# 인공지능 데이터 구축·활용 가이드라인
# - 도서자료 요약 데이터 -

| | | |
|---|---|---|
| **인공지능 데이터 구축** | 사업 총괄 | 바이브컴퍼니 |
| | 데이터 설계 | 케이티하이텔(주) |
| | 원천데이터 수집 및 정제 | 딥네츄럴 |
| | 데이터 가공 | 딥네츄럴 |
| | 데이터 검수 | 딥네츄럴, 연세대학교, 경북대학교 |
| | 크라우드 소싱 | 딥네츄럴 |
| | 저작도구 개발 | 딥네츄럴 |
| | AI모델 개발 | 케이티하이텔(주) |
| | 응용 서비스 개발 | 포티투마루 |
| **가이드라인 작성** | 케이티하이텔(주) | 홍수안 팀장 |
| **가이드라인 버전** | ver 2.3<br>2021.03.01 | |

# 목 차

# 1. 데이터 명세 정보

## 1.1 데이터 정보 요약

| | | |
|---|---|---|
| 데이터 이름 | 도서자료 요약 데이터 | |
| 활용 분야 | 생성요약 | |
| 데이터 요약 | 인공지능 생성요약 모델 학습을 위한 대규모 한국어 데이터셋 | |
| 데이터 출처 | 국립중앙도서관 제공 정책 관련 도서 | |
| 데이터 이력 | 배포버전 | 1.0 |
| | 개정이력 | 신규 |
| | 작성자/배포자 | 전영민 |

## 1.2 데이터 포맷

본 데이터셋은 텍스트 생성요약 데이터셋이므로 모두 텍스트 형태이다. 수요처의 필요에 따라 xml, txt, json 등 모델 학습 및 배포에 적절한 포맷으로 저장하여 제공한다.


```
{
    "passage_id":"123456_0001",
    "metadata":
        {
            "doc_id":"123456",
            "doc_type":"도서",
            "doc_name":"북미정상회담:창조적 블랙홀이 될 것인가?",
            "author":"정성윤",
            "publisher":"통일연구원",
            "published_year":"2018",
            "kdc_label":"사회과학",
            "kdc_code":"300"
        },
    "chapter":null,
    "passage":"최근 정세 변동의 의미와 평가 북핵 문제는 지난 25 년 동안 다양한 철학과 접근법 그리고 전략의 동원에도 불리지
    않았던 난제 중 난제이다. 그 결과 북핵 문제는 한반도와 동북아의 모든 이슈를 삼키고 가두어 버리는 블랙홀이 되어 버렸다. 그러나
    북핵 문제가 새로운 국면에 진입하고 있다. 결정적 계기는 향후 개최될 남북 정상회담과 북미 정상회 담이 될 것이다. 특히 북핵 위기
    25 년 만에 처음으로 북미 정상이 직접 담판을 하게 됨으로써, 북핵 문제의 획기적 진전에 대한 기대가 높아지고 있다. 두 차례
    정상회담이 합의된 가장 큰 배경은 북한의 태도 변화, 미국의 대화 호응, 우리 정부의 강력한 남북관계 진전 의지와 외교력이다. 이
    중 김정은 스스로가 비핵화 의지를 밝힌 것이 정세 변환의 핵심이자 촉발 동인이다. 이러한 북핵 정세 변화 가 추동하고 있는 구조적
    변화 양상, 이를 가능하게 만든 북한의 전략전환 이유, 북미 정상회담 전후의 정세 방향, 그리고 한국의 정책적 고려사항을 제시한다.",
    "summary":"북핵 문제는 지난 25 년 동안 다양한 노력에도 해결되지 않은 난제로 한반도와 동북아를 모두 아우르는 주요 이슈가
    되었다. 그러나 북핵 문제가 남북 정상회담과 북미 정상회담을 계기로 새로운 양상에 돌입한다. 이는 북한, 미국, 우리 정부의 태도
    변화 및 대화 의지가 바탕이 되었고, 이 중 특히 김정은의 비핵화 의지 표명이 가장 결정적이다. 이러한 구조적 변화와 숨은 의도,
    예상 정세 및 한국의 정책 고려사항을 제시한다."
}
```

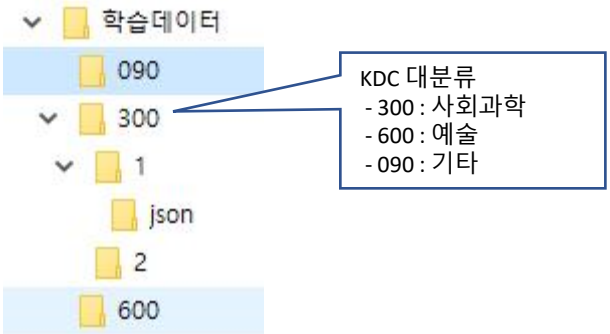

## 1.3 어노테이션 포맷

각 어노테이션 데이터가 지닌 속성은 아래와 같다.

