

문서인식 서비스 개발

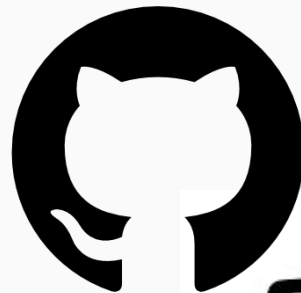
2차 중간발표

리마인드 - 목표 산출물

- 문서인식 프로그램 구현
 - Microsoft Azure 및 Naver Clova에 대응 가능하도록 구현
- 프로그램 저작권 등록

리마인드 - 프로젝트 관리

- 소스 코드는 Git과 Github로 관리
- 자료 및 일정 공유는 Notion을 사용중



프로젝트 분석 - UiPath란?

- Robot Process Automation (RPA) 업무 자동화 플랫폼
- 이메일을 통한 고객 응대 등 자동화 가능
- **AI를 활용**해 거래 명세서, 화물 송장 등 문서 처리 파이프라인 구현 가능
- Platform as a Service (PaaS) 제공


UiPath Document Understanding

- 아날로그 문서를 Optical Character Recognition(OCR)을 통해 인식
- AI를 이용해 문서의 양식 분류 및 처리
- 미리 정의되고 학습된 양식이 아니면 정확한 처리가 힘들

UiPath Document Understanding (Contd.)

- 미리 정의되고 학습된 양식이 아니면 정확한 처리가 힘들
- 각각의 나라, 산업 분야마다 조금씩 다른 양식을 사용함

Business & Talent Aligned

 **Right Management Group**

Rec No. 16611

30 November 2018
FOR PROFESSIONAL SERVICES

PO: NP74006735
TAX INVOICE 163-1802292-000

Melissa Bagnato
Motorola Australia Pty Ltd
10 Wesley Court
Tolly Ho Business Park
EAST BURWOOD VIC 3151

PROGMAM **FE**

RightChoice One Month	\$1,282.00
RightChoice Two Month	\$1,884.00
RightChoice Three Month x 2	\$5,320.00
RightChoice Six Month	\$5,186.00
GST 10%	\$1,367.20
GST INCLUSIVE FEE	\$15,039.20

Methods of Payment:

Cheque: Please make payable to: Right Management Consultants Pty Ltd

Direct Credit:

Bank: Citibank NA, Australia
Branch: 2 Park Street, STONEY NSW 2000
BSB: 245 000
Account Number: 230 103 003
Swift Code: CITI2AU2
Fax remittance advice to: 03 8554 2211
Email remittance advice to: melbourne.ar@right.com

PAYABLE WITHIN 15 DAYS OF RECEIPT

Right Management Consultants Pty Ltd ABN 62 306 132 163
Level 6, 112 St Marks Road, Melbourne VIC 3004 Australia
T +61 3 8554 2200 F +61 3 8554 2211 www.right.com

Hisco
Fast Friendly Service

12310 Tejon Street
Westminster, CO 80234-2321
Phone: (303) 387-4843
Fax: (303) 387-4009

INVOICE
US Dollar

CUST #	INVOICE DATE	INVOICE NUMBER
18276538	11/11/19	11810519-00
P.O. NO.	PAGE #	
0974081	1 of 3	

Send All Payments To:
HIS COMPANY, INC.
PO BOX 679472
DALLAS, TX 75267-9472

BILL TO:

HACH Co.
ATTN: ACCOUNTS PAYABLE
PO Box 389
Loveland, CO 80539-0389

SHIP TO:

