



ICR 소개자료

(Intelligent Character Recognition)

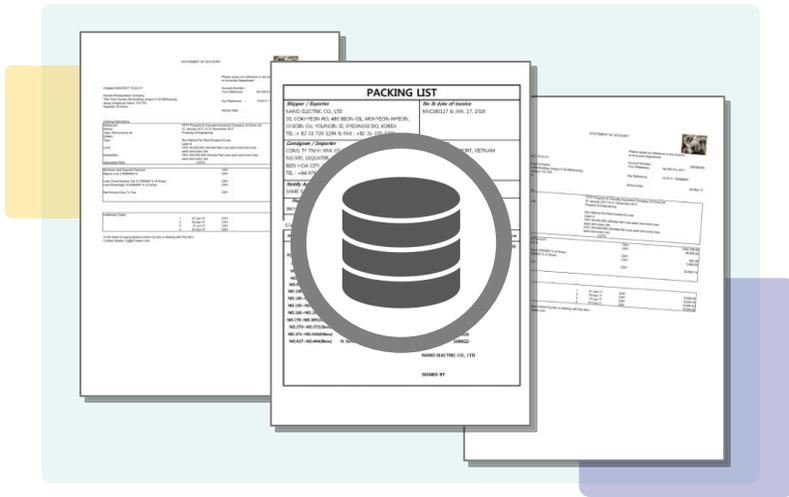
목 차 Contents

1. ICR 소개	3
2. System 구성도	
2.1. Cloud 구성방식	4
2.2. On-premise 구성방식	5
3. 진료비 영수증, 진단서 인식 Process	
3.1. 전처리 구현방안(선에 걸침, 뒤틀림 등의 보정방안)	6
3.2. 후처리 구현방안(검증)	7
4. 사례	
4.1. 코리안리	8
4.2. 대림산업	9
4.3. 캐롯손해보험	10

1. ICR 소개

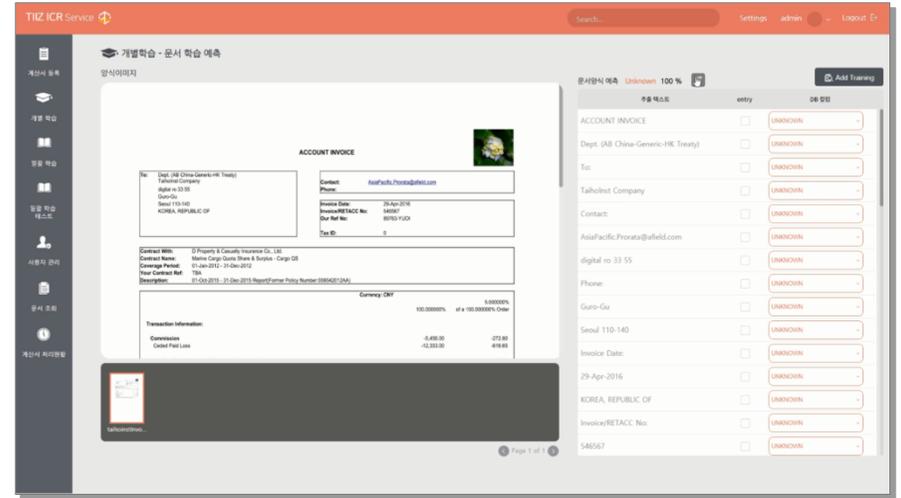
TIIZ ICR은 상용OCR의 1차 결과를 바탕으로 자체 개발한 ML의 기능을 적용한 시스템으로 보다 명확하고 빠른 문서 내용의 DB화 및 분석이 가능하도록 한 솔루션입니다.

상용 OCR Engine의 클라우드 버전은 Microsoft의 OCR 엔진을 사용하여 당사의 TIIZ솔루션과 통합 환경을 구축하여 사용할 수 있습니다.



#저장소

저장소에 저장된 다양한 스캔 이미지를 불러와 사용합니다.



