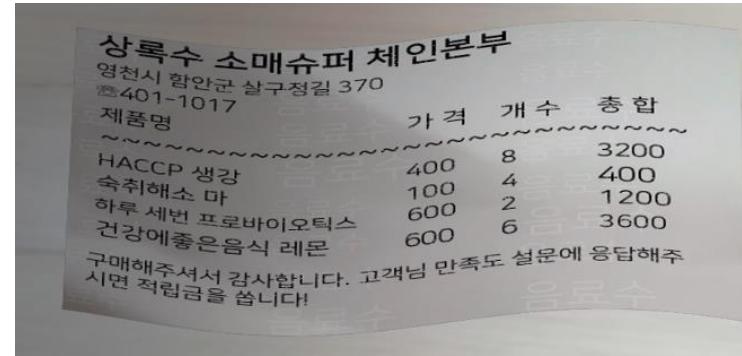
Document generation

문서 이미지 생성 방법

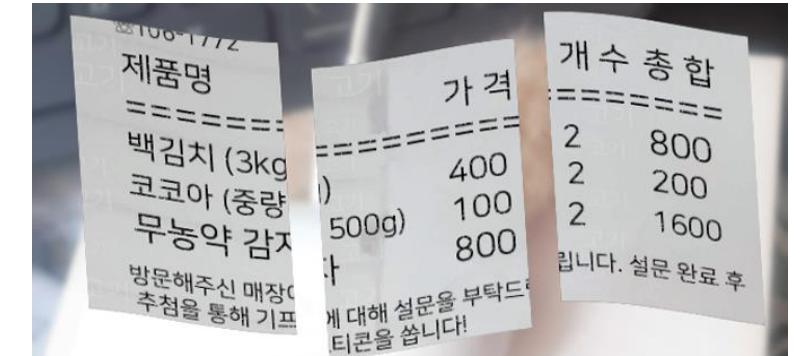
실제 영수증 이미지를 위한 노력 + 더 어려운 영수증을 위한 노력



Watermark