HACH COMPANY
5608 LINDENBUSH DR
LOVELAND, CO 80538-8842

INSTRUCTIONS		TERMS	
Collect		1% 10 NET 30	
SHIP POINT		SHIP VIA	
HISCO-DENVER		UPS GROUND	11/11/19

LINE	PRODUCT AND DESCRIPTION	QUANTITY ORDERED	QUANTITY B.O.	QTY SHIPPED	QTY UM	UNIT PRICE	PRICING UNIT	AMOUNT (NET)
*****NO LOG SKETCHES*****								
SHIP ALL ORDERS COLLECT: UPS ACCURANDY FREIGHT BILLED ON AN INVOICE WILL NOT BE PAID. 15P OR 05P. C, B, F, C MUST BE PACKAGED IN THE BOX. DO NOT ATTACH TO THE PACK SLP. ALL LOCITE AND DOW CORNING MUST SHIP WITH A C OF C. PLEASE PACKAGE IN A SEPARATE BOX.								
1	STTC-038-33789 20 TIP CRISSEL 1.5MM 38 DEG BBFF FOR MX SYSTEMS Country of Origin: UNITED STATES Customer Prod: STTC-0838	20	0	20	EA	24.26	EA	485.20
2	STTC-126-33789 10 TIP BENT SHARP .4MM 30 DEG 708F FOR MX SYSTEMS Country of Origin: UNITED STATES Customer Prod: STTC-126	10	0	10	EA	23.68	EA	236.80
3	STTC-126-33789 5 TIP CRISSEL 1MM 30 DEG 708F FOR MX SYSTEMS Country of Origin: UNITED STATES Customer Prod: STTC-126	5	0	5	EA	24.97	EA	124.85
4	SFP-0805-33789 2 TIP CRISSEL BENT .5MM STANDARD FOR WFR-HSR Country of Origin: UNITED STATES Customer Prod: HSC0000 HISC 916533	2	0	2	EA	15.28	EA	30.40
5	STTC-138-33789 5 TIP CRISSEL 1.5MM 38 DEG 708F FOR MX SYSTEMS Country of Origin: UNITED STATES Customer Prod: STTC-126	5	0	5	EA	23.12	EA	115.60
6	SHTC-160-33789 5 METCAL 700 SHTC RW CART LONG BLADE 10MM Country of Origin: UNITED STATES Customer Prod: STTC-126	5	0	5	EA	85.95	EA	429.75
7	SHTC-162-33789 3 TIP BLADE LONG 22.1MM 708F FOR MX SYSTEMS Country of Origin: UNITED STATES Customer Prod: STTC-126	3	0	3	EA	86.61	EA	259.83
8	SHTC-164-33789 4 BLADE TIP CARTRIDGE 1MM Country of Origin: UNITED STATES Customer Prod: STTC-126	4	0	4	EA	89.68	EA	358.72
9	TT-95-50391 5 SOLDER TIP TINNER .70 OZ LEAD FREE RESIDUE FREE	5	0	5	EA	9.98	EA	49.90

Page 1 of 3
All sales are made subject to Hisco's Standard Sales Terms and Conditions at www.hisco.com/terms

INVOICE

Sirius Cybernetics Corp.

4592 Bell Street
New York, NY 10018

BILL TO **SHIP TO** **INVOICE #**

CHOAM CHOAM Invoice Date 27/01/2016

27 Shield Wall Ave, 27 Shield Wall Ave, P.O.# 16012633
Carthag, CH 1965 Carthag, CH 1965
Anakis Anakis Due Date 26/02/2016

Qty	Description	Unit Price	Amount
10	Nutrimatic Drinka Dispenser	4,200.00	42,000.00
17	Shipboard Computer "Eddie"	8,402.00	142,834.00
3	Happy Vertical People Transporters	21,000.00	63,000.00
Subtotal			247,834.00
VAT 19.0%			47,088.46
Total			\$294,922.46

