

한국은행 문서 디지털화 사업수행계획서

작 성 일	2024.08.02
작 성 자	(주)크라우드웍스
	박정근 PM

목차

1	사업 개요	3
1.1	추진배경.....	3
1.2	추진 목표.....	3
1.3	추진 전략.....	3
1.4	기대 효과.....	3
2	과업 내역	4
2.1	과업 범위.....	4
2.2	과업 수행 계획.....	5
2.3	구축 데이터 명세	7
2.4	데이터 구축 프로세스	11
2.5	데이터 구축 상세	12
2.6	데이터 품질 관리	15
3	추진 일정	17
3.1	주요 추진 일정	17
3.2	세부 추진 일정	17
4	프로젝트 관리 지원	18
4.1	수행 조직.....	18
4.2	비상 대책.....	18

1 사업 개요

1.1 추진배경

- 생성형 인공지능의 빠른 발전과 함께 AI 챗봇, 문서 검색·요약 등 AI 언어모델 활용 수요가 확대됨에 따라, 한국은행 업무에 특화된 AI 언어모델 기반 서비스를 개발 필요
- 2000 년 이후 한국은행 홈페이지 수록 문서를 대상으로 언어모형 학습 등에 활용하고 의미검색기능을 구현하기 위한 문서 디지털화 작업 필요

1.2 추진 목표

- 시범 사업 수행 경험을 통한 홈페이지 수록 문서 디지털화 수행
- 언어모형 학습 등에 활용될 수 있는 AI 학습용 데이터셋 구축
- 한국은행 홈페이지 수록 문서 디지털화 (PDF to JSON(Markdown))

1.3 추진 전략

- 문서 디지털화 전용 도구에 기반한 Human Parsing 전략
- 문서 디지털화 전용 Annotation Tool 활용 전략
- 데이터 항목별 프로젝트 구분 작업 전략
- 주제별, 챕터별 문서 분류 작업 전략

1.4 기대 효과

- 한국은행 내 시스템과 원활하게 호환되고, 효과적으로 연계되는 작업 지원
- 데이터 접근성(검색), 처리(변환, 분석 등) 효율성, 호환성 증대
- 타 시스템과의 통합 용이, 데이터 활용성 증대
- AI 언어모델 기반 서비스 고도화

2 과업 내역

2.1 과업 범위

- 2000 년 이후 한국은행에서 생산되고 홈페이지에 공개된 문서 중 조사·연구 관련 주요 보고서 약 12 만 페이지를 문서 디지털화 가이드라인에 따라 JSON 형식의 Markdown 형식으로 변환
- 문서별 작업 수량

문서 유형	문서 수	페이지 수
BOK 경제연구	707 부	40,440
지급결제보고서	23 부	2,910
연차보고서	29 부	5,280
경제전망보고서	49 부	4,270
BOK 이슈노트	230 부	8,000
국민계정리뷰(논고, 자료만)	386 부	24,241
금융안정보고서	42 부	7,030
조사통계월보(논고만)	663 부	9,945
지역경제보고서	46 부	12,531
통화신용정책보고서	64 부	7,990
BOK 경제리뷰	49 부	1,695
BOK 경제브리프	6 부	190
합계	2,294 부	124,522

- 사업기간 내 발간되는 신규 문서는 모두 작업 대상에 포함합니다.
- 상기 수량은 문서 유형별 최근 5 개 문서에 대한 평균 수량이므로 금액 내 최대 125,000 페이지, 사업 예산 내 최대 132,000 페이지를 진행합니다. (기 작업한 약 8 천 페이지를 제외한 수량)