Perspective & Wave effect



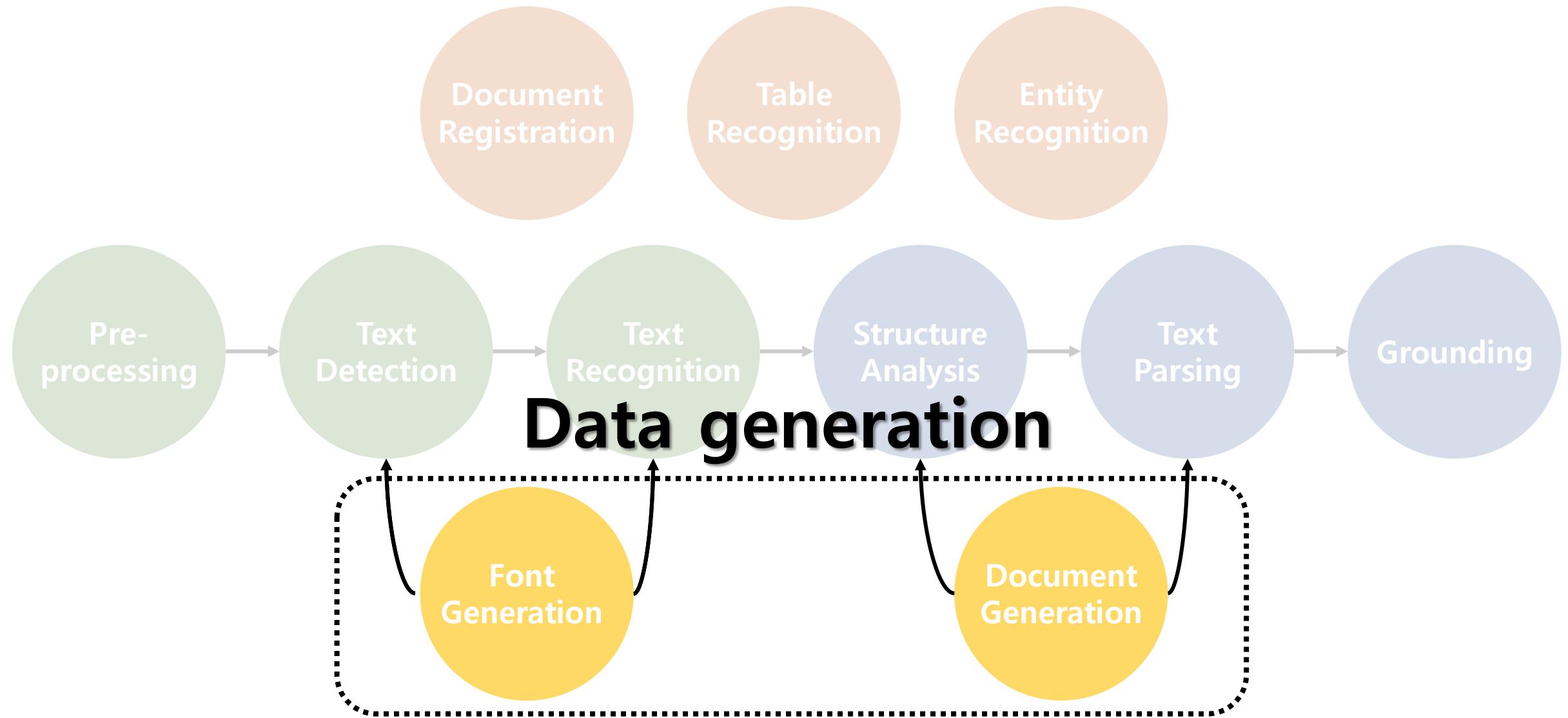
Split

Document generation

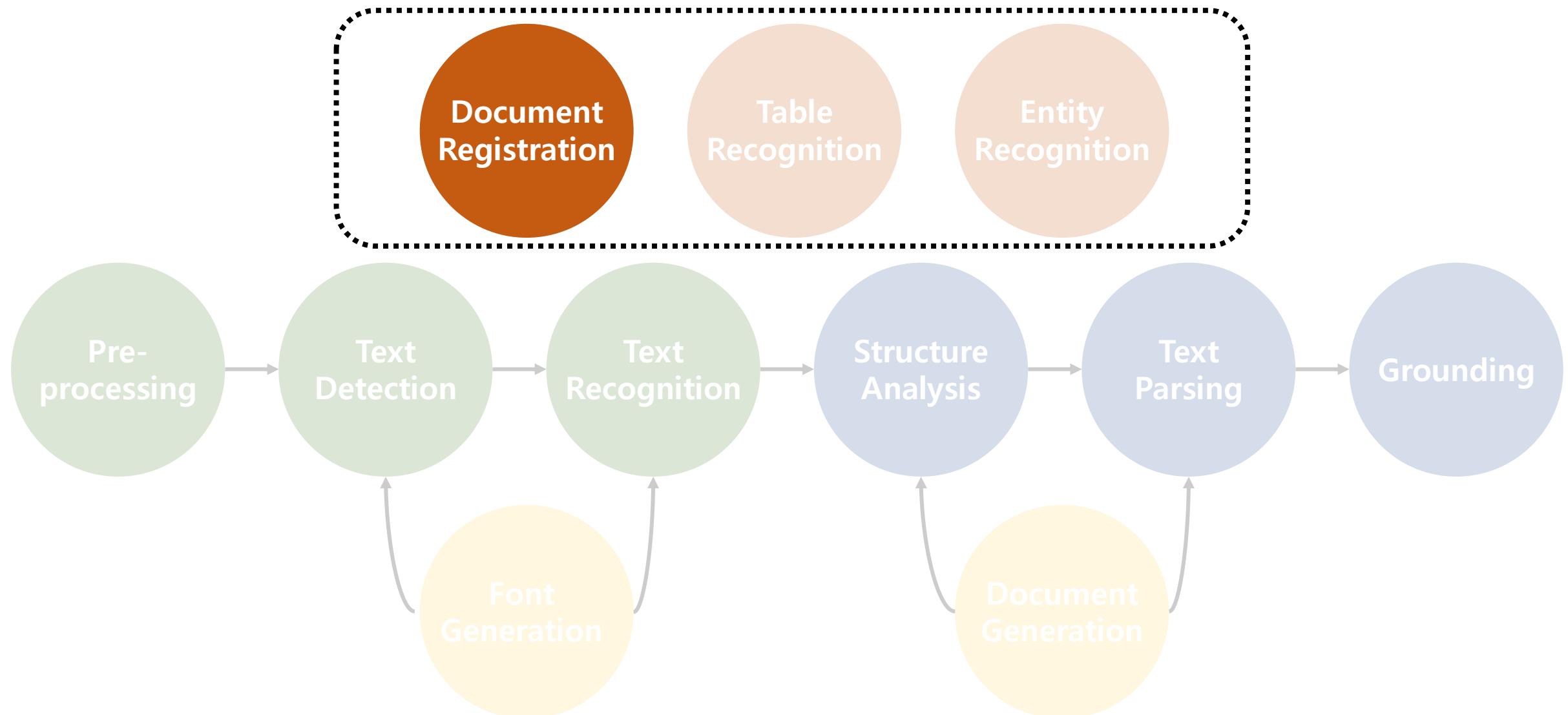
새로운 글자를 실제 글자처럼 보이게 만들 수 없을까?