Terms & Conditions
Payment is due within 30 days

Thank you

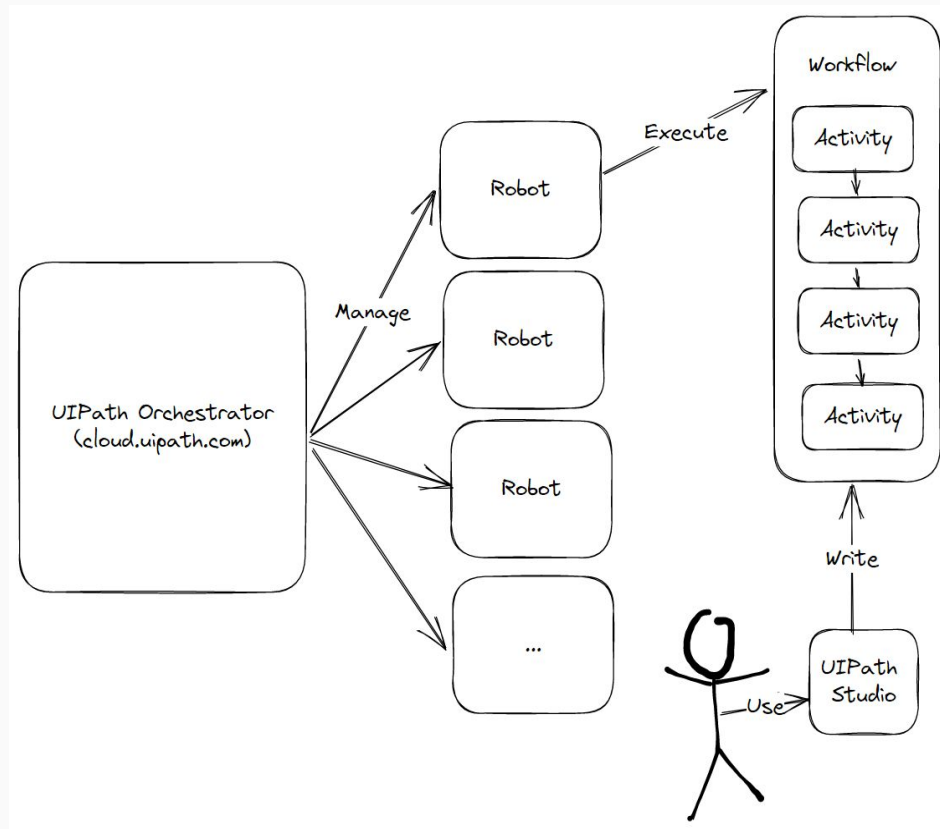
프로젝트 목표

더 많은 종류의 문서를 처리할 수 있도록 서드파티 문서 인식 서비스를 활용할 수 있도록 만들자.

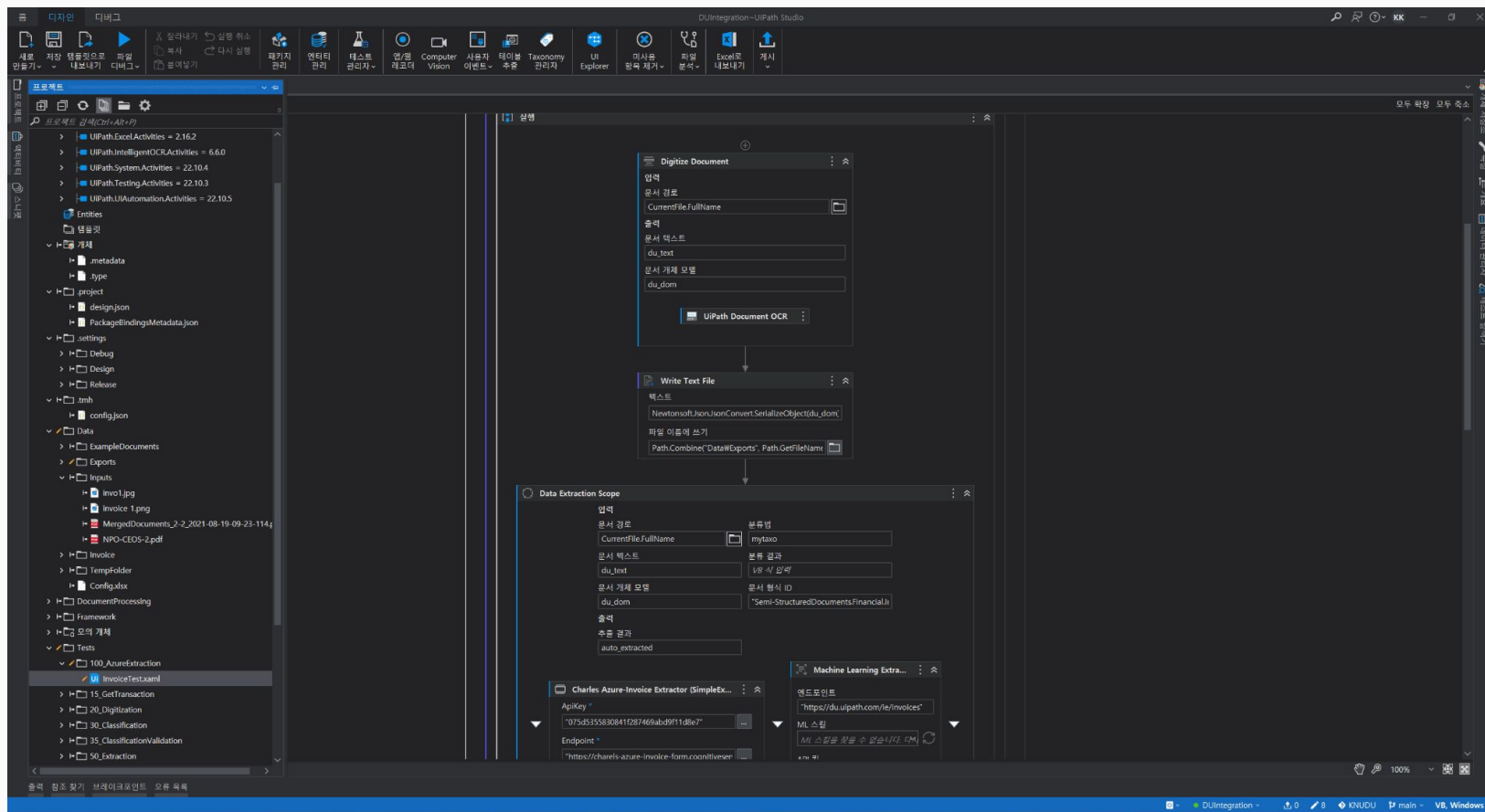
- **Microsoft Azure Document Understanding**
- *Naver Clova OCR*