#양식분석

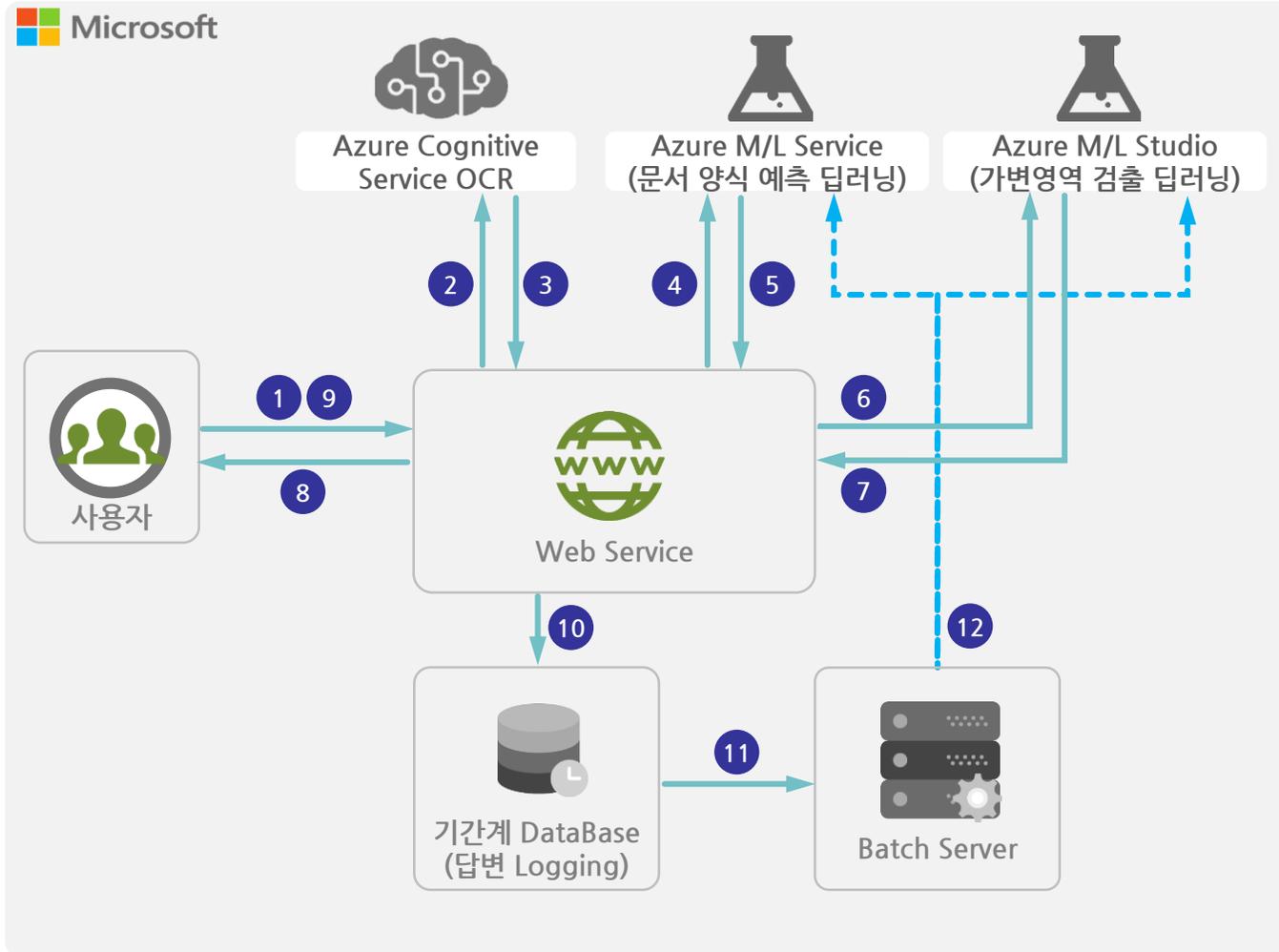
저장소의 파일을 머신러닝 분석으로 ICR 처리/진행합니다.