○ 문서 디지털화 작업 기준

- 문서 기본 정보
 - 문서 기본 정보를 메타 데이터로 구축
 - 문서유형(doc_type), 문서번호(doc_no), 파일명(file_path), 문서제목(title), 문서발간일(publish_datetime), 문서저자(authors), 참고문헌(references), 부연설명(disclaimer), 파일일련번호(file_part)
- 문서 내용 구조화
 - 문서 전체 내용을 항목별 구조화
 - 초록, 머리말(preface), 개요(abstract), 본문(body), 참고(annexes), 부록(appendix), 각주(footnote)
- 표, 수식, 이미지 처리
 - 단순한표는 Markdown 형식으로 변환
 - 복잡한 표(셀병합)는 별도의 이미지 파일로 저장하여 관리
 - 단순한 수식(문장內)은 LaTeX 문법에 맞게 Markdown 형식으로 변환
 - 복잡한 수식(문장外)은 별도의 이미지 파일로 저장하여 관리
 - 문서에 포함된 모든 이미지(차트는 별도의 이미지 파일로 저장하여 관리
- 문서 계층 구조 처리
 - 문서 전체 내용의 타이틀/챕터 간의 계층정보가 나타날 수 있도록 대단원, 중단원, 소단원 등의 계층구조에 따라 Markdown 형식(#, ##, ###, ...)에 맞게 변환
 - Markdown 형식의 글머리번호/기호를 활용하여 계층 구조에 따라 변환

2.2 과업 수행 계획

○ 프로젝트 수행 계획

- 시범 사업을 통한 요구사항 분석 및 문서별 상세 내용 분석 완료

구분	01. BOK(이슈노트)	02. BOK(발행인구)	03. 국제경제자료	04. 조사통계자료	05. 통치신상정보보고서	06. 경제인상정보보고서	07. 지역경제정보고서	08. 금융통계정보고서	09. 국민경제정보	10. 연차보고서	11. 기업정보보고서
doc_type	BOK(이슈노트)	BOK(발행인구)	국제경제자료	조사통계자료	통치신상정보보고서	경제인상정보고서	지역경제정보고서	금융통계정보고서	국민경제정보	연차보고서	기업정보보고서
doc_no	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000
file_path	파일명	파일명	파일명	파일명	파일명	파일명	파일명	파일명	파일명	파일명	파일명
title	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000
publish_datetime	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000
authors	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000
file_ext	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000
abstract	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000
acknowledgment	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000
body	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000
appendix	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000
reference	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000
disclaimer	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000	100-1000

<한국은행 발간 문서별 상세 분석 내용>

- 작업 가이드 준비 및 작업 화면 디자인 완료, 숙련된 작업자 보유
- 프로젝트 착수와 동시에 메타 데이터 구축/문서 내용 구조화, 전처리 작업 진행
- 메타 데이터 구축 프로젝트

doc_no

doc_type

title

authors (name, mail)

publish_datetime

disclaimer

references (author, year, title, source)

disclaimer

구분	값	단위	산출 근거
1 시간 작업량	5.5	부	시범 사업
1 일 작업량	22	부	5.5(부) * 4(h)
1 개월 작업량	484	부	22(부) * 22(d)
필요 인원(작업)	5	명	2300(부) / 484(부)
필요 인원(검수)	2.5	명	5(명) * 0.5
선발 인원(작업+검수)	15	명	필요 인원(7.5)의 2 배수

- 문서 내용 구조화, 전처리 프로젝트



구분	값	단위	산출 근거
1 시간 작업량	17	페이지	시범 사업
1 일 작업량	68	페이지	17(p) * 4(h)
1 개월 작업량	1,496	페이지	68(p) * 22(d)
3 개월 작업량	4,488	페이지	1,496(p) * 3(m)
필요 인원(작업)	27	명	120,000(p) / 4,488(p)
필요 인원(검수)	13	명	27(명) * 0.5
선발 인원(작업+검수)	80	명	필요 인원(40)의 2 배수