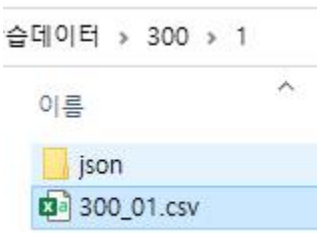
| No. | 속성명 | 영문명 | 길이 | 타입 | 필수여부 | 설명 |
|---|---|---|---|---|---|---|
| 1 | 원문ID | passage_id | 100 | String | Required | 원문에 부여되는 고유번호<br>'문서ID_분리순서' 포맷 |
| 2 | 메타데이터 | metadata | - | Object | Required | 서지 정보에서 추출한 메타데이터 |
| 2-1 | 문서ID | doc_id | 100 | String | Required | |
| 2-2 | 문서유형 | doc_type | 100 | String | Required | '도서' 및 '논문'으로 문서유형 구분 |
| 2-3 | 문서명 | doc_name | 100 | String | Required | |
| 2-4 | 발행자 | author | 100 | String | Optional | |
| 2-5 | 발행처 | publisher | 100 | String | Optional | |
| 2-6 | 발행연도 | published_year | 4 | String | Optional | |
| 2-7 | 주제분류 | kdc_label | 100 | String | Required | 해당 원문의 KDC 분류명 |
| 2-8 | 분류기호 | kdc_code | 3 | String | Optional | 해당 원문의 KDC 분류코드 |
| 3 | 챕터 | chapter | 100 | String | Optional | 해당 원문이 소속된 챕터명 |
| 4 | 원문 | passage | 1000 | String | Required | 구축 대상 원문 문단 |
| 5 | 요약문 | summary | 300 | String | Required | 원문 문단에 대한 생성요약 |

## 1.4 데이터 구성

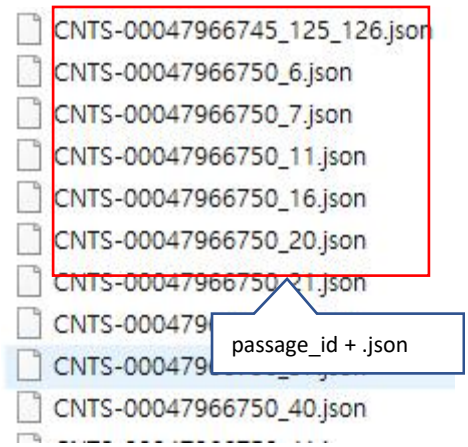
데이터 저장소에 저장되는 데이터의 구성은 아래와 같이 진행된다. (협의 후 변경가능)
학습데이터 아래에는 KDC 대분류 기준으로 폴더가 생성이 되며, 학습데이터가 10만개가 넘는 경우, 추가적으로 하위 폴더를 생성한다.




학습이터는 10만개 단위로 json 파일의 폴더와 csv 파일의 형식으로 구성된다. csv 파일의 경우 너무 크지 않게 일정 규격으로 잘라서 업데이트 한다.




json 폴더의 하위에는 passage_id + .json의 규칙을 가진 json 파일이 포함된다.




## 1.5 데이터 통계

### 1.5.1 데이터 구축 규모

본 데이터 구축 사업의 최종 산출 데이터는 텍스트 형식의 문단(각 300-1000자) 20만 건과 각 문단 별 요약문 20만 건이다. 각 문단과 이를 요약한 요약문은 한쌍으로 모델에 학습되며, 총 20만 쌍의 원문 문단-요약문 구축을 목표로 한다. 이는 JSON 파일 1건으로 용량은 약 2~3KB이다.

### 1.5.2 데이터 분포

수급이 완료된 데이터의 주제별 권수는 KDC(한국십진분류법, Korean Decimal Classification) 기준으로 아래와 같다. 다만, 권의 개수가 적은 자연과학, 역사, 어학, 문학, 철학, 종교, 순수과학, 언어등은 기타의 항목과 합하여 표기한다.
학습데이터 원문의 비율은 1권에서 발췌할 수 있는 원문의 개수가 다를 수 있으나, 최대한 수집데이터의 비율도 아래의 비율에 맞출수 있도록 가공한다.

| 대분류 | 사회과학 | 기술과학 | 예술 | 기타 | 합계 |
|---|---|---|---|---|---|
| 권수 | 144,489 | 29,707 | 17,363 | 8,441 | 200,000 |
| 비율(%) | 72.25 | 14.85 | 8.68 | 4.22 | 100 |

## 1.6 원시데이터 특성

### 1.6.1 대상분류

구축 대상 원시데이터는 국립중앙도서관이 보유한 정책 관련 도서로 '실제' 데이터에 해당한다.