- 여러 종류의 문서를 분야에 맞게 문서 분류
- 분류한 문서의 사용 용도에 따라 원하는 값 추출

- 문서별 분류 현황, 처리 건수, 사용자별 처리 현황 등에 대해 대시보드 형식으로 제공하여 간편하게 ICR 현황 관리

2. System 구성도

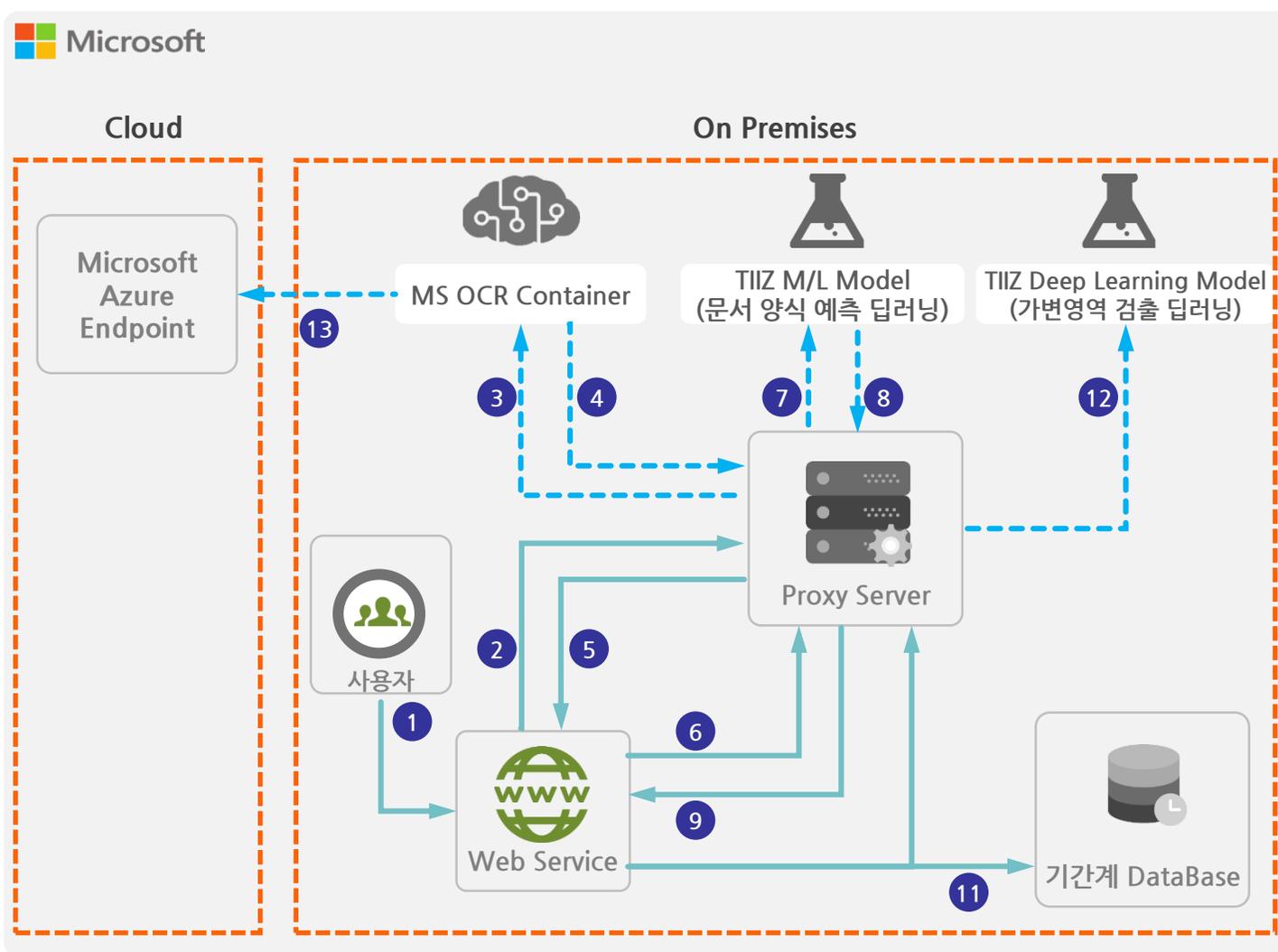
1. Cloud 구성방식



- 1 분석할 이미지를 전송
- 2 Cognitive Service OCR로 이미지 전송
- 3 OCR 데이터 Web Service로 송신
- 4 OCR 데이터 M/L Service로 전송
- 5 OCR 데이터 오타 판단하여 수정 후 데이터를 Web Service로 송신
- 6 텍스트의 오타유무 판단 및 수정하여 Web Service로 송신
- 7 OCR 데이터를 이용하여 유효한 영역 판단하여 검출 데이터 Web Service로 송신
- 8 가공한 OCR 데이터를 사용자에게 송신
- 9 사용자는 수정사항이 있을 경우 수정 후 Web Service로 송신
- 10 사용 이력을 기간계 DB로 로깅
- 11 추가학습 사항이 있을 경우 Batch Service로 송신
- 12 정해진 시간에 추가학습 진행