○ 인력 투입 계획

- 시범 사업 인력 40 명 중 우수 인력 25 명 선발
- 유사 프로젝트 참여 인력 중 140 명 모집
- 모집 인원에게 대한 교육 훈련 진행
- 교육 훈련 우수 인력 70 명 선발(0.5 배수)
- 총 95 명(25+70) 프로젝트 인력으로 투입

2.3 구축 데이터 명세

○ 데이터 포맷

데이터명	원시데이터 포맷	라벨링 데이터 포맷
------	----------	------------

한국은행 문서 디지털화 데이터	PDF	텍스트: JSON 이미지: PNG
---------------------	-----	-----------------------

○ 라벨링 데이터 어노테이션 명세 (주요 속성 발췌)

어노테이션 방법	주요 어노테이션 속성	속성 설명	라벨링 데이터 포맷
Tagging, Parsing	Doc_type	문서 유형	JSON
	Doc_no	문서 번호	
	Title	문서 제목	
	Publish_datetime	문서 발간일	
	authors	문서 저자	
	References	참고문헌	
	Disclaimer	기타 부연설명	
	preface	초록, 머리말	
	Abstract	개요	
	Body	본문	
	annexes	참고	
	Appendix	부록	
	Formula	수식	JSON PNG
	Image	이미지(차트 포함)	
	table	표	

○ 라벨링 데이터 예시

```
{
  "doc_type": "BOK 이슈노트",
  "doc_no": "제 2024-5 호",
  "file_path": "BOK 이슈노트 제 2024-5 호_빅데이터와 기계학습 알고리즘을  
활용한 실시간 인플레이션 전망(real-time inflation forecasting).pdf",
  "title": "빅데이터와 기계학습 알고리즘을 활용한 실시간 인플레이션  
전망(real-time inflation forecasting)",
}
```



```
"publish_datetime": "2024-02-06 00:00:00",
"authors": [
  {
    "name": "이창훈",
    "mail": "changhoon@bok.or.kr"
  },
  {
    "name": "홍지연",
    "mail": "jyhong@bok.or.kr"
  },
  {
    "name": "이현창",
    "mail": "hyunchang.yi@bok.or.kr"
  }
],
"file_part": "2",
"preface": "",
"abstract": "",
"disclaimer": [
  "본 자료의 내용은 한국은행의 공식견해가 아니라 집필자 개인의",
  "견해라는 점을 밝힙니다....(중략)"
],
"body": "# Ⅲ. 실시간 인플레이션 전망¶n 이 절에서는 실시간 인플레이션",
  "전망을 위한 데이터와 전망모형에 대해 살펴보고, ...(중략)",
"annexes": [
  ""
],
"appendix": "# <부록 1> 트리 기반 기계학습 기법에 사용된 예측변수",
  "목록 ...(중략)",
"references": [
  {
    "author": [
      "강규호",
      "김도완"
    ]
  },
]
```

```

        "year": "2023",
        "title": "대규모·비선형 베이지안 VAR 모델을 활용한 한국 거시경제
        전망 및 시나리오 분석",
        "source": "BOK 경제연구, 제 2023-17 호."
    }, ...(중략)
],
    "images": [
        {
            "id": "2_pic1",
            "caption": "<그림 1> 실시간 전망 프로세스",
            "text": "",
            "file_path": "images/2_pic1.png"
        },
        {
            "id": "2_pic2",
            "caption": "<그림 2> 표본외 예측력: 전망시계별
            MAE(2016.1~2023.9)",
            "text": "",
            "file_path": "images/2_pic2.png"
        }, ...(중략)
    ],
    "tables": [
        {
            "id": "2_table1",
            "caption": "<표 1> 예측변수 데이터셋",
            "text": "",
            "file_path": "tables/2_table1.png"
        },
        {
            "id": "2_table2",
            "caption": "<표 2> 표본 외 예측력: 앙상블 모형 MDA",
            "text": "",
            "file_path": "tables/2_table2.png"
        }, ...(중략)
    ],

