

문서인식 서비스 개발

유아이패스코리아
산학협력프로젝트 8팀

목차

Contents

1. 팀원 소개
2. 프로젝트 분석 및 설계내용
3. 진행 상황
4. 이슈사항 및 해결방안
5. 향후 일정



팀원 소개



이재영

컴퓨터학부

Azure 개발
외부 솔루션 연동
데모 소프트웨어 제작



명노아

컴퓨터학부

Azure 개발
데모 소프트웨어 제작
세부 기능 구현



김기태

컴퓨터학부

패키지파일 생성
Azure 개발
외부 솔루션 연동



박경모

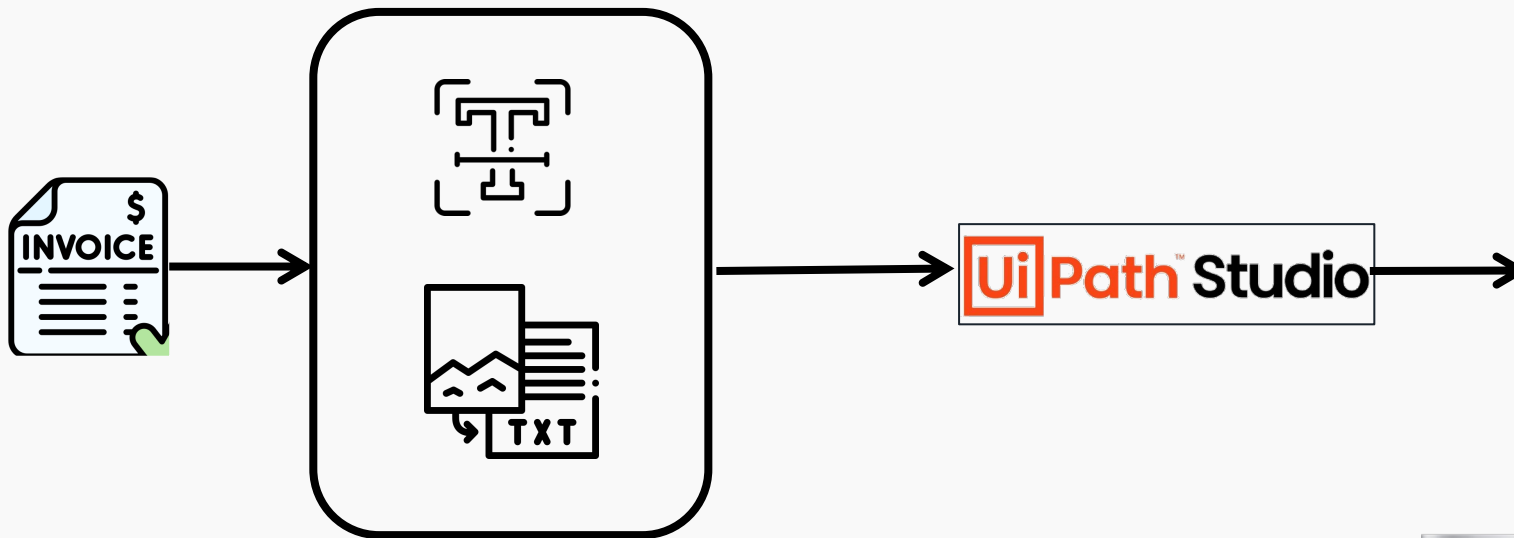
컴퓨터학부

데이터필드 수집
외부 솔루션 연동
데모 소프트웨어 제작

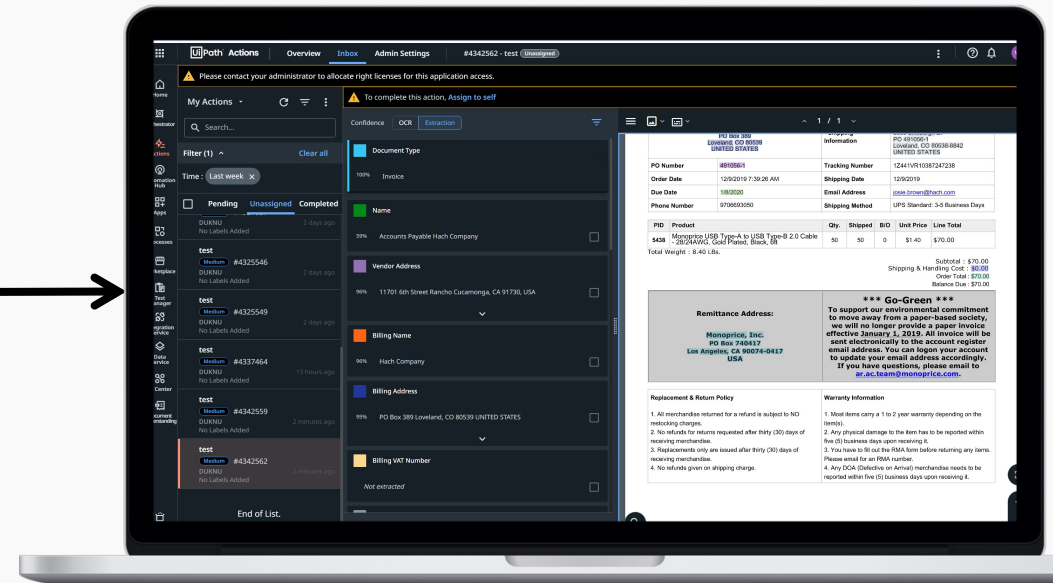


프로젝트 분석

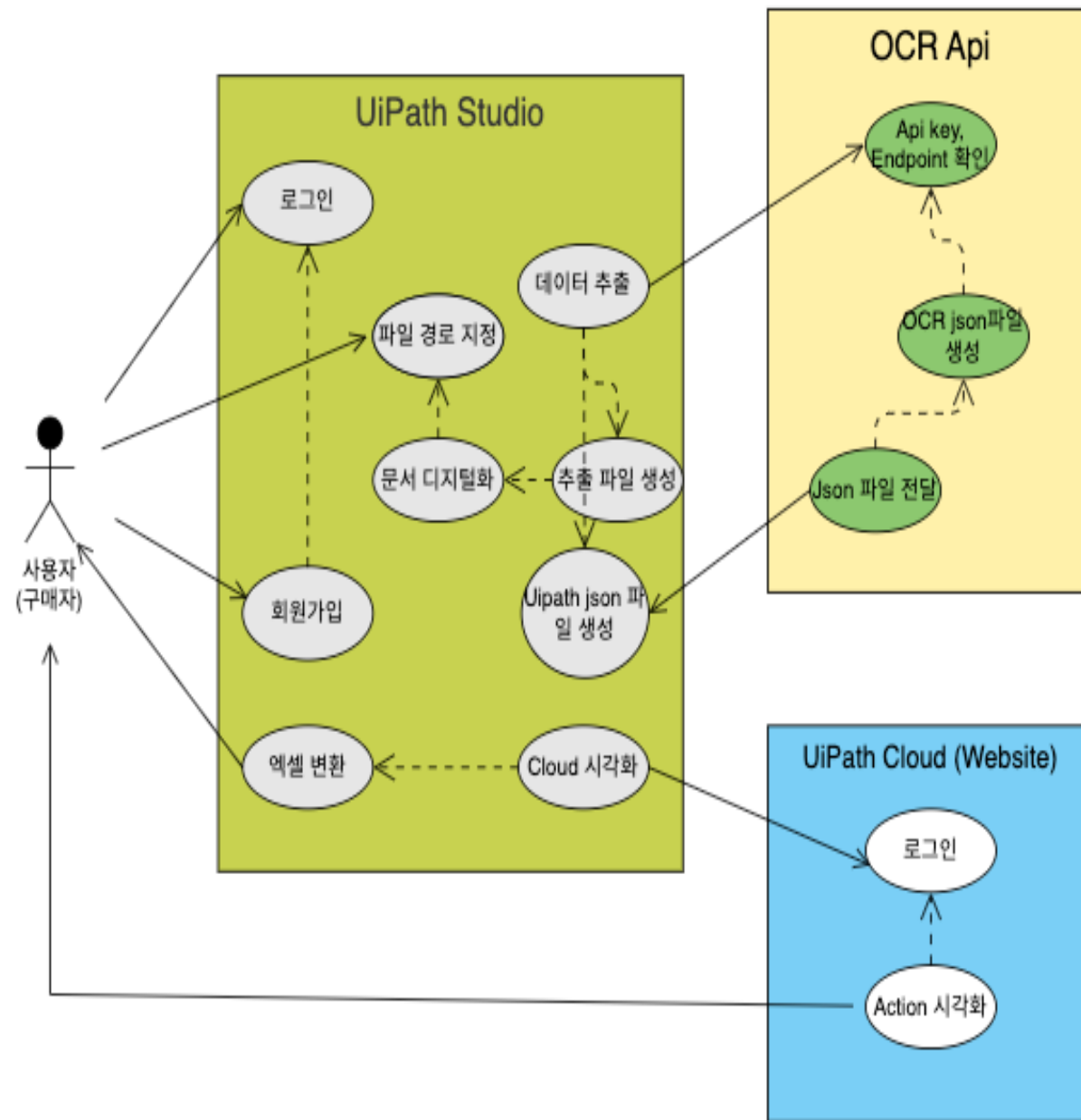
기존의 OCR API를 이용 -> UiPath Studio에 맞게 형식 변환 ->
정보 분류 및 시각화 처리 (일부 미구현)