### 1.6.2 제약조건

수집 대상 원시데이터는 저작권 문제가 해결된 것으로 제한한다. 도서자료 중 시나 소설과 같이 내용에 해석에 차이가 있고 짧은 콘텐츠는 배제한다. 책 전체를 요약하는 것이 아닌 단락별로 내용을 요약한 데이터셋을 구축하며, 단락의 길이는 500자 이상 1000자 이내로 한정된 '제약있음(constrained)' 데이터에 해당한다.

### 1.6.3 속성

원시데이터는 플레인 텍스트(plain text) 형태로, 요약에 부적절한 이미지, 도표 등은 제거된 형식이다.

## 1.7 기타정보

### 1.7.1 포괄성

요약에 적절한 형식의 원문을 확보하기 위해 국립중앙도서관 보유 정책자료로 한정하나, 내용 측면에서는 보건사회, 생명, 조세, 환경, 지역사회 개발, 무역, 경제, 노동 등 다양한 분야에서 수집하여 포괄성을 최대한 확보한다.

### 1.7.2 독립성

공공에 출간된 도서를 제공받음으로써 개인정보 등 민감한 정보가 포함되어 있지 않은 데이터를 수급받을 수 있으며, 국립중앙도서관과의 협약을 통해 저작권이 소멸된 도서를 수급함으로써 저작권 문제를 해결한다.

### 1.7.3 유의사항

본 데이터는 유일한 한국어 도서 대상 대규모 공개 생성요약 데이터로 민간 및 공공에서 자유롭게 사용될 수 있도록 공개되어, 추후 우리나라의 생성요약 연구 및 서비스 발전에 기여할 것으로 예상된다.

다만, 본 데이터는 생성요약 모델링에 적합한 데이터로 추출요약 알고리즘을 구성하고 학습시키는 데는 적절하지 않다. 또한 구축 대상 원시데이터가 500자 이상 1000자 이하이므로, 이보다 짧거나 긴 문서를 요약하는 모델을 학습시키는 데는 유의해야 한다.

# 2. 데이터 구축 가이드

## 2.1 데이터 구축 개요

텍스트 정보가 PC 및 스마트폰과 같은 문자에서 스마트스피커, 커넥티드카 인포테인먼트, VR 등 다양한 기기로 확장되고 전자책 시장이 성장함에 따라 많은 양의 텍스트에서 빠르게 정보를 추출하여 전달할 수 있는 요약 서비스에 대한 수요가 증가하고 있다. 생성요약 데이터셋 및 모델이 민간 기업 및 일반 개인에게 공개된다면, 일반 대중들이 이를 활용한 서비스를 이용하며 부가가치가 높은 업무에 집중함으로써 사회경제적 발전을 이룩할 수 있으리라 기대된다. 국가 차원에서는 4차 산업혁명의 핵심인 인공지능 분야에서 한국어 모델을 확보하는 것이 중요한 시점이다. 이에 한국어 생성요약 모델을 구축하기 위한 대규모 데이터의 필요성도 증대되었고, 본 가이드라인은 본 사업을 통해 구축될 데이터 및 추후 새롭게 구축될 데이터셋에 대한 작업공정과 상세 지침을 제공하고자 한다.