2. System 구성도

2. On premise 구성방식



- 1 분석할 이미지를 전송
- 2 수신 이미지 프록시 서버 전송
- 3 TIIZ OCR Service로 전송
- 4 메타 데이터 프록시 서버 전송
- 5 수신 메타 데이터 Web Service 전송
- 6 비식별화 등 전처리 데이터 프록시 서버 전송
- 7 수신 전처리 데이터 ML Model 전송
- 8 가공한 OCR 데이터 프록시 서버 전송
- 9 수신 OCR 데이터 Web Service 전송
- 10 수정 데이터 있을 시 프록시 서버 전송
- 11 최종 데이터 기간계 DB 저장
- 12 추가 학습 데이터 Deep Learning Model 전송 후 추가 학습
- 13 과금을 위한 사용량 정보가 Microsoft Cloud로 전송

※ 현재 영문만 가능합니다. 한글의 경우 2020년 말경에 Preview Version 출시 예정입니다.

3. 진료비 영수증, 진단서 인식 Process

1. 전처리 구현방식

Image adjustment ML을 통하여 문서의 기울기, 회전 및 뒤집기, 선에 걸침 등에 대한 이미지 보정을 수행한다.



Image pre-processing

- 기울기 수정
- 구부러진 텍스트 라인 수정
- 노이즈 및 흐릿함 제거
- 문서 배경색상을 흰색으로 변경
- 부등변 사각형 일그러짐 보정
- 사다리꼴 왜곡을 보정하며 유용한 데이터를 포함하지 않은 이미지 모서리 제거
- 회전 및 뒤집기
- 색상 마크 제거



추가 개체 인식 수행

- Object detection - Check Box, Radio Box
- Mark recognition

3. 진료비 영수증, 진단서 인식 Process

2. 후처리 구현방식

(1) 오자 검증 및 수정

1차 인식된 문자의 정합성 검증과 정확도 향상을 위하여 다양한 후처리 방식을 적용한다.

적용예시)

1step - 오자 검증, □ 과 o, l과 I 등과 같이 잘 못된 Detection이 발생할 확률이 높은 단어에 대한 Library를 구성하여 오자를 수정

2step - Anomaly(변칙) detection ML 검증 방법을 사용하여 필드데이터 길이, 타입 등에 대한 이상을 감지 후 수정

1222.00

1540

3000.00

2400.00

The screenshot shows the TIIZ ICR Service interface. On the left is a navigation menu with icons for '계산서 등록' (Invoice Registration), '개별 학습' (Individual Learning), '모집 학습' (Batch Learning), '사용자 관리' (User Management), '문서 조회' (Document Search), '계산서 처리현황' (Invoice Processing Status), '문서 등록' (Document Registration), and '문서 업로드' (Document Upload). The main area is titled '문서 업로드' (Document Upload) and contains a 'Single Record' form with fields for Buzer, PO Number, PO Date, Delivery Address, Total Price, Currency, and Requested Delivery Date. To the right is a '문서 이미지' (Document Image) section showing a scanned document with a 'Voucher No' and 'Voucher Date' field. Below the image is a 'Multi Record' table with columns for Material, EAN, Quantity, Unit Price, Item Total, and Serial Number. The table contains 6 rows of data, with the 5th row highlighted.

Material	EAN	Quantity	Unit Price	Item Total	Serial Number
20	HG48EE690DBXXU/Samsung 49inch Full HD Smart Comm TV	407	364.00	148148.00	20
21	HG48E690UBXXU/DEAL43inch 4K Commercial TV DEAL	150	340.00	51000.00	21
22	HG48E690UBXXU/49inch 4K Smart Comm TV	20	406.00	8120.00	22
23	HG48E690UBXXU/Samsung Smart UHD comm TV	32	390.00	12480.00	23
24	LH375HFLBB/EN37in Stretched Black Dlx Slim LED	0	902.50	8025.00	24
25	LH43PHFPMGC/EN43in Black Dlx Slim LED 24/7 700cd	3	660.25	1980.75	25
26	LH46JHFCCLB/EN46in Videowall 24/7 700cd	10	802.75	8027.50	26

3. 진료비 영수증, 진단서 인식 Process

2. 후처리 구현방식

(2) 대사처리

전처리, 후처리 과정을 모두 완료한 후에도, 100% 인식되지 않는 글자나 문구에 대하여 대사처리를 하여야 하는 경우가 발생합니다. 이러한 과정을 빠르고 신속하게 지원할 수 있도록 다음과 같이 지원합니다.