Microsoft Azure, Naver CLOVA



프로젝트 설계 Use-Case





To-Do List

Azure OCR API가 제공하는 Field명과 Uipath의 Field를 매칭

```

"IndexInText": 37,
"Language": "eng",
"Length": 17,
"Rotation": 0,
"SkewAngle": -0.1,
"Type": 1,
"WordGroups": [
  {
    "IndexInText": 37,
    "Length": 17,
    "Type": 4,
    "Words": [
      {
        "Box": [ 77.0, 1992.0, 272.0, 58.0 ],
        "Polygon": [ 1992.0, 77.0, 2264.0, 77.0, 2264.0, 135.0, 1992.0, 135.0 ],
        "IndexInText": 37,
        "OcrConfidence": 1.0,
        "Text": "INVOICE",
        "VisuallineNumber": 0,
        "TextType": 1
      }
    ]
  }
],

```

- IndexInText
- Type
- Text
- Texttype
- Polygon

```
{
  "FieldId": "Semi-StructuredDocuments.Financial.Invoice.BillingName",
  "FieldName": "Billing Name",
  "FieldType": 0,
  "IsMissing": false,
  "DataSource": 0,
  "Values": [
    {
      "Components": [],
      "Value": "Hach Company",
      "UnformattedValue": "Hach Company",
      "Reference": {
        "TextStartIndex": 454,
        "TextLength": 12,
        "Tokens": [
          {
            "TextStartIndex": 454,
```

- FieldType
- Values
- Value
- Boxes

→
"Attribute 변환"

To-Do List

읽어들인 값의 위치를 구분하여
출력 (Json 시각화)

Vendor Address

96% 11701 6th Street Rancho Cucamonga, CA 91730, USA

11701 6th Street
Rancho Cucamonga, CA 91730, USA

Address Line 1
11701 6th Street

Address Line 2

Address Line 3

City
Rancho Cucamonga

State / County / Province
California

Country
United States

Zip Postal Code
91730

Billing Name

96% Hach Company

Billing Address

INVOICE

Monoprice, Inc.
11701 6th Street
Rancho Cucamonga, CA 91730, USA
www.monoprice.com
TEL: 877-271-2592 ; FAX: 909-989-0078

INVOICE
Invoice Number : 19583668
Online Store: https://www.monoprice.com

Invoice Number 19583668

Accounts Payable
Hach Company
PO Box 389
Loveland, CO 80539
UNITED STATES

Shipping Information

Receiving Department
Hach Company
5600 Lindbergh Dr
PO 491056-1
Loveland, CO 80538-8842
UNITED STATES

PO Number 491056-1

Tracking Number 1Z441VR10387247238

Order Date 12/9/2019 7:39:26 AM

Shipping Date 12/9/2019

Due Date 1/8/2020

Email Address josie.brown@hach.com

Phone Number 9706693050

Shipping Method UPS Standard: 3-5 Business Days

PID	Product	Qty.	Shipped	B/O	Unit Price	Line Total
5438	Monoprice USB Type-A to USB Type-B 2.0 Cable - 28/24AWG, Gold Plated, Black, 6ft	50	50	0	\$1.40	\$70.00

Total Weight : 8.40 LBs.

Subtotal : \$70.00

Shipping & Handling Cost : \$0.00

Order Total : \$70.00

Balance Due : \$70.00

Remittance Address:
Monoprice, Inc.
PO Box 740417
Los Angeles, CA 90074-0417
USA

*** Go-Green ***
To support our environmental commitment to move away from a paper-based society, we will no longer provide a paper invoice effective January 1, 2019. All invoice will be sent electronically to the account register email address. You can logon your account to update your email address accordingly. If you have questions, please email to ar.ac.team@monoprice.com.

LINE NO.	PRODUCT AND DESCRIPTION	QUANTITY ORDERED	QUANTITY B.O.	QTY. SHIPPED	QTY U/M	UNIT PRICE	PRICING UNIT	AMOUNT (NET)
*****NO WOG SHIPMENTS*****								
SHIP ALL ORDERS COLLECT:								
UPS ACCT#8A900V								
FREIGHT BILLED ON AN INVOICE WILL NOT BE PAID.								
IBF OR OBF.								
C OF C MUST BE PACKAGED IN THE BOX. DO NOT ATTACH TO THE								
PACK SLIP.								
ALL LOCTITE AND DOW CORNING MUST SHIP WITH A C OF C.								
PLEASE PACKAGE IN A SEPARATE BOX.								
1	STTC-838-33709 TIP CHISEL 1.5MM 30 DEG 800F FOR MX SYSTEMS Country of Origin: UNITED STATES Customer Prod: STTC-838	20	0	20	EA	24.26	EA	485.20
2	STTC-126-33709 TIP BENT SHARP .4MM 30 DEG 700F FOR MX SYSTEMS Country of Origin: UNITED STATES Customer Prod: STTC-126	10	0	10	EA	23.68	EA	236.80
3	STTC-125-33709 TIP CHISEL 1MM 30 DEG 700F FOR MX SYSTEMS Country of Origin: UNITED STATES Customer Prod: STTC-125	5	0	5	EA	24.97	EA	124.85
4	SFP-CNB05-33709 TIP CONICAL BENT .5MM STANDARD FOR MFR-HSR Country of Origin: UNITED STATES Customer Prod: 6901000 MISC 916533	2	0	2	EA	15.20	EA	30.40
5	STTC-138-33709 TIP CHISEL 1.5MM 30 DEG 700F FOR MX SYSTEMS Country of Origin: UNITED STATES	5	0	5	EA	23.12	EA	115.60
6	SMTc-160-33709 METCAL 700 SMTc RWK CART LONG BLADE 10MM Country of Origin: UNITED STATES	5	0	5	EA	85.95	EA	429.75
7	SMTc-162-33709 TIP BLADE LONG 22.1MM 700F FOR MX SYSTEMS Country of Origin: UNITED STATES	3	0	3	EA	86.61	EA	259.83
8	SMTc-164-33709 BLADE TIP CARTRIDGE 5MM Country of Origin: UNITED STATES	4	0	4	EA	89.68	EA	358.72
9	TT-95-50391 SOLDER TIP TINNER .70 OZ LEAD FREE RESIDUE FREE	5	0	5	EA	9.90	EA	49.50