UiPath 프로그램 구조

- UiPath Studio로 Workflow 프로그래밍
- Robot은 Workflow를 실행하는 주체로 여럿 있을 수 있음
- Orchestrator가 Workflow를 배포하고 Robot을 관리
- Orchestrator와 Robot은 직접 배포하거나 PaaS 서비스를 이용할 수 있음 (cloud.uipath.com)



UiPath 프로그램 구조 (Contd.)



UiPath 프로그램 구조 (Contd.)

- Workflow는 여러개의 Activity로 이루어져 있다.
- UiPath Studio에서 Activity를 Drag&Drop해 Workflow를 구성한다.
- Activity는 C#으로 작성되었고 UiPath측에서 다양하게 제공한다.

프로젝트 목표

더 많은 종류의 문서를 처리할 수 있도록 서드파티 문서 인식 서비스를 활용할 수 있도록 만들자.

즉? 서드파티 문서 인식 서비스를 활용하는 **Activity**를 작성하자.

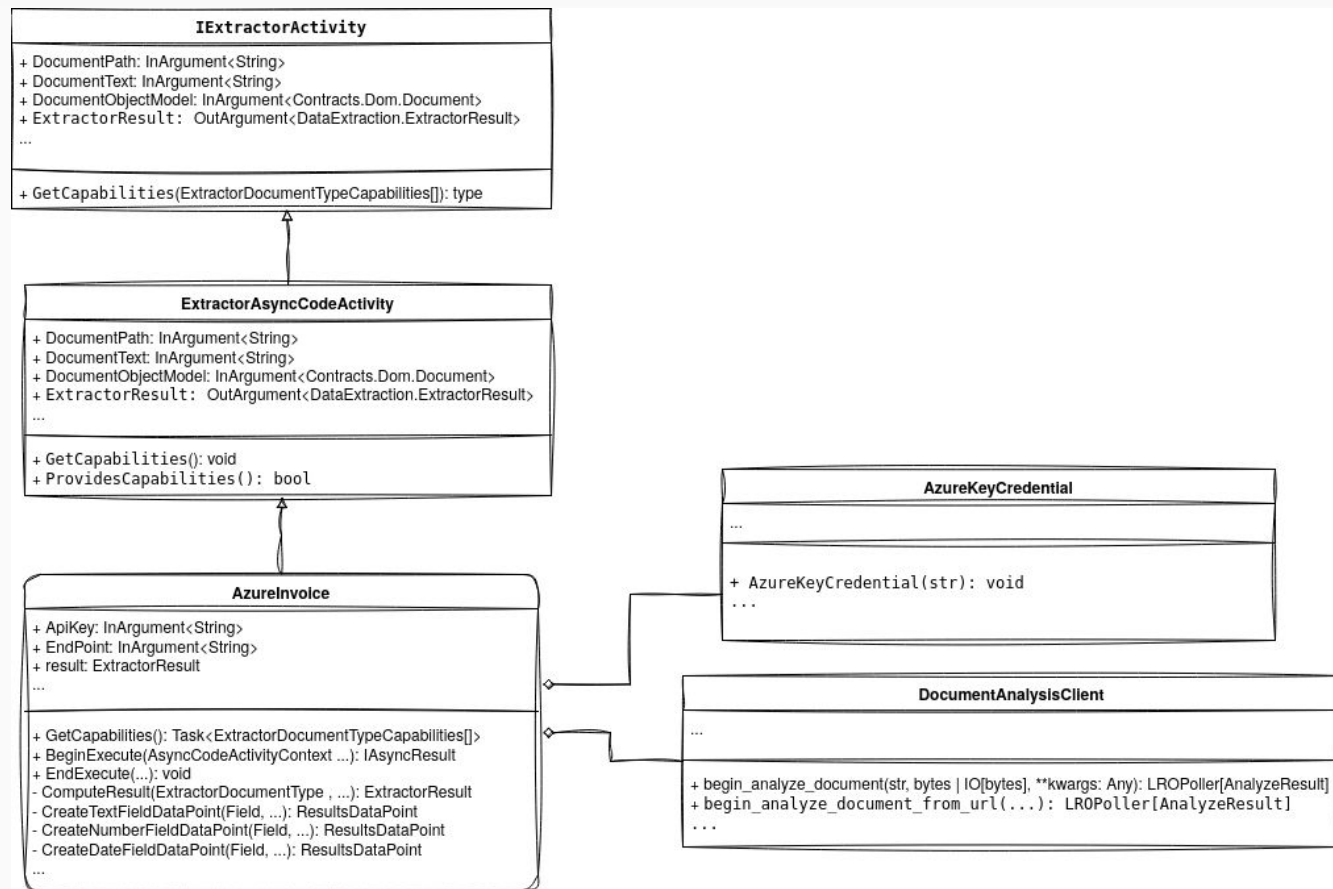
Activity for Microsoft Azure

```
35 using System.Diagnostics;
36 //using System.Diagnostics.Metrics;
37 using System.Linq;
38 using System.Threading.Tasks;
39 using UIPath.DocumentProcessing.Contracts;
40 using UIPath.DocumentProcessing.Contracts.DataExtraction;
41 using UIPath.DocumentProcessing.Contracts.Dom;
42 using UIPath.DocumentProcessing.Contracts.Results;
43 using UIPath.DocumentProcessing.Contracts.Taxonomy;
44 using System.IO;
45 using System.Drawing;
46
47 namespace SampleActivities.Basic.DataExtraction
48 {
49     class PageLayout
50     {
51         public double Width { get; set; }
52         public double Height { get; set; }
53         public string Unit { get; set; }
54
55         public PageLayout(double width, double height, string unit) {
56             Width = width;
57             Height = height;
58             Unit = unit;
59         }
60     }
61
62     [DisplayName("Charles Azure-Invoice Extractor")]
63     public class AzureInvoice : ExtractorAsyncCodeActivity
64     {
65         [Category("Server")]
66         [RequiredArgument]
67         [Description("ML모델 서비스 endpoint 정보")]
68         public InArgument<string> Endpoint { get; set; }
69
70         [Category("Server")]
71         [RequiredArgument]
72         [Description("ML모델 서비스 endpoint Api key 정보 ")]
73         public InArgument<string> ApiKey { get; set; }
74
75         Object lockObj = new Object();
76
77         ExtractorResult result;
78         List<PageLayout> pages;
79         public override Task<ExtractorDocumentTypeCapabilities[]> GetCapabilities()
80         {
81             #if DEBUG
82             Debug.WriteLine("GetCapabilities called");
83             #endif
84