다른 그림 찾기 퀴즈~





Form recognition



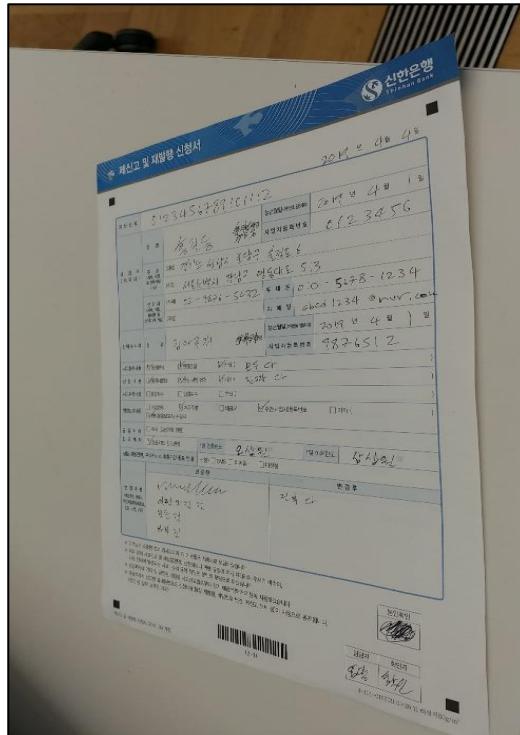
Document registration

NAVER LINE

Registration 목적

고정된 Template에 대해서 손쉽게 Key-value 추출

문서 Template /Annotation



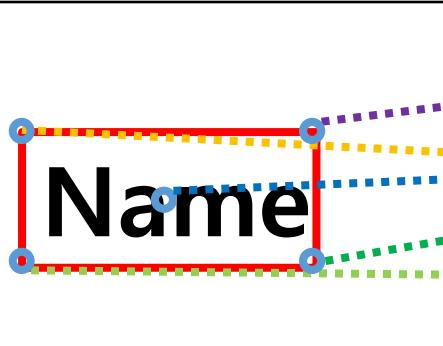
입력 이미지

Registration 완료된 결과 /Value 영역 위치

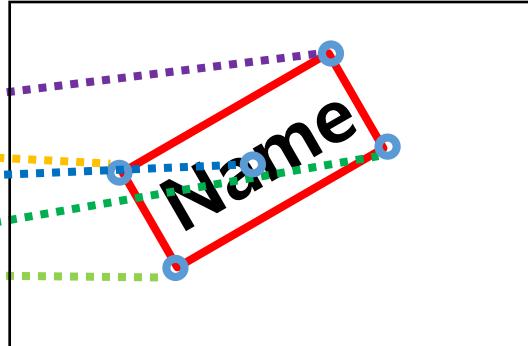
Document registration

Registration 방법

OCR text을 기반한 Image Keypoints를 대응시켜 변환 행렬 구함



Template image

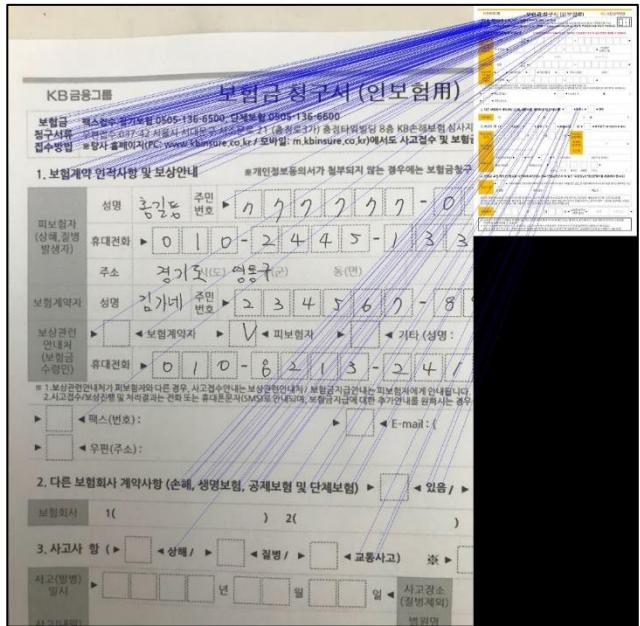


Input image

○ : Keypoints

-----: 대응점

문서 Template

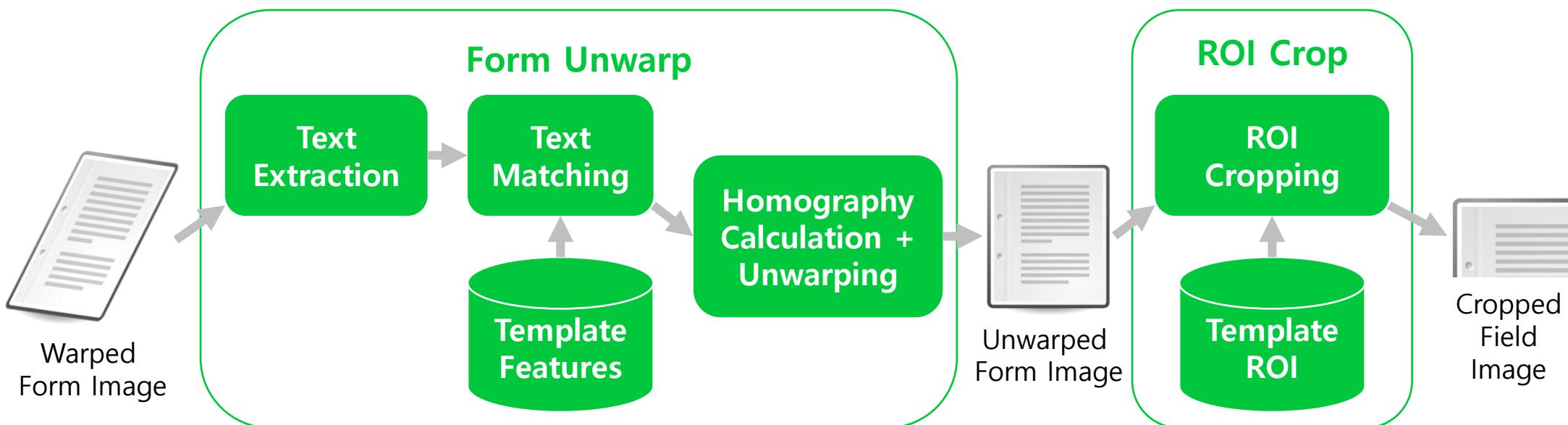


입력 이미지 및 대응점

Document registration

Registration 방법

Image keypoints via OCR n-gram



Document registration

Applications

Fixed-Form Recognition

NAVER CLOUD PLATFORM

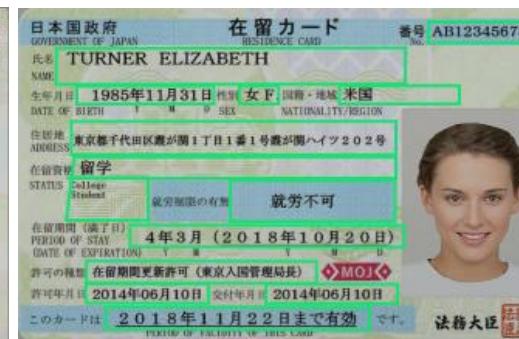
<https://www.ncloud.com/product/aiService/ocr>

- 사용자가 Template을 등록해서 고정형 문서 인식 서비스 제공
- 입력 이미지에서 원하는 key-value 추출



Document registration

Applications



총 5개 종류 인식 가능
(운전면허증, MyNumberCard, 재류자격증 등)

ID card recognition



OCR 대만어 확장과 함께 제공 (21.1Q)



견본
2014. 2. 1.
서울특별시 종로구청장

Key text Parsing(20.4Q)