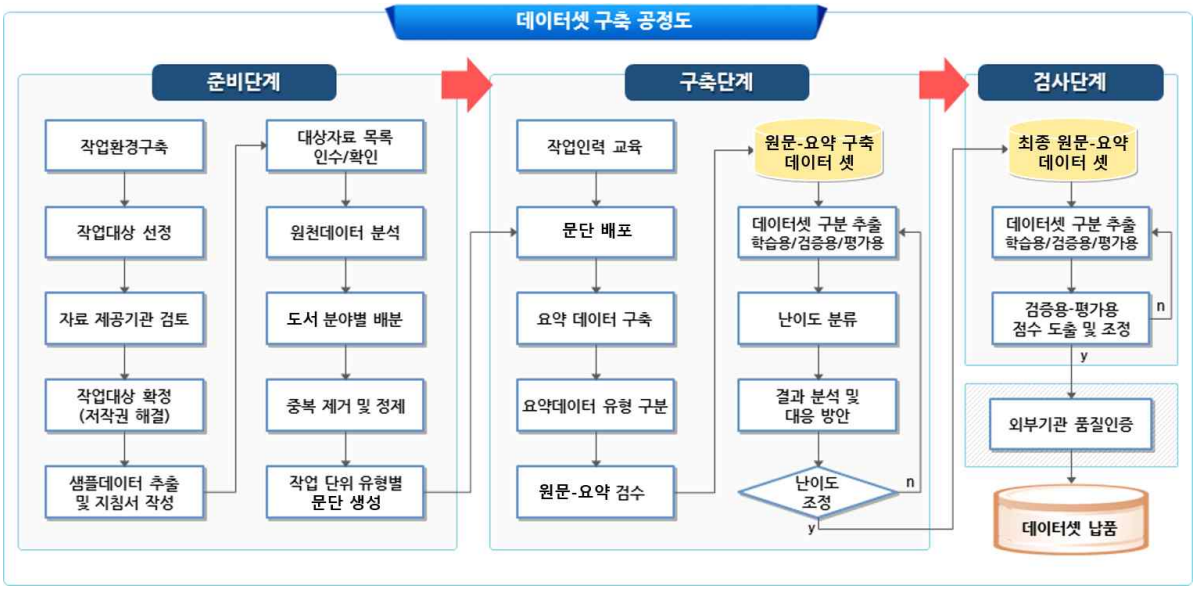
본 데이터셋은 국립중앙도서관 제공 도서 약 5만권을 디지털화하여 문서 요약 분야의 기초 자료로 활용할 수 있는 **도서자료 데이터 20만건 이상** 구축 완료를 목표로 하며, 구축 데이터에 대해서는 자유로운 배포와 이용이 가능하도록 저작권을 해결한다.

데이터셋을 구축하기 위한 프로세스는 아래와 같이 데이터 획득·정제(준비), 구축, 검수(검사) 단계로 이루어진다.

준비단계에서는 실제 데이터 구축 작업에 앞서 필요한 요소들을 구비한다. 구축 작업을 진행하기에 앞서 구축 도구 등 작업환경을 구축하고 작업인력을 채용하며, 데이터 분석 및 정제를 통해 작업에 필요한 원시데이터를 확보한다.

구축단계에서는 실제로 데이터 구축 작업을 진행한다. 구축 작업에 대한 작업인력 교육을 실시한 후 실제 데이터 구축에 돌입한다. 이때 검수담당 인력이 구축된 요약문을 수시로 모니터링하여 데이터의 수량과 품질을 유지한다.

검사단계에서는 구축된 전체 데이터를 기존에 정의한 방법에 따라 검증한다. 먼저 데이터셋을 학습용/검증용/평가용으루 구분하여 각 목적에 맞게 활용할 수 있게 한다. 학습을 통한 모델 점수 및 외부기관 검수를 통해 데이터셋이 목표 수량과 품질을 달성했는지 확인한 후 최종적으로 납품한다.




## 2.2 문제정의

### 2.2.1 임무 정의

본 데이터 구축 및 모델 개발 사업은 생성요약이라는 자연어처리 분야의 근간을 구축함으로써 4차산업 시대의 혁신 동력을 제공하는 데 가장 큰 목적이 있다. 현재 시장에서 볼 수 있는 요약 서비스는 본문에서 중요한 내용을 그대로 가져오는 추출요약(Extractive Summarization) 기술을 적용한 것인데, 본 사업은 본문에서 중요한 문장을 하나의 새로운 요약문으로 재창조하는 생성요약(Abstractive Summarization)을 위한 데이터셋을 구축하고 이를 실제 모델에 학습시키는 데 더 큰 의의가 있다. 추출요약에 비해 생성요약은 난이도가 훨씬 더 높은 과제인 대신 그만큼 더 자연스럽고 다양한 형태로 적용이 가능하다는 강점이 있다. 즉, 생성요약에 대한 필요성은 점점 대두되고 있는 만큼 본 사업을 통해 대규모 데이터를 구축하기 어려운 민간 기업 및 일반 개인까지 데이터를 활용할 수 있도록 공개하고 여러 산업 분야에 적용하도록 도모하여 한국어 생성요약 분야가 발전하도록 기여할 것이다. 일반 대중은 이를 활용한 서비스를 이용함으로써 정보 탐색에 소요되는 시간을 절약하고 생산성 향상을 누릴 수 있을 것으로 기대된다. 이처럼 부가가치가 높은 업무에 집중함으로써 사회경제적으로도 발전하는 데도 크게 기여할 것이다. 국가 차원에서는 아직 영문으로도 발전이 더딘 생성요약 분야에서 한국어 모델을 정부 주도로 개발함으

로써 세계적으로 4차 산업을 주도한 선례를 마련하게 될 것이다.