```

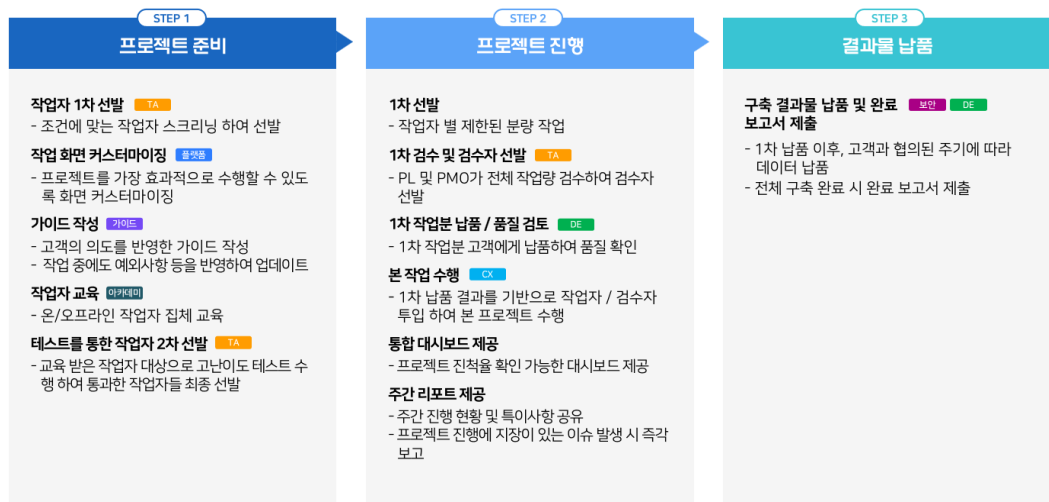
```

"formulas": [
  {
    "id": "2_formula1",
    "caption": "예측력 평가 기준: RMSE, MAE, MDA",
    "text": "",
    "file_path": "formulas/2_formula1.png"
  },
  {
    "id": "2_formula2",
    "caption": "",
    "text": "",
    "file_path": "formulas/2_formula2.png"
  }
]
}

```

2.4 데이터 구축 프로세스

○ 데이터 구축 프로세스 개요



○ 프로젝트 준비

- 시범 사업 프로젝트 참여자, 문서 디지털화 유사 프로젝트 참여자로 하여 작업자 1 차 선발
- 온/오프라인 작업자 집체 교육 및 우수 교육자 대상 최종 선발

- 원시 데이터 수집(한국은행 홈페이지 내 다운로드) 및 작업 가이드 작성
- 문서 유형별 문서 분석 및 원천 데이터로 정제(PDF 분리)
- Markdown editor 를 활용한 작업 화면 커스터마이징

○ 프로젝트 진행

- 메타 데이터 구축 프로젝트 오픈
- 문서 유형별 문서 전처리 프로젝트 순차 오픈
- 샘플 데이터 검토 및 품질 검수 후 납품
- 프로젝트 모니터링

○ 프로젝트 완료(결과물 납품)

- 문서 유형별 주기에 따라 결과 데이터 납품
- 완료 보고서 제출

2.5 데이터 구축 상세

○ 메타 데이터 구축 프로젝트

- 데이터 라벨링 도구(Annotation tool)를 활용하여 원본 문서를 보고 작업할 수 있는 직관적인 작업화면 디자인
- 작업 화면 내 왼쪽 원본 문서를 보고 문서 유형, 문서 번호, 문서 제목, 문서 발행일, 저자(이름, 이메일), 기타 부연설명, 참고문헌 태깅