Form recognition

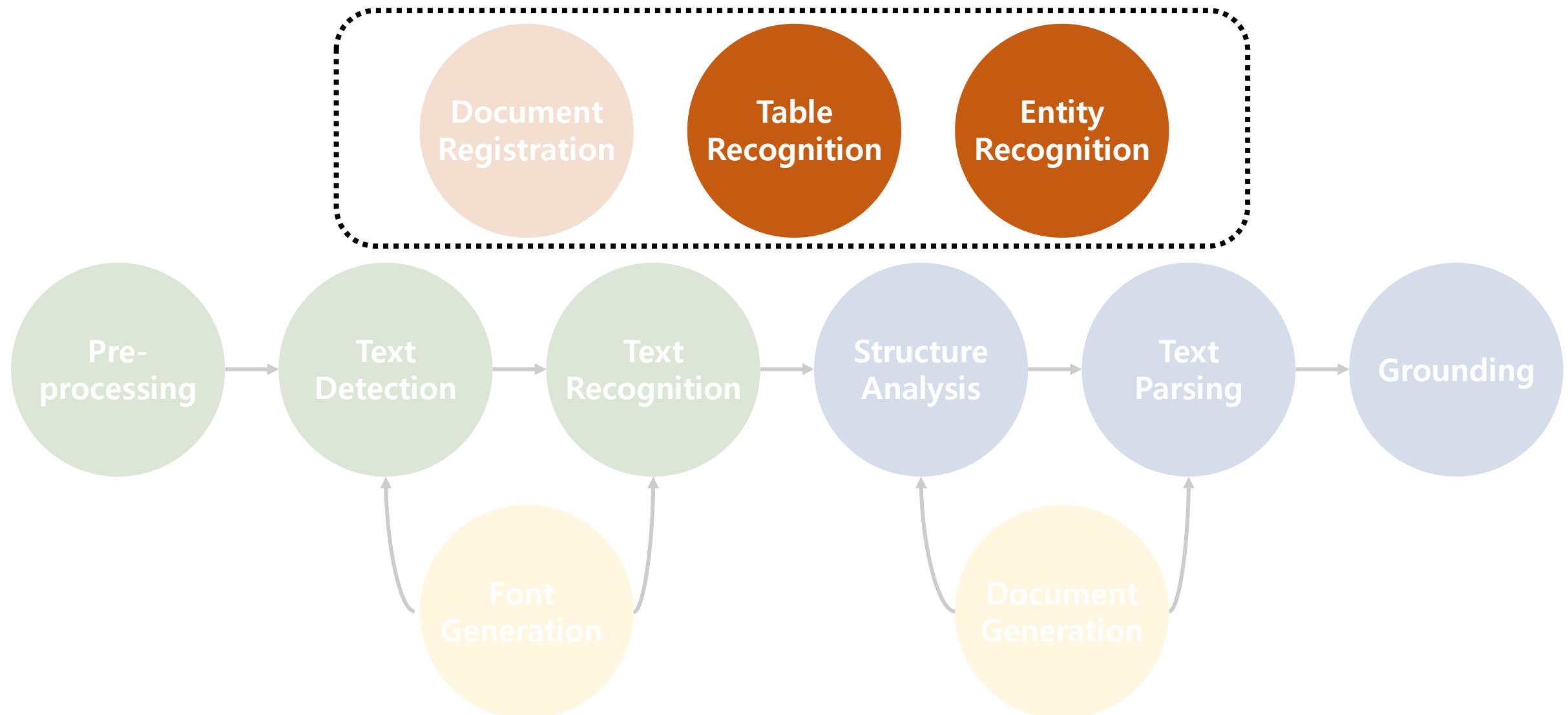


Table recognition

NAVER LINE

테이블 인식 요구사항

- 테이블은 문서내 구조화된 정보를 담고 있음
 - 다양한 문서에 테이블 포함 (의료비영수증, 세부내역서, 납품서, Invoice, BL 등등)

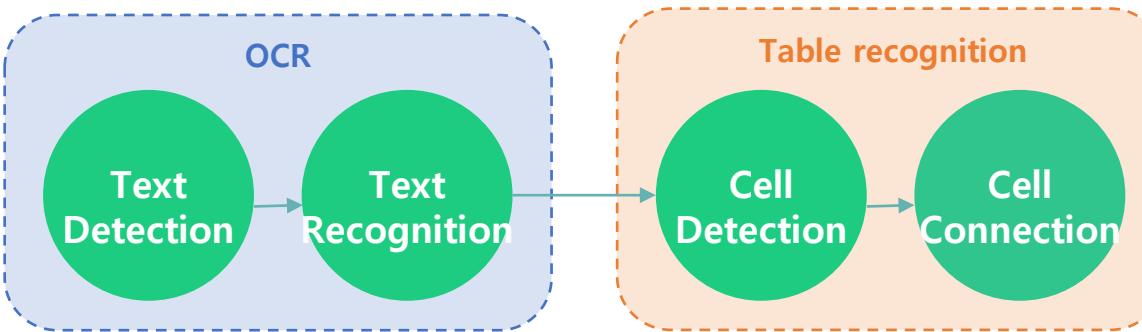