85             //Azure Form Recognizer Invoice fields definition
86             List<ExtractorFieldCapability> fields = new List<ExtractorFieldCapability>();
87
88             fields.Add(new ExtractorFieldCapability { FieldId = "CustomerName", Components = new ExtractorFieldCapability[0], SetValues = new string[0] });
89             fields.Add(new ExtractorFieldCapability { FieldId = "CustomerId", Components = new ExtractorFieldCapability[0], SetValues = new string[0] });
90             fields.Add(new ExtractorFieldCapability { FieldId = "PurchaseOrder", Components = new ExtractorFieldCapability[0], SetValues = new string[0] });
91             fields.Add(new ExtractorFieldCapability { FieldId = "InvoiceId", Components = new ExtractorFieldCapability[0], SetValues = new string[0] });
92             fields.Add(new ExtractorFieldCapability { FieldId = "InvoiceDate", Components = new ExtractorFieldCapability[0], SetValues = new string[0] });
93             fields.Add(new ExtractorFieldCapability { FieldId = "VendorName", Components = new ExtractorFieldCapability[0], SetValues = new string[0] });
94             fields.Add(new ExtractorFieldCapability { FieldId = "VendorTaxId", Components = new ExtractorFieldCapability[0], SetValues = new string[0] });
95             fields.Add(new ExtractorFieldCapability { FieldId = "VendorAddress", Components = new ExtractorFieldCapability[0], SetValues = new string[0] });
96             fields.Add(new ExtractorFieldCapability { FieldId = "CustomerAddress", Components = new ExtractorFieldCapability[0], SetValues = new string[0] });
97             fields.Add(new ExtractorFieldCapability { FieldId = "CustomerTaxId", Components = new ExtractorFieldCapability[0], SetValues = new string[0] });
98             fields.Add(new ExtractorFieldCapability { FieldId = "CustomerAddressRecipient", Components = new ExtractorFieldCapability[0], SetValues = new string[0] });
99             fields.Add(new ExtractorFieldCapability { FieldId = "BillingAddress", Components = new ExtractorFieldCapability[0], SetValues = new string[0] });
100             fields.Add(new ExtractorFieldCapability { FieldId = "BillingAddressRecipient", Components = new ExtractorFieldCapability[0], SetValues = new string[0] });
101         }
102     }
103 }
```