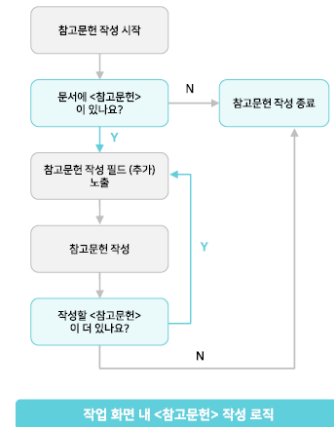
문서별 메타 데이터 구축 작업 화면 예시

```

{
  "doc_type": "BOK이슈노트",
  "doc_no": "제2022-10호",
  "file_path": "BOK이슈노트 제2022-10호 물가상승압력 확산 동향 평가.F.pdf",
  "title": "물가상승압력 확산 동향 평가",
  "publish_datetime": "2022-02-13 00:00:00",
  "authors": [
    {
      "name": "오강현",
      "mail": "khoh@bok.or.kr"
    },
    {
      "name": "김윤경",
      "mail": "Younkyong@bok.or.kr"
    },
    {
      "name": "이재진",
      "mail": "jjlee1121@bok.or.kr"
    },
    {
      "name": "김재훈",
      "mail": "jhkang@bok.or.kr"
    }
  ]
}
```

결과 데이터 예시

- 참고 문헌 작업 시 조건부 로직(Conditional Logic)을 설정하여 예외사항 대응



○ 문서 내용 구조화, 전처리 프로젝트

- 내용 구조화: 문서를 구성하는 요소에 맞게 개요, 초록/머리말, 본문, 참고, 부록으로 구조화하고 해당하는 영역을 태깅하는 작업 수행

```

"preface": "",
"abstract": "본 KEY TAKEAWAYS 독립전자는 유권권이나 정책으로 예시지  
수급이 분리된 개입에 급이전성 제공으로, 중국 등 대외적으로 문제가  
기대하면서 02 국기를 통해 부동하게 적용받은 것으로 예상한다. 02  
경제에서는 이러한 경제상황에 단계에 관련된 여러 가지 측면에서 다시  
문헌의 용지(sick man of Europe)로 전이될 수 있다는 우려를 제기하고 있다...",
"disclaimer": [
  "본 자료의 내용은 한국은행의 공식견해가 아니라 집필자 개인의  
견해라는 점을 밝힙니다."
],
"body": "본 자료는 독립전자의 발행(020000)에 따라 독립전자의 발행을  
최우선으로 목표로 설정하는 등 독립과 조화를 이루고 있다. 또한 020000부터  
2분기 연속 독립전자의 발행에 힘쓰고 있다. 020000부터 2분기 연속  
독립전자의 발행에 힘쓰고 있다. (1) 독립 전자의 발행(020000, 02): 22.4/  
4분기 0.1 - 22.1/4분기 0.1 - 22.1/4분기 0.1, 020000, 020000, 020000, 020000  
020000, 020000, 020000, 020000, 020000, 020000, 020000, 020000, 020000, 020000  
이러한 경제상황에 단계에 관련된 여러 가지 측면에서 다시 독립전자의 발행  
(sick man of Europe)로 전이될 수 있다는 우려를 제기하고 있다...",
"answers": [
  "본 자료는 독립전자의 발행(020000)에 따라 독립전자의 발행을  
최우선으로 목표로 설정하는 등 독립과 조화를 이루고 있다. 또한 020000부터  
2분기 연속 독립전자의 발행에 힘쓰고 있다. 020000부터 2분기 연속  
독립전자의 발행에 힘쓰고 있다. (1) 독립 전자의 발행(020000, 02): 22.4/  
4분기 0.1 - 22.1/4분기 0.1 - 22.1/4분기 0.1, 020000, 020000, 020000, 020000  
020000, 020000, 020000, 020000, 020000, 020000, 020000, 020000, 020000, 020000  
이러한 경제상황에 단계에 관련된 여러 가지 측면에서 다시 독립전자의 발행  
(sick man of Europe)로 전이될 수 있다는 우려를 제기하고 있다...",
  "appendix": ""
]
  