(O)
(X)

To-Do List

단일 개체로서의 값이 아닌 행
단위로 구분 가능한 알고리즘
작성

*현재 포함, 비포함 알고리즘 사용

진행상황

- 개발환경 세팅 및 코드 분석 완료
- OCR Field와 Uipath Studio에서 사용하는 Field Table을 생성하는 InvoiceAzure.cs 개발 진행 중

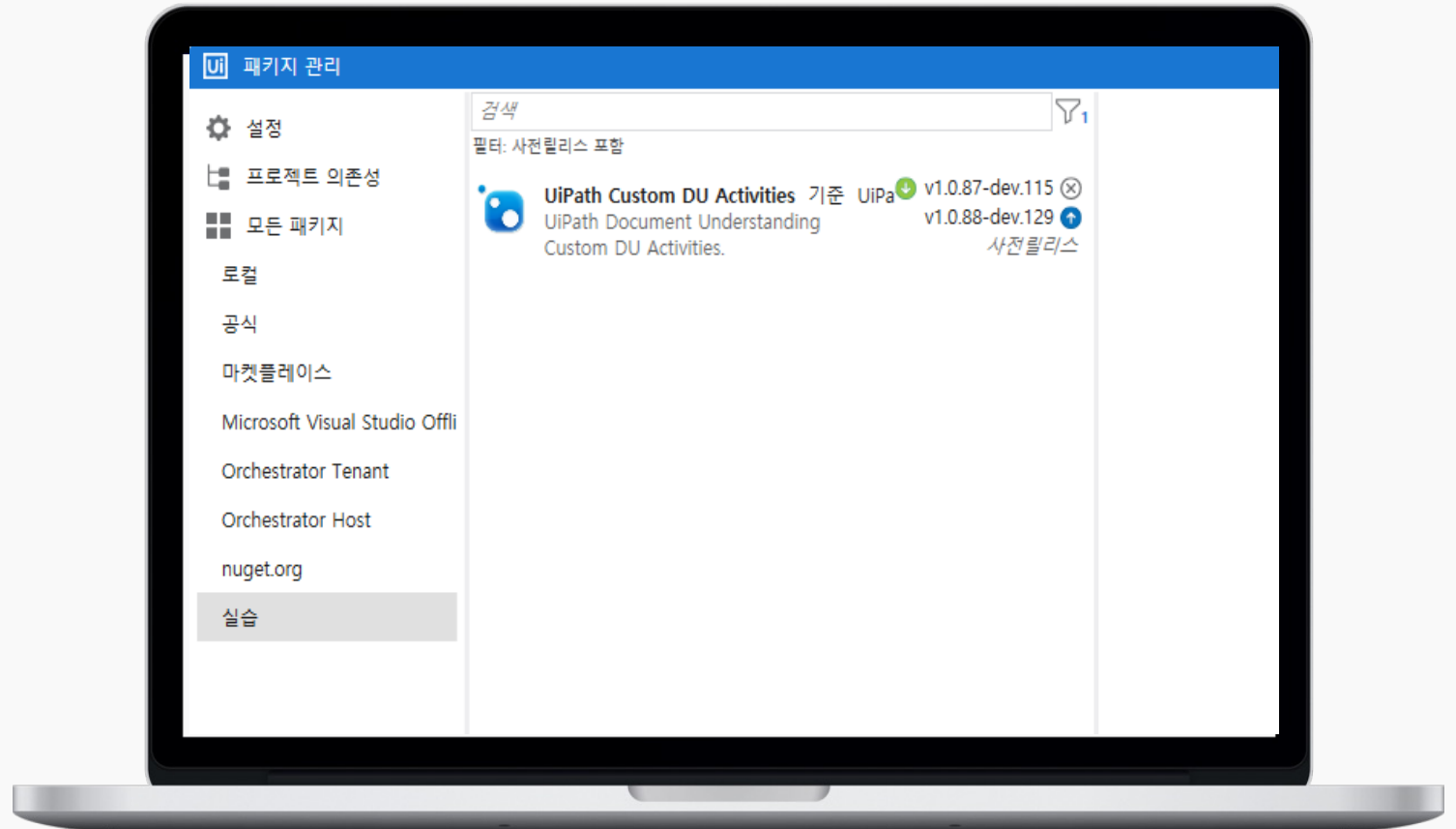


이슈사항 및 해결방안

패키지 파일 불러오기(해결완료)