### 2.2.2 데이터 구축 유의사항

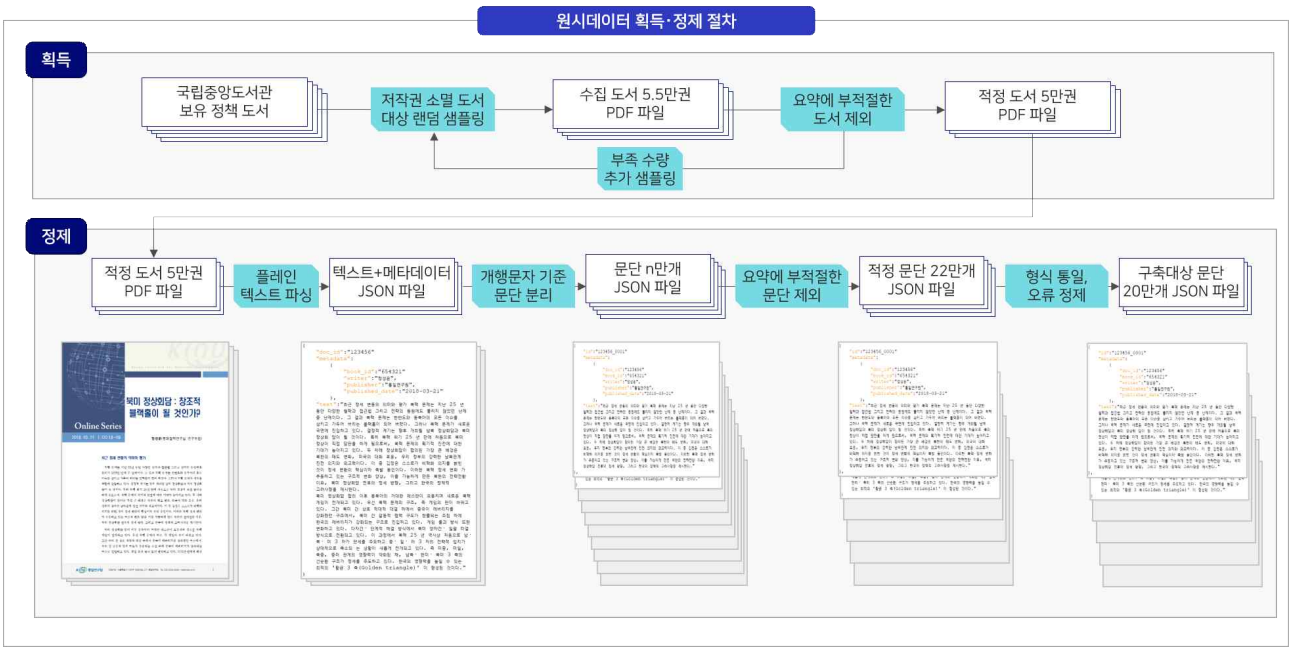
데이터 획득 시 대상 데이터는 도서자료를 대상으로 한 텍스트 데이터로 초상권에 대한 문제는 부재하며, 국립중앙도서관의 협조를 통해 저작권 소멸된 데이터를 전달받아 저작권 문제를 해결한다. 또한 데이터 구축 전 데이터 제공 및 활용에 따른 법률적 문제점 검토를 진행하여 향후에 발생 가능한 법률 문제에 대비한다. 내용 측면에서는 개인의 의견이 담긴 도서가 아닌 대중에게 공개된 정책 관련 문서를 대상으로 하여 사회적으로 민감한 정보는 배제한다.
정제 시에는 명확한 기준을 제시하여 어노테이션에 적절한 원시데이터만을 선별하고, 사후 오류보정을 통해 어노테이션이 원만하도록 준비한다. 어노테이션 작업 시에는 작업자 및 검수자를 대상으로 구축 방식에 대한 교육을 진행하고, 구축 및 검수 작업이 원활하도록 어노테이션 툴을 지속적으로 관리한다.