→ Key-value 추출하는 기법으로는 다양한 테이블의 구조화 정보를 표현 한계

Table recognition

NAVER LINE

테이블 인식 방법 (1/2)

개별Cell을 검출하고 연결하는 방법



입력 이미지

OCR 결과

Cell detection

Cell connection

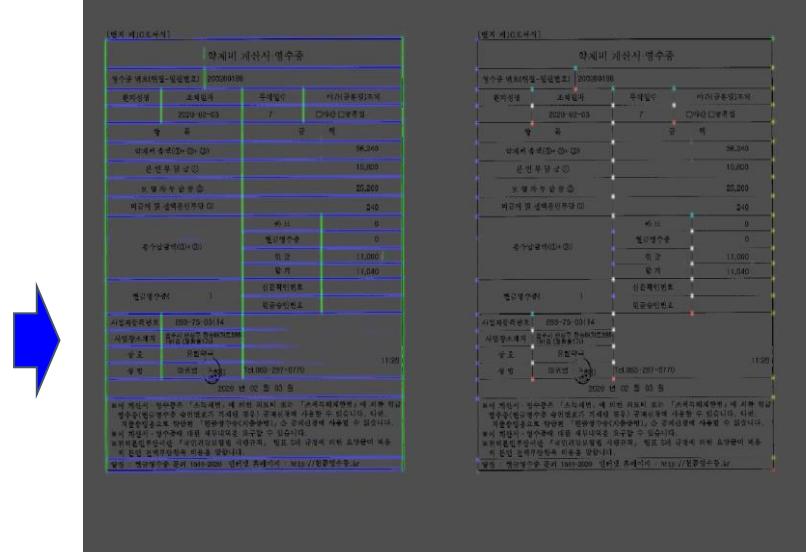
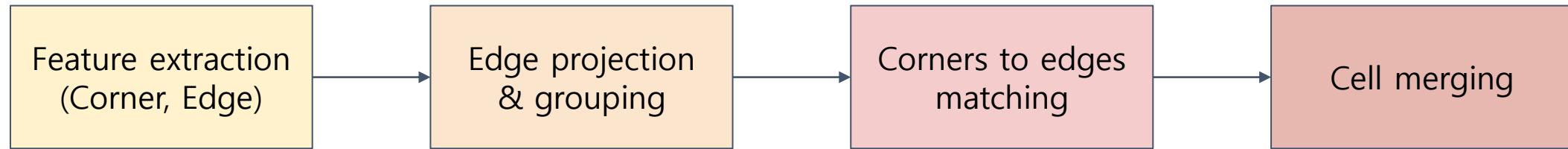
→ 다수의 데이터 필요, 범용화하기 어려움