Visual Studio에서 빌드한 패키지 파일이 Uipath
Studio에서 갖고오지 못하는 현상

Uipath.UIAutomation.Activities 패키지 설치 후
경로를 재등록하여 문제 해결

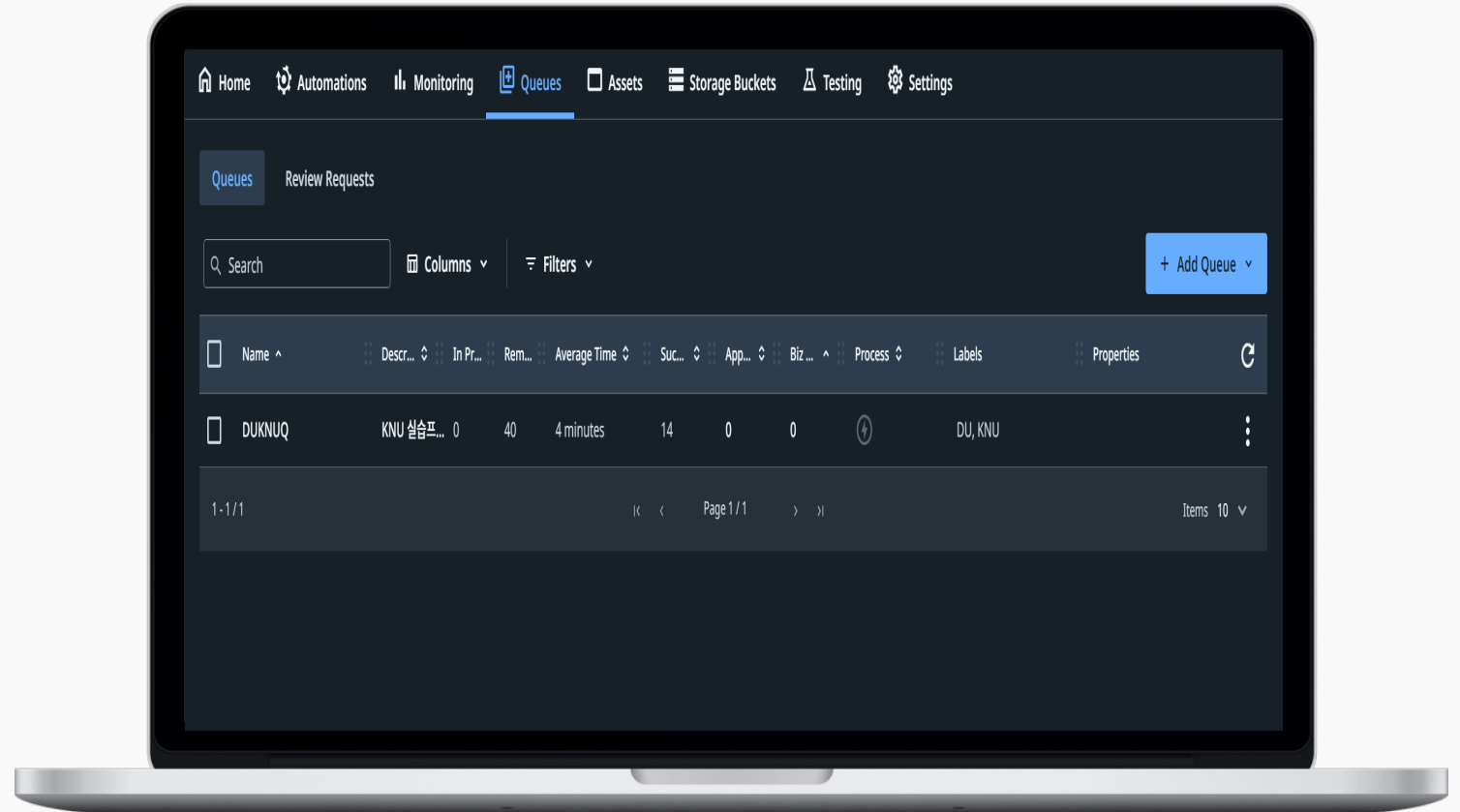


이슈사항 및 해결방안

Activity Action pending(해결 완료)

UiPath Studio - uploadFielstoQuete.xaml 실행 중
파일 선택 후 pending된 파일을 여러 팀원들이
무분별 사용하는 문제 발생

InvoiceTest.xam 파일로 Activity를 구성하여 무분별
공유 문제 해결



이슈사항 및 해결방안

Dom 객체 속성 오류(해결 완료)