대사전 ML 적용

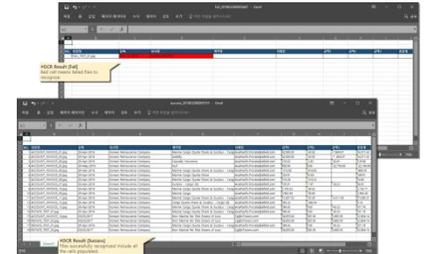
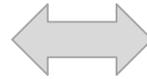
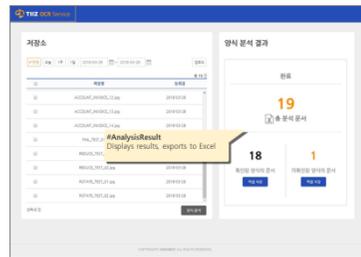
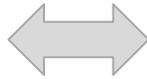
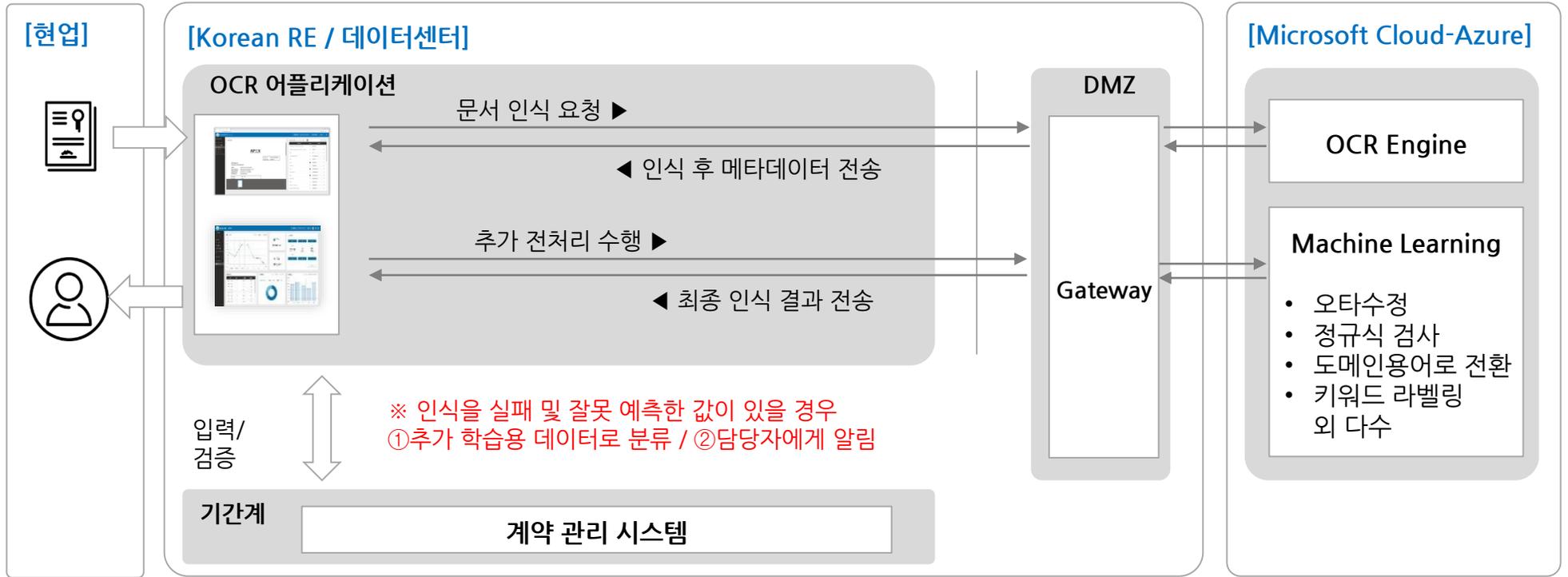
정확도 판정

- 변환된 Text의 정확도를 ML을 적용하여 % 단위로 판정
- 이정 범위 이하의 Column값은 Highlight, 별도 검증요구 등을 통한 사용자 재확인 절차
- 대사 처리 시, 원본 문서를 Upload하여 확인
- 별도 재학습 범위로 설정하여 재학습