Table recognition

NAVER LINE

테이블 인식 방법 (2/2)

테이블 구성 요소를 검출하여 형태를 재구성 하는 방법



Input image

Corner + Edge

부기 제 [10]서식			
약체기] 계산서-영수증			
영수증 번호(단점-일련번호)	200200188	부당인수	이전(공통임)조세
할인상점	조제일자	무당인수	□부과 □부과 면제
	2020-02-28	7	□부과 □부과 면제
할 류		금 액	
약세비 충액(①+②+③)		36,540	
본인부담금 ①		10,890	
보험자부담금 ②		25,290	
비급여 및 혼약판이부담 ③		240	
총수금금액(①+③)		카드 0 은행영수증 0 현금 11,030 할인 11,040 신용회원번호 입금승인번호	
현금영수증()			
사업자등록번호	893-75-00114		
사업장소개지	동화면 풍남리 정성학교길 256 전화번호(국번+전화번호)		
상 호	유한국교		141
성 명	이귀분 (姓)	tel.053-287-8770	
	2020년 02 월 03 일		
보이지 않는 영수증은 「소노세티」에 의뢰 되어 있는 「e-영수증제작방법」에 의뢰 하여 제작됩니다. 「e-영수증제작방법」은 「e-영수증제작방법」을 사용할 수 있습니다. (단, 단 계정이 있는 경우에만 가능합니다.) 「e-영수증제작방법」은 「e-영수증제작방법」을 제작, 제작된 영수증을 「e-영수증제작방법」을 통하여 할 수 있습니다. (단, 단계 로 인한 결제취약성이 있는 「e-영수증제작방법」 사용방법은, 필요 50% 규격이며 비록 의 분이 결제취약성을 이유로 제작을 달성이 어렵습니다.)			
인터넷 주소	http://www.senoseti.com	인터넷 주소	http://www.senoseti.com
인터넷 주소	http://www.senoseti.com	인터넷 주소	http://www.senoseti.com

Cell result

	A	B	C	D	E	F	G
1	A1:G1						
2	A2:B2		C2:G2				
3	A3:A3	B3:C3		D3:D3	E3:G3		
4	A4:A4	B4:C4		D4:D4	E4:G4		
5	A5:C5			D5:G5			
6	A6:C6			D6:G6			
7	A7:C7			D7:G7			
8	A8:C8			D8:G8			
9	A9:C9			D9:G9			
10				D10:E10		F10:G10	
11							
12				D11:E12		F11:G12	
13				D13:E13		F13:G13	
14	A10:C14			D14:E14		F14:G14	
15				D15:E15		F15:G15	
16	A15:C16			D16:E16		F16:G16	
17	A17:A17	B17:C17		D17:G17			
18	A18:A18	B18:C18		D18:G18			
19	A19:A19	B19:C19		D19:G19			
20	A20:A20	B20:C20		D20:G20			
21	A21:G21						
22	A22:G22						
23	A23:G23						
24							

xlsx export

Table recognition

테이블 인식 방법 비교

	방식/결과형태	특징	현황
Cell detection + connection	<ul style="list-style-type: none"> Cell 검출하고 연결하는 형태 Key-value 형태로 결과 제공 	<ul style="list-style-type: none"> 특정 요구에 맞춘 결과 제공 학습에 많은 양의 데이터가 필요 	<ul style="list-style-type: none"> 흥국생명 진료비영수증 대응 방식 (20.4Q)
Table reconstruction	<ul style="list-style-type: none"> 모서리,선을 이용해서 테이블 재구성 xlsx 파일로 테이블 결과 제공 	<ul style="list-style-type: none"> 범용성 확보 사용자가 테이블로 시스템 연동 필요 	<ul style="list-style-type: none"> NCP 테이블 인식 솔루션 제공 (21.1Q)

Entity recognition

문서 이해에 필요한 다양한 요소

KB금융그룹 보험금 청구서 (인보험用)

보험금 페스션수: 강기보험 050-136-6500, 단체보험 050-136-6500
청구서 우편번호: 037 42 서울시 서초구 서초로 71 (송정동 37) 송정호텔딩 8층 KB손해보험 실무지원센터 (우편번호: 08130)
인터넷 홈페이지(PC: www.kbsure.co.kr / 모바일: m.kbsure.co.kr)에서 시고행수 및 보험금청구서를 접수가 가능합니다.