Uipath Studio에서 건네주는 dom 객체와
Azure OCR API간 호환성 문제 발생

Dom 참조 방식을 간접 참조로 바꾸어 해결

```
private ExtractorResult ComputeResult(ExtractorDocumentType documentType, ResultsDocumentBounds documentBounds,
                                     string text, Document dom, string documentPath, string endpoint, string apiKey)
{
    var credential = new AzureKeyCredential(apiKey);
    var client = new DocumentAnalysisClient(new Uri(endpoint), credential);
    var extractorResult = new ExtractorResult();
    var resultsDataPoints = new List<ResultsDataPoint>();

    AnalyzeDocumentOperation operation = client.AnalyzeDocument(WaitUntil.Completed, "prebuilt-invoice", File.OpenRead(documentPath));
    AnalyzeResult result = operation.Value;
    // call kakao ml extractor with documentPath using endPoint and apiKey
    // create result data
    foreach (var x in result.Pages)
    {
        this.pages.Add(new PageLayout((double)x.Width, (double)x.Height, x.Unit.ToString()));
    }

    for (int i = 0; i < result.Documents.Count; i++)
    {
        AnalyzedDocument azformdoc = result.Documents[i];
        foreach (var du_field in documentType.Fields)
        {
            if (azformdoc.Fields.TryGetValue(du_field.FieldId, out DocumentField out_az_field))
            {
                // ...
            }
        }
    }
}
```

⑧ DUIntegration 실행이 시작되었습니다.

⑧ DUIntegration 실행이 종료되었습니다. in: 00:00:23

이슈사항 및 해결방안

파일 실행 시 포맷 에러 발생(해결 중)

디버깅 하기 힘든 포맷 오류가 발생하여 결과값에서
다 읽히지 못한 값들이 남아있음을 발견

Extractor Table에서 OCR Field 이름과 Uipath Field
이름을 매칭 예정

추출기 구성

추출기 구성

각 문서 형식 및 필드에 적용할 추출기를 구성합니다.
특정 필드에 추출기를 활성화하려면 구성자에 대한 적절한 상자를 선택합니다.
구성 및 데이터 추출에 정의된 분류법을 사용하는 추출기의 경우 문서 형식 및 필드의 고유 ID를 입력하는 것은 선택 사항입니다.
자체 내부 분류법을 사용하는 추출기의 경우 추출기를 활성화할 문서 형식 및 필드에 대해 모두 내부 분류법 고유 ID를 제공합니다.

문서 형식 및 필드	Charles Azure-Invoice Extractor (SimpleExtractor)	Machine Learning Extractor
	프레임워크 별칭 <input type="text"/> 최소 신뢰도 % <input type="text"/>	프레임워크 별칭 <input type="text"/> 최소 신뢰도 % <input type="text"/>
▶ Receipt	<input type="checkbox"/> azure.invoice.charlesdem.	<input type="checkbox"/> 해당 없음
▼ Invoice	<input checked="" type="checkbox"/> azure.invoice.charlesdem.	<input checked="" type="checkbox"/> 해당 없음
Name	<input type="checkbox"/> <input type="text"/>	<input checked="" type="checkbox"/> name <input type="text"/>
Vendor Address	<input type="checkbox"/> <input type="text"/>	<input checked="" type="checkbox"/> vendor-addr <input type="text"/>
Billing Name	<input type="checkbox"/> <input type="text"/>	<input checked="" type="checkbox"/> billing-name <input type="text"/>
Billing Address	<input type="checkbox"/> <input type="text"/>	<input checked="" type="checkbox"/> billing-addr <input type="text"/>
Billing VAT Number	<input type="checkbox"/> <input type="text"/>	<input checked="" type="checkbox"/> billing-vat-no <input type="text"/>
Shipping Address	<input type="checkbox"/> <input type="text"/>	<input checked="" type="checkbox"/> shipping-addr <input type="text"/>
Payment Address	<input type="checkbox"/> <input type="text"/>	<input checked="" type="checkbox"/> payment-addr <input type="text"/>
Vendor VAT Number	<input type="checkbox"/> <input type="text"/>	<input checked="" type="checkbox"/> vendor-vat-no <input type="text"/>
Date	<input type="checkbox"/> <input type="text"/>	<input checked="" type="checkbox"/> date <input type="text"/>
DueDate	<input type="checkbox"/> <input type="text"/>	<input checked="" type="checkbox"/> due-date <input type="text"/>
Invoice Number	<input type="checkbox"/> <input type="text"/>	<input checked="" type="checkbox"/> invoice-no <input type="text"/>
PO Number	<input type="checkbox"/> <input type="text"/>	<input checked="" type="checkbox"/> po-no <input type="text"/>
Payment Terms	<input type="checkbox"/> <input type="text"/>	<input checked="" type="checkbox"/> payment-terms <input type="text"/>
Shipping Charges	<input type="checkbox"/> <input type="text"/>	<input checked="" type="checkbox"/> shipping-charges <input type="text"/>
Tax Amount	<input type="checkbox"/> <input type="text"/>	<input checked="" type="checkbox"/> tax <input type="text"/>
Net Amount	<input type="checkbox"/> <input type="text"/>	<input checked="" type="checkbox"/> net-amount <input type="text"/>
Total	<input type="checkbox"/> <input type="text"/>	<input checked="" type="checkbox"/> total <input type="text"/>
Discount	<input type="checkbox"/> <input type="text"/>	<input checked="" type="checkbox"/> discount <input type="text"/>
Currency	<input type="checkbox"/> <input type="text"/>	<input checked="" type="checkbox"/> currency <input type="text"/>
▼ Items	<input type="checkbox"/> <input type="text"/>	<input type="checkbox"/> <input type="text"/>