Activity for Microsoft Azure

- UiPath측에서 정의한 **Activity** 클래스의 컨벤션이 있음
- 다양한 종류의 **Activity** 클래스가 있고 각기 다른 여러 제약이 있음
- **Dependency Injection (DI)** 클래스 개발 작업과 닮은 부분이 있음
- **Azure**측에서 제공하는 **C# SDK** 사용해 **Rest API**에 연결

Activity for Microsoft Azure - 클래스 다이어그램



진행상황



개발 중인 프로그램 스크린샷

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 / Country / 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 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.

```
DocumentId: "invo1.jpg"
ResultsVersion: 0
▼ ResultsDocument:
  ▼ Bounds:
    StartPage: 0
    PageCount: 1
    TextStartIndex: 0
    TextLength: 2288
    Language: "eng"
    DocumentGroup: "Semi-StructuredDocuments"
    DocumentCategory: "Financial"
    DocumentTypeId: "Semi-StructuredDocuments.Financial.Invoice"
    DocumentTypeName: "Invoice"
    DocumentTypeDataVersion: 0
    DataVersion: 0
    DocumentTypeSource: 0
  ▼ DocumentTypeField:
    Components: []
    Value: "Invoice"
    UnformattedValue: ""
  ▼ Reference:
    TextStartIndex: 0
    TextLength: 0
    Tokens: []
    DerivedFields: []
    Confidence: 1
    OperatorConfirmed: false
    OcrConfidence: -1
  ▼ Fields:
    ▼ 0:
      FieldId: "Semi-StructuredDocuments.Financial.Invoice.Name"
      FieldName: "Name"
      FieldType: 0
      IsMissing: false
      DataSource: 0
      ▼ Values:
        ▼ 0:
          Components: []
          Value: "Hach Co.\nATTN: ACCOUNTS PAYABLE"
          UnformattedValue: ""
```


프로그램 명세서

나. 기능적 요구사항 (Functional Requirements)

1) 정보 흐름 (Information Flows)

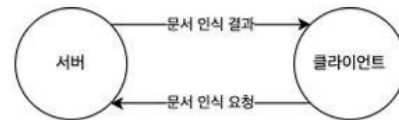


그림 1. 서버-클라이언트의 자료 흐름도(레벨 1)