1. 보험계약 인식사항 및 보상안내

설명 주민 번호: 123456-7891011
휴대전화: 010-1234-56789 직장명/ 주부
주소: 이천 성동 주안 개나리 아파트 502동 408호
보험계약자 성명: 장도연 주민 번호: 123456-7891011
보험관련 안내서: 0 보험금 수령일
휴대전화: 010-1234-56789
체크(번호): E-mail: ()
우편주소: 서울 강남구 삼성동 현대아이파크 아파트 432동 609호

2. 다른 보험회사 계약사항 (손해, 생명보험, 공제보험 및 단체보험) □ 있음 / □ 없음

보험회사: 1() 2() 3()

3. 사고사 항 (▶ 상해 / ▶ 0() 질병 / ▶ 교통사고) ▶ 추가첨부 (추가첨부는 제출되는 경우에만 가능)
사고(발생) 일시: 2019년 2월 16일
사고(내용) 경위: 차량 충돌
사고장소: 차량 충돌 장소
교통사고: 차량 충돌
본인차량번호: 차량번호 위치: 운전석 조수석 뒷좌석 보상금 기타:
보상금 수령 계좌 (보험금을 타인에게 위임하는 경우 보험금청구서 외 별도 '위임장'과 인감증명서를 제출해야 합니다)
은행명: 국민은행 계좌번호: 123-456-789 예금주: 한 해진
보증금: 1000000 원
작성일자: 2019년 2월 16일 (날짜) 보험금청구서 작성자: 이지연 (피보험자)
※ 소유자(고의사고, 하이킹, 힐링, 친환경, 미세먼지, 사고후 보상거액 등)는 평의하여, 일반에 의해 10년 미만의 징역이나 2년만의
이하의 형금에 처해질 수 있습니다.
※ 반드시 피보험자가 서명하거나 미성년자 금지자와의 경우 그 친권자는 또는 법정 후견인이 서명하시기 바랍니다.
미성년자의 경우 친권자가 서명하여, 다른 일정한 회사에 전화하지 않는 한 본인부모 일정부모로 공조명으로 통의 및 서명할 수 있습니다.



칸에 작성 안내를 위해 형광펜등으로 둘그리미 표시를 하시면 청구서 오인식의 원인이 될 수 있습니다.

주민번호	1	2	3	4	5	6	7	8	9	10	11
의료급여 수급권자	False	True	□ 예	□ 아니오	단체 보험 가입여부	multi_2	True	False	□ 예	□ 아니오	
주민번호	1	2	3	4	5	6	7	8	9	10	11
연락처	0	10	15	20	25	30	35	40	45	50	55
계약자	True	False	False	False	False	False	False	False	False	False	False
E-MAIL / 주소 상세 기재 (메일/우편 요청시)											

- **MultiBox 인식** : 글자칸이 주어지고 그 안에 적힌 글자 인식
- **CheckBox 인식** : Check 여부 인식

NOTE TO ALL APPLICANTS: Please completely fill out this application or fail to submit required documents indicated in instructions, USCIS may deny your application.

art 14. Interpreter's Contact Information, Certification, and Signature

Provide the following information about the interpreter.

Interpreter's Full Name

Interpreter's Family Name (Last Name) Interpreter's Given Name (First Name)
Interpreter's Business or Organization Name (if any)

Interpreter's Mailing Address

Street No. ✓ Ste. #
City or Town State/Prov.
Province

art 14. Interpreter's Contact Information, Certification, and Signature

Provide the following information about the interpreter.

Interpreter's Full Name

Interpreter's Family Name (Last Name) Interpreter's Given Name (First Name)
Interpreter's Business or Organization Name (if any)