```

결과 데이터 예시

- 각주 처리: 각주 번호마다 각주 내용을 별도 작업하고 데이터 프로세싱을 통해 본문 내 각주 위치에 맞게 각주 내용을 삽입하여 처리

각주 번호마다
각주 내용 별도 작업

각주 작업화면 예시

본문에 각주 내용이 삽입된 결과 데이터 예시

본문(body) 내 각주 위치에
해당하는 각주 내용을 삽입하여 처리

- 문서 내 위계(계층 구조) 처리: 원본 문서의 계층 구조에 맞게 Markdown 형식(#, ##, ###...)으로 위계 처리

원본 문서

계층 구조 작업 예시

계층 구조 Markdown 형식 변환 예시

- 표, 수식, 이미지 처리:
 - 간단한 표는 Markdown editor 를 활용하여 직접 그려서 Markdown 형식으로 작업 수행
 - 문장 내 포함된 간단한 수식은 OCR 을 활용하여 텍스트와 수식을 추출하고 LaTeX 문법이 적용된 Markdown 형식으로 작업 수행

〈표 2〉 표본의 체계적·임상적 모형 MDA

연도구간	A=0	A=3	A=12
2010.1 ~ 2023.9	0.68 (0.85)	0.65 (0.85)	0.72 (0.85)
2016.1 ~ 2020.12	0.65 (0.78)	0.58 (0.67)	0.73 (0.87)
2021.1 ~ 2023.9	0.75 (0.93)	0.78 (1.00)	0.75 (1.00)

주: 1) 연도 구간의 각 기간 내의 일정한 연도(연도) 1) 해당 시점의 데이터를 제공한다.

간단한 표 원본

문장 내 간단한 수식 예시

분류

표본

연도구간

A=0

A=3

A=12

2010.1 ~ 2023.9

2016.1 ~ 2020.12

2021.1 ~ 2023.9

0.68 (0.85)

0.65 (0.85)

0.72 (0.85)

0.65 (0.78)

0.58 (0.67)

0.73 (0.87)

0.75 (0.93)

0.78 (1.00)

0.75 (1.00)

간단한 표 작업 예시

텍스트

표본

연도구간

A=0

A=3

A=12

2010.1 ~ 2023.9

2016.1 ~ 2020.12

2021.1 ~ 2023.9

0.68 (0.85)

0.65 (0.85)

0.72 (0.85)

0.65 (0.78)

0.58 (0.67)

0.73 (0.87)

0.75 (0.93)

0.78 (1.00)

0.75 (1.00)

텍스트를 포함한 간단한 표 결과 데이터 예시

OCR

표본

연도구간

A=0

A=3

A=12

2010.1 ~ 2023.9

2016.1 ~ 2020.12

2021.1 ~ 2023.9

0.68 (0.85)

0.65 (0.85)

0.72 (0.85)

0.65 (0.78)

0.58 (0.67)

0.73 (0.87)

0.75 (0.93)

0.78 (1.00)

0.75 (1.00)

OCR을 활용한 간단한 수식 작업 예시

LaTeX

표본

연도구간

A=0

A=3

A=12

2010.1 ~ 2023.9

2016.1 ~ 2020.12

2021.1 ~ 2023.9

0.68 (0.85)

0.65 (0.85)

0.72 (0.85)

0.65 (0.78)

0.58 (0.67)

0.73 (0.87)

0.75 (0.93)

0.78 (1.00)

0.75 (1.00)

LaTeX 문법이 적용된 간단한 수식 결과 데이터 예시

- 복잡한 표, 복잡한 함수식, 차트 등 이미지는 별도의 이미지 파일(.PNG)로 저장하여 관리함

이미지 처리

표본

연도구간

A=0

A=3

A=12

2010.1 ~ 2023.9

2016.1 ~ 2020.12

2021.1 ~ 2023.9

0.68 (0.85)

0.65 (0.85)

0.72 (0.85)

0.65 (0.78)

0.58 (0.67)

0.73 (0.87)