## 2.3 획득·정제

### 2.3.1 원시데이터 선정

원시데이터는 국립중앙도서관과의 협조하에 제공받은 저작권 문제가 없는 도서자료를 대상으로 하며, 이는 다양한 형식의 비정형 텍스트 데이터이다. 유형 면에서는 문학과 같이 해석이 명확하지 않고 길이가 짧은 자료는 제외하며, 정책자료, 연구문 등 객관적인 해석이 가능한 자료를 선정하여 요약 모델에 대한 의미있는 평가를 진행한다. 내용 면에서는 보건사회, 생명, 조세, 환경, 지역사회 개발, 무역, 경제, 노동 등 다양한 분야에서 수급하여 특정 분야에 편향되지 않은 데이터를 확보한다.
국립중앙도서관에서 수급가능한 도서의 수량이 목표 수량인 5만권보다 적은 경우 출판사 등과 협약하여 저작권이 소멸된 도서를 추가로 확보한다.

### 2.3.2 획득·정제 절차




국립중앙도서관에서 충분한 양의 저작권 소멸 도서가 확보된 경우 이 중 목표수량인 5만권의 약 1.1배인 5.5만권을 임의로 샘플링하여 수집한다. 수집한 5.5만권의 파일을 분석하여 요약에 부적절한 도서를 제외하여 총 5만권의 도서를 확보한다. 또한 서지정보를 기준으로 제목, 발행자, 발행처, 발행년도 등 모든 정보가 동일한 경우 중복된 도서로 파악하여 구축 대상에서 제외한다. 요약에 부합하는 도서가 5만권 미만인 경우, 추가로 부족한 수량의 1.1배를 다시 샘플링하여 수집하고, 이와 같은 방식으로 적정도서 5만권을 최종 확보한다.
목표한 수량의 도서를 획득한 후에는 구축을 위한 형태로 정제하는데, 가장 먼저 PDF파일을 분석과 요약이 용이하도록 플레인 텍스트 형식으로 파싱하여 메타데이터와 함께 JSON 형태로 저장한다. 텍스트 형태로 변환 후에는 요약 작업을 수월하게 진행할 수 있도록 문단 단위로 분리하여 각각을 JSON 형태로 분리해 저장한다. 분리된 문단 중에 요약에 부적절한 문단의 경우 구축 대상에서 제외하여 총 22만개의 문단을 선정하고, 최종적으로 형식 통일, 파싱 오류 정제 등을 통해 구축에 투입될 수 있는 20만개의 문단으로 확보한다.

### 2.3.3 획득·정제 기준

원시데이터 획득 시에는 편향 방지를 위해 국립중앙도서관에 저작권 소멸된 정책자료의 메타데이터에서 확인할 수 있는 정보를 활용하여 분야, 발행기관별 보유 현황을 파악 후 균등한 비율로 대상 도서자료를 무작위 샘플링한다.
도서를 선정한 후에는 각 도서의 형식과 내용이 요약에 적절한지 아래와 같은 기준에 따라 판단하여 최종 정제 대상 도서를 선정하며, 부족한 경우 동일한 분야에서 부족한 수량만큼 다시 랜덤 샘플링하여 보충한다.

**※ 대상 도서 선정 기준**

| | |
|---|---|
| 적절 | - 산문으로 문단 구분이 가능한 문서 (예: 연구, 보고서, 연구원 기고) |
| 부적절 | - 이미지나 표가 지나치게 많음 (예: PPT, 사업계획서)<br>- 산문이 아닌 짧은 문장으로 구조화 (예: 보도자료, 운용계획, 회의록, 사실 나열)<br>- 국문 외에 한문, 영문이 지나치게 많이 혼용되어 의미 파악 어려움 |

선정한 데이터는 효율적인 정제와 가공을 위해 플레인 텍스트(plain text) 형태로 파싱하며, 함께 저장되어 있는 서지 정보(서명, 저자, 발행기관, 발행일, 주제분야 등)을 메타데이터로 저장하여 관리가 용이하도록 한다. 이때 인코딩은 UTF-8 인코딩으로 통일한다. 텍스트 변환이 완료되면 요약문을 작성할 수 있도록 좀 더 작은 단위의 텍스트로 분리해야 하는데, 개행문자와 개행문자 사이의 한 단락을 하나의 지문으로 생성한다. 이때 분리된 순서대로 '출처도서의 id_추출순번' 형태로 시리얼 넘버를 부여해 손쉽게 출처 도서를 특정하고 데이터를 관리할 수 있도록 한다.
도서 내에 존재하는 제목/소제목 등은 실제 구축 대상 문단에서는 제거하며, 메타데이터의 "chapter" 항목으로 저장한다. 메타데이터는 실제로 모델에 학습되는 데이터는 아니므로 기호를 정제할 필요는 없으나, 동일한 도서 내에서는 챕터 별로 동일한 글머리 기호를 사용하도록 유의한다.
문단 분리가 완료되면 아래와 같은 기준으로 요약에 적절한 문단만 취하고 적절하지 않은 문단은 제외하여 최종 구축 대상 문단을 선정한다. 이때 머리말, 추천사, 책머리에 등 원문 초입에 등장하는 글 및 각주, 인용문, 예시 등은 그 자체로 아래 기준을 충족하는 경우 요약 대상 원문으로 포함한다. 대상도서가 총 5만권임을 고려하여 도서당 약 4-5개의 문단을 선정하되, 문서의 길이가 짧아 목표 개수보다 적은 경우 정제된 모든 문단을 취한다. 또한 도서당 약 4-5개의 문단을 선정했음에도 목표 원문 수량에 도달하지 않은 경우 이보다 많은 문단을 원문으로 포함해도 무방하다.