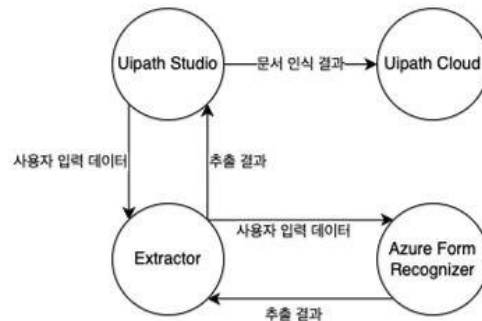


그림 2. 소프트웨어 인터페이스간 자료 흐름도(레벨 2)

이름	외부 문서 추출기 Integration 프로그램
목적 / 내용	문서 추출기로 생성된 결과값을 UiPath 프로그램에 통합 후 UiPath Cloud로 전송
입력 주체 / 출력 목적지	UiPath Studio / UiPath Cloud
범위 / 정확도 / 허용오차	외부 문서 추출기의 결과값에 따른 범위 / 외부 문서 추출기의 정확도를 따름 / 외부 문서 추출기의 허용 오차를 따름
단위	문서 입력
시간 / 속도	프로그램 실행 속도
타 입출력과의 관계	입력 내용에 따라 클라이언트에서 처리 후 Cloud 서버로 명령 요청
화면 형식 및 구성	IDE 형식의 구성
윈도우 형식 및 구성	해당 없음
데이터 형식	xaml
명령 형식	파일 실행
종료 메시지	실행 후 자동 종료

프로그램 명세서 (Contd.)

3) 소프트웨어 인터페이스

이름	RestAPI를 이용한 문서 추출기
목적 / 내용	RestAPI를 통해 Azure Form Recognizer 서비스로 문서를 전송하고 결과값을 생성
입력 주체 / 출력 목적지	Uipath Studio / Uipath Studio
범위 / 정확도 / 허용오차	인식하고자 하는 문서의 양식에 따른 입력 범위 / Azure Form Recognizer의 정확도를 따름 / Azure Form Recognizer의 허용 오차를 따름
단위	문서 입력
시간 / 속도	Azure Form Recognizer의 실행 속도
타 입출력과 관계	입력 내용에 따라 클라이언트에서 처리
화면 형식 및 구성	해당 없음
윈도우 형식 및 구성	해당 없음
데이터 형식	c# 프로그램을 빌드하여 생성되는 DLL파일
명령 형식	빌드
종료 메시지	실행 후 자동 종료

가. 외적 인터페이스 요구사항 (External Interface Requirements)

1) 사용자 인터페이스 (User Interfaces)

이름	Uipath Studio를 통한 입력 처리
목적 / 내용	Uipath사의 클라이언트가 Uipath Studio를 통하여 인식하고자 하는 문서 전달
입력 주체 / 출력 목적지	사용자 / Uipath Cloud
범위 / 정확도 / 허용오차	인식하고자 하는 문서의 양식에 따른 입력 범위 / 문서의 품질에 따른 입력 정확도 / Uipath studio의 파일 인식률에 따른 허용 오차
단위	문서 입력
시간 / 속도	즉각적인 사용자 명령 수행
타 입출력과 관계	입력 내용에 따라 클라이언트에서 처리하거나 서버로 명령 요청
화면 형식 및 구성	IDE 형식의 구성
윈도우 형식 및 구성	그래픽 유저 인터페이스
데이터 형식	인쇄 및 손글씨 양식, PDF, 이미지 파일
명령 형식	파일 실행 명령
종료 메시지	실행 후 자동 종료

이슈사항 및 해결현황 - 해결됨

1. UiPath 클라우드 서비스에 로그인에 불가능한 상황 발생
 - a. 서버쪽 문제로 추측. 해결됨

이슈사항 및 해결현황 - 진행중

1. 하나의 **Field**안의 데이터가 띄어쓰기를 기준으로 여러 **Word**로 분리됨
 - a. 따라서 이것들을 합쳐 하나의 데이터로 만들 필요가 있음
2. **Azure** 측에서 반환하는 다양한 데이터 타입을 처리할 필요가 있음
 - a. 테이블 데이터, 리스트 데이터, 숫자 값 등 각각 핸들링할 필요가 있음

QnA