0.75 (0.93)

0.78 (1.00)

0.75 (1.00)

이미지 처리 작업 예시

이미지 처리 결과

표본

연도구간

A=0

A=3

A=12

2010.1 ~ 2023.9

2016.1 ~ 2020.12

2021.1 ~ 2023.9

0.68 (0.85)

0.65 (0.85)

0.72 (0.85)

0.65 (0.78)

0.58 (0.67)

0.73 (0.87)

0.75 (0.93)

0.78 (1.00)

0.75 (1.00)

이미지 처리 결과 데이터 예시

이미지 처리 결과

표본

연도구간

A=0

A=3

A=12

2010.1 ~ 2023.9

2016.1 ~ 2020.12

2021.1 ~ 2023.9

0.68 (0.85)

0.65 (0.85)

0.72 (0.85)

0.65 (0.78)

0.58 (0.67)

0.73 (0.87)

0.75 (0.93)

0.78 (1.00)

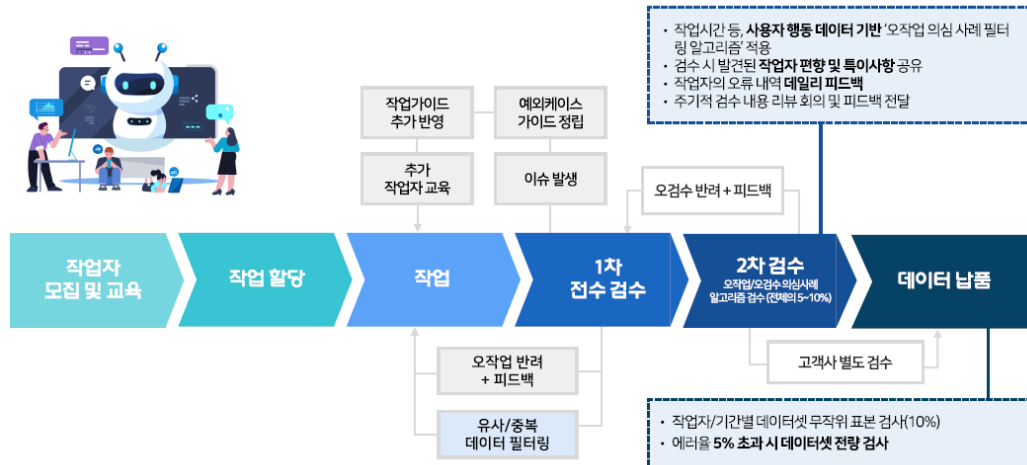
0.75 (1.00)

이미지 처리 결과 데이터 예시

2.6 데이터 품질 관리

○ 데이터 품질 관리 개요

- 1 차 작업물에 대한 전수 검수 수행
- 검수 시 오작업에 대한 반려 + 피드백 수행
- 유사/중복 데이터 필터링 수행
- 전수 검수 시 이슈 발생 상황에 따라 예외 케이스 가이드 정립, 추가 작업자 교육 등 수행 후 재작업 진행
- 1 차 전수 검수 후 2 차 검수(샘플링 자동 모니터링) 수행



○ 2 차 검수(샘플링 자동 모니터링)