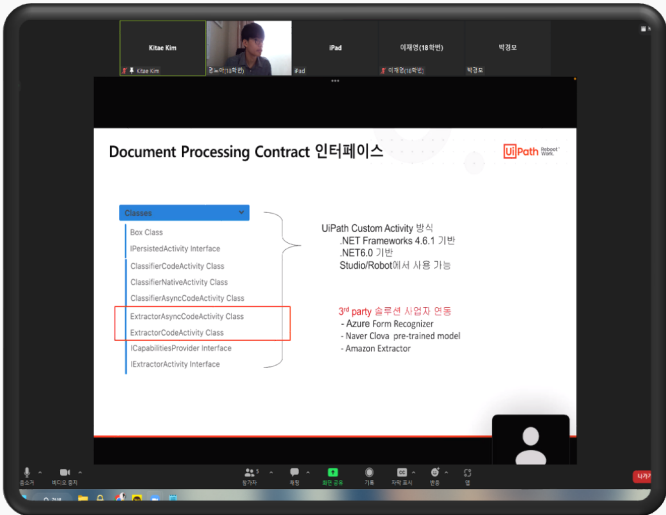
저장 취소

향후 계획

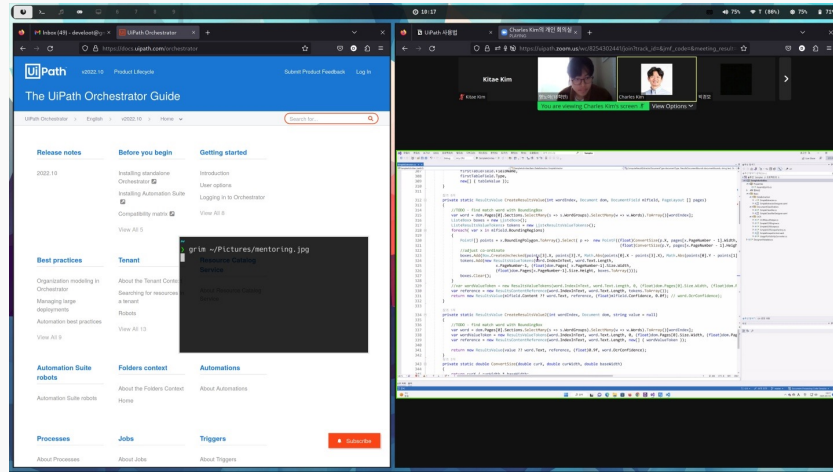
Our Mission

1. 추출한 구성에 개발한 Extractor의 Field 수정
2. 좌표 출력 기능 구현
3. Invoice(송장) 이외에 다른 모델 추가 구현
4. Microsoft Azure 구현 완료 후, 네이버 Clova OCR로 구현