**※ 요약 대상 문단 선정 기준**
1. 공백을 포함한 글자수가 최소 300자에서 최대 1000자인 문단만 포함
2. 제거된 이미지/표/인용문 없이 내용 파악이 불가능한 경우 제외
3. 선행 내용 없이 원문 단독으로 내용 파악이 불가한 경우 제외

최종적으로 선정된 문단은 아래와 같이 다양한 형식을 통일하고 부적합한 형식 및 오류 사항을 보정하여 원활하게 원문 문단을 이해하고 요약문을 작성할 수 있도록 정제한다.

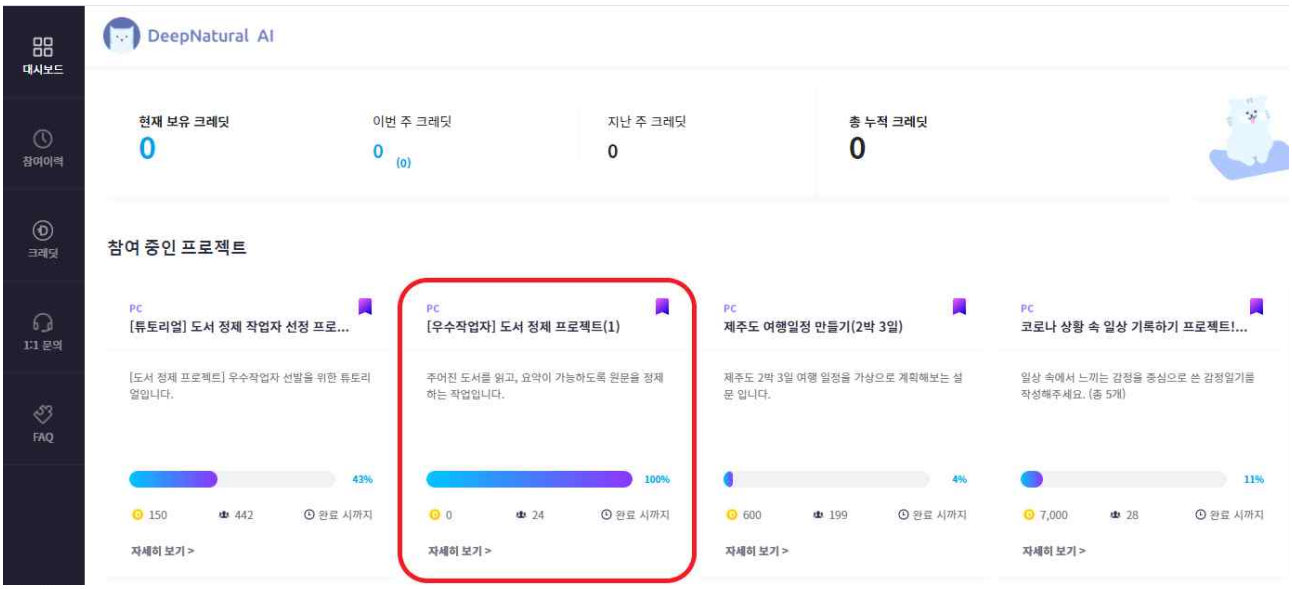
**※ 형식 통일 및 오류 정제 기준**
1. 분절 오류 단락 삭제 및 수정
예) (전략) 그 제품의 교역 및 운송체계를 유지함으로써 달성할 > 달성할 수 있다.
2. 고유명사 처리 기준 미달 문장 삭제 (개인정보 등)
예) 01*-****-****, ****@****.com > 삭제
3. 심각한 비문 삭제 및 수정
예) 제시된 환경영향 평가방법과 평가지침이 체계적으로 시행할 수 있는 제품의 환경영향 평가제도를 제안한다. > 제시된 환경영향 평가방법과 평가지침을 체계적으로 시행할 수 있는 제품의 환경영향 평가제도를 제안한다.
4. 오타, 인코딩 오류 등 교정
예) '상하(上下)로/유엔(UN)에'가 '상하(??)로/유엔(??)에'로 인코딩 > '상하로/유엔에'로 괄호 내용 제거
예) '上下로/UN에'가 '??로/??에'로 인코딩 > '상하로/유엔에' 또는 '上下로/UN에'로 기존 한자/영어 또는 한국어 음독으로 교정