The screenshot displays the TIIZ ICR Service interface. On the left, there is a sidebar with navigation options like '문서 업로드', '개별 학습', '신규문서 등록', and 'ICR 문서관리'. The main area shows a table of documents with columns for '파일명', '원형코드', '날짜', '비고', '공급자', and '사업자'. One document entry is highlighted in red, corresponding to the date '2019-11-23 19:04:16'. To the right, a detailed view of a document is shown, titled '레디믹스트 콘크리트 납품서'. This view includes a header with '일자: 2019-11-23' and '일자: 2019-11-23', and a table of data. A red box highlights the date '2019-11-23' in the document view, which matches the highlighted date in the table. Below the table, there are various technical specifications and a signature area.

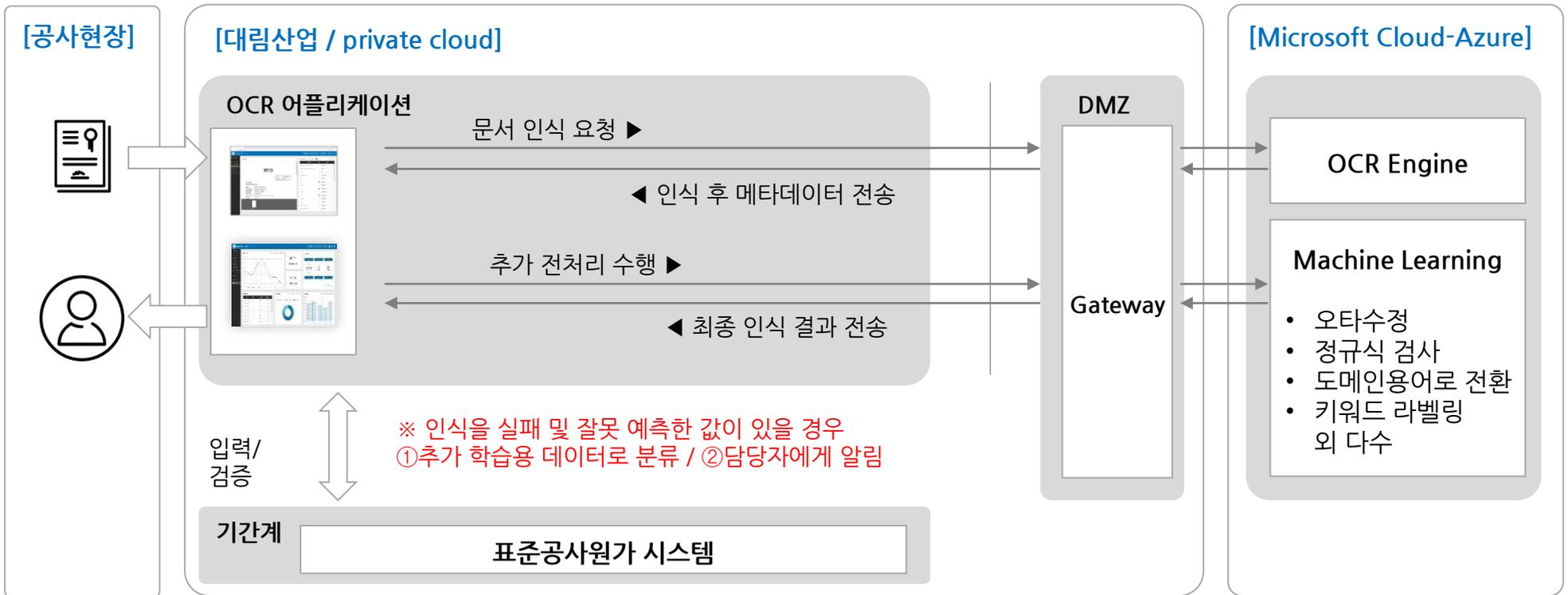
4. 사례 - 코리안 리

글로벌하게 수집되는 여러 타입의 계약서에 대하여 계약 관리 시스템에 입력하는 업무에 대한 자동화 사례입니다.



4. 사례 - 대림산업

아파트 건설 현장에서 입수되는 여러 타입의 송장(레미콘, 철근 일반자재)에 대하여 건설표준공사원가 데이터를 수집하는 업무에 대한 자동화 사례입니다.



4. 사례 - 캐롯손해보험

캐롯손해보험의 업무에 대한 자동화 사례입니다.