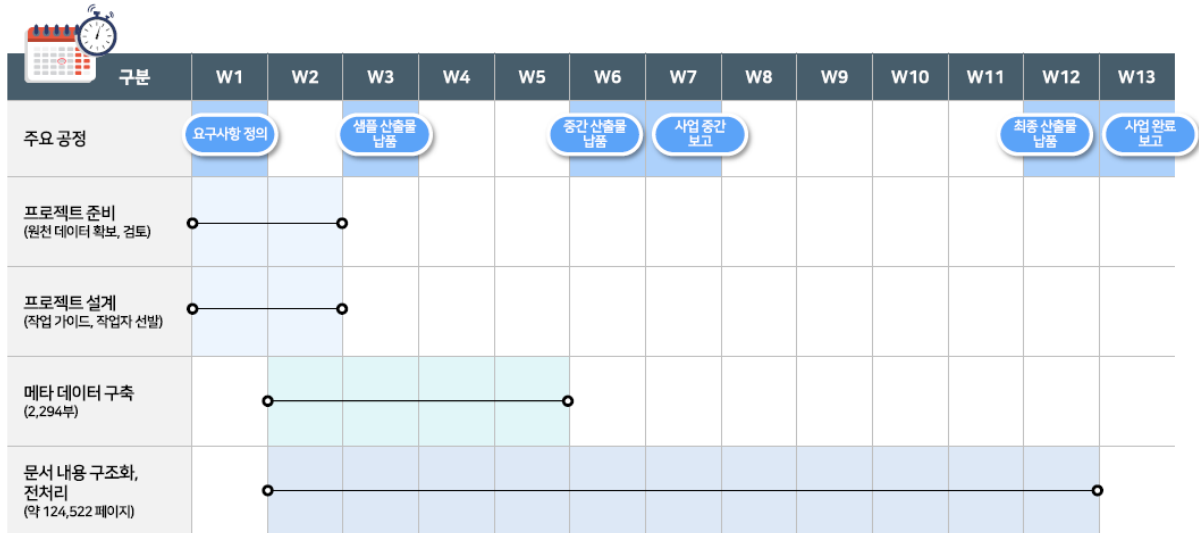
- 전수 검수 완료된 데이터를 대상으로 알고리즘 기반 2 차 샘플링 검수 수행
- 검수 과정 중 발생할 수 있는 오검수 및 불성실 검수에 대한 가설 수립
- 여러 가설들을 바탕으로 상황에 유연한 작업자 행동 기반 데이터 평가 수행

○ 데이터 납품 후 관리

- 데이터 하자 발생 시 발생 건에 대해 5 영업일 이내 보완 후 재납품 수행
- 전체 데이터 수량 대비 에러율 5% 초과 시 데이터 전량 검사 진행

3 추진 일정

3.1 주요 추진 일정



3.2 세부 추진 일정

기간	내용	비고
8/7~8/9	원시 데이터 확보	홈페이지 다운로드
8/7~8/12	작업자 선발, 작업 가이드 수립, 작업 설계	기존 작업자 활용
8/13~9/6	메타 데이터 구축 프로젝트 수행	
8/15~10/25	문서 유형별 전처리 프로젝트 수행 (BOK 이슈노트~)	문서 유형별 난이도 고려하여 주 단위 순차 오픈
8/29	샘플 데이터 납품(BOK 이슈노트)	
9/2~10/30	결과 데이터 납품	프로젝트 진행 상황 고려하여 문서 유형별 순차 납품
9/19	프로젝트 중간 보고	
11/7	프로젝트 결과 보고	

4 프로젝트 관리 지원

4.1 수행 조직



4.2 비상 대책

- 작업 진척률 저하
 - 작업 진척률 개선 시까지 일일 진척률 보고 진행
 - 예비 작업자 및 프로젝트 운영 보조 인력 투입
 - 프로젝트 기간 동안 작업자 상시 모집(예비 작업자 사전 확보)
 - 작업량 인센티브 제도 적용 운영(우수 작업자 대상)
- 데이터 품질 저하
 - 데이터 품질 저하 원인 파악 및 불량 작업자 확인 + 작업 가이드 개선
 - 작업자 교체 및 불량 작업자의 작업물 전량 검수(예비 작업자 및 프로젝트 운영 보조 인력 투입)
 - 자동 모니터링 + 수동 모니터링 동시 진행
 - 불량 작업자 별도 교육 및 교육 수료 여부에 따라 재투입 결정
- 데이터 유실
 - 정기 백업(주간 단위, 월간 단위) 및 비정기 백업 진행