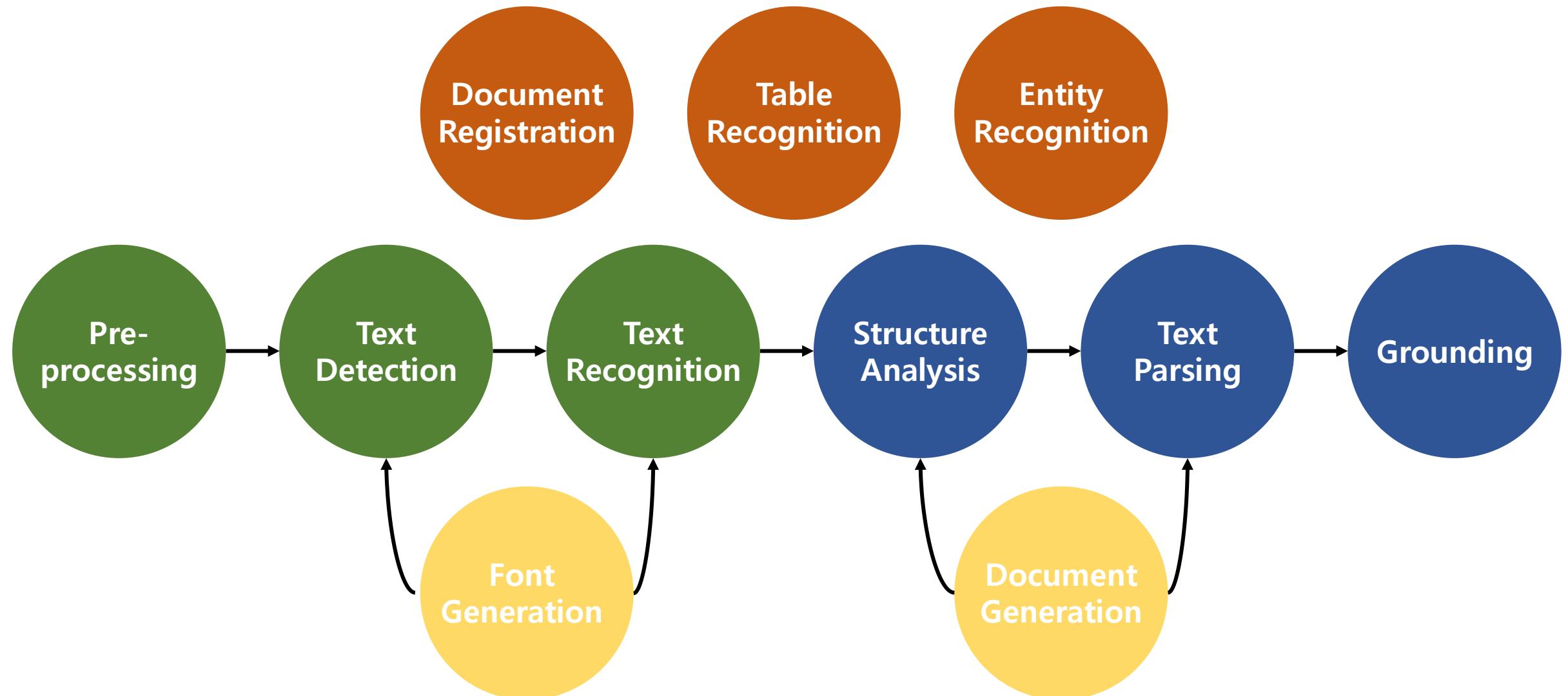
Interpreter's Mailing Address

Street No. ✓ Ste. #
City or Town State/Prov.
Province

- **Text attribute 인식** : 활자체/필기체 글자 속성, 도장 인식

인식 대상 문서 예제

Document Understanding



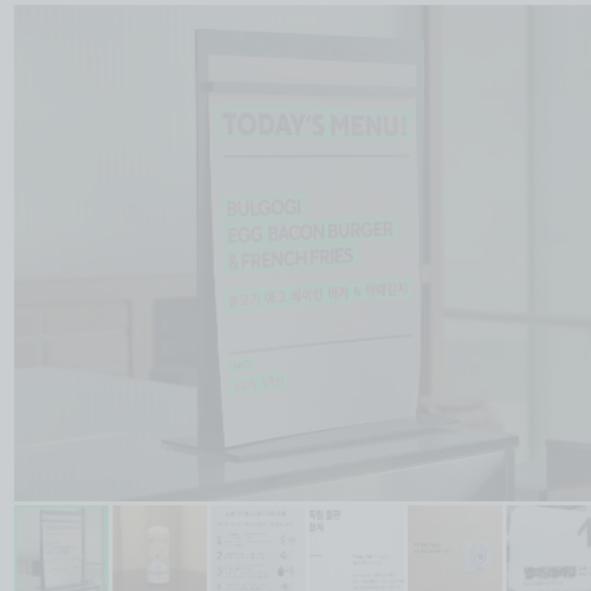
Thank you for listening!

CLOVA OCR

다양한 언어 지원은 물론, 문서별로 최적화된 모델을 제공하여 정확한 결과를 지원합니다.

지금 바로 체험해보세요

Web demo available at
→<https://clova.ai/ocr>



Text json

TODAY'S MENU!
BULGOGI
EGG BACON BURGER
&FRENCH FRIES
불고기 에그 베이컨 버거 & 막대감자
INFO.
소고기: 호주산