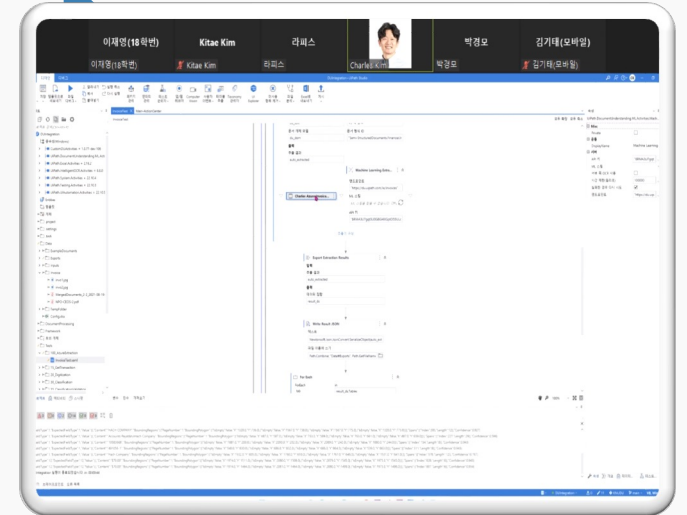




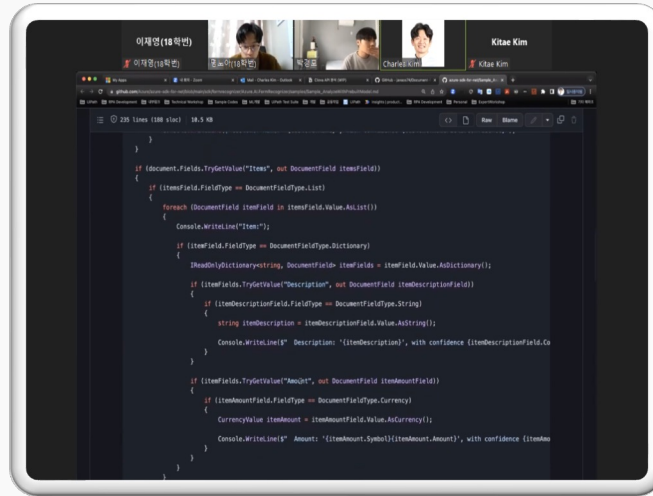
3.11일 회의(오리엔테이션)



3.18일 회의(디버깅 설명)



3.23일 회의(코드 구조 설명)



3.29일 회의(추가 Git소스코드, 테이블 매칭)



팀원회의 및 개발 후 식사)

회의 사진

일정 & 코드 관리

위키

자료

- 시간표
- 연락처
- 2023S_CDP2(OT).pdf 707.2KB
- 2023S SW분야 학부생중심 산학협력프로젝트 제안서_UIPath(김형수).pdf 183.6KB

기술 문서

- UiPath 사용법
- Azure Form Recognizer Endpoint 및 키 (보안 철저)
- Visual Studio 2022 설치 및 구성
- UiPath Studio X 사용법 (WIP)
- UiPath 분석 (WIP)
- Activity 분석 (WIP)
- Clova API 분석 (WIP)
- Azure API 분석 (WIP)

발표 초안

- 수행계획 발표
- 1차 중간 발표
- 최종 발표

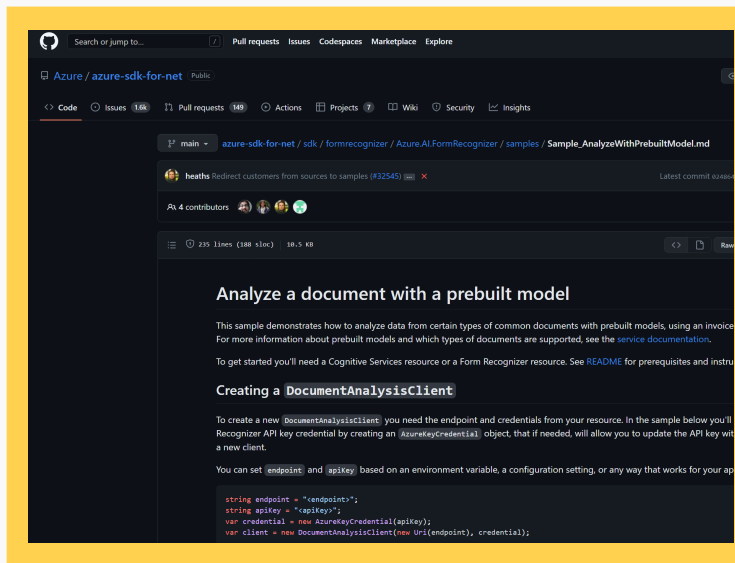
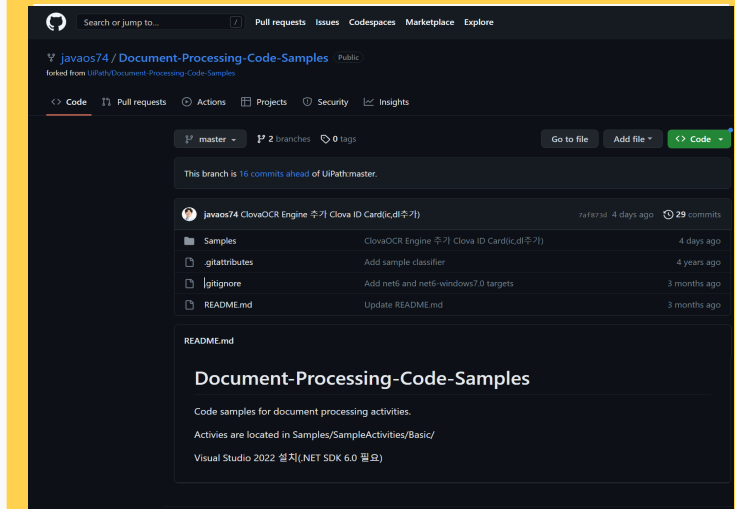
산출물

- 수행계획서 (초안).hwp 411.5KB

외부 링크

- Github: <https://github.com/UiPath-Document-Understanding-Project>
- 예제 코드: [Document-Processing-Code-Samples](#)
- 공식 문서: <https://docs.uipath.com/>
- 아자르 : https://github.com/Azure/azure-sdk-for-net/blob/main/sdk/formrecognizer/Azure.AI.FormRecognizer/samples/Sample_AnalyzeWithPrebuiltModel.md

Notion으로 일정 및 지식 공유



Github로 코드 공유 및 관리



Q&A