### *2.3.4* 획득·정제 조직

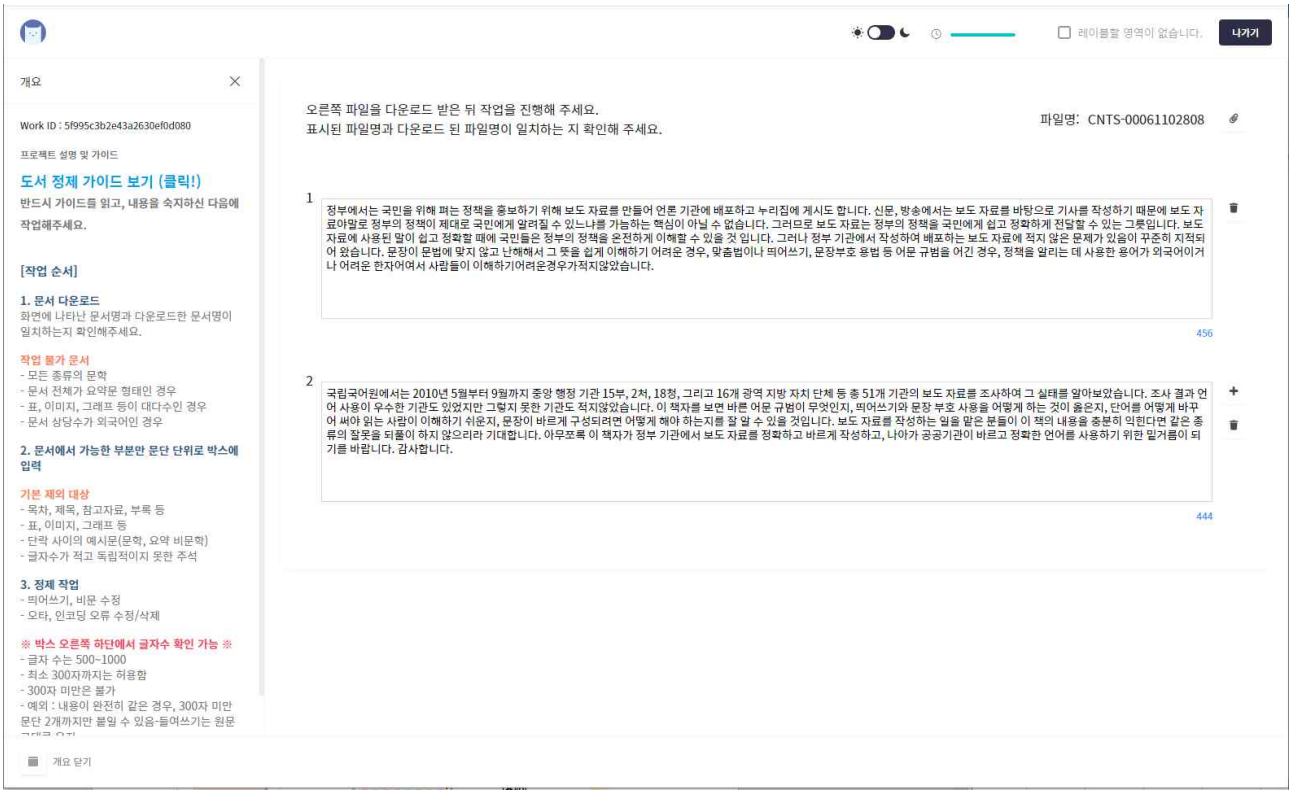
정제를 위한 조직은 다수의 크라우드 작업자로 구성되며, 이중 검수는 검수 관련 교육을 받은 소수의 크라우드 작업자 위주로 구성이 된다.

### *2.3.5* 획득·정제 도구

크라우드 작업자가 프로젝트에 참여할 경우 아래와 같은 웹기반 크라우드 작업 플랫폼에서 작업을 한다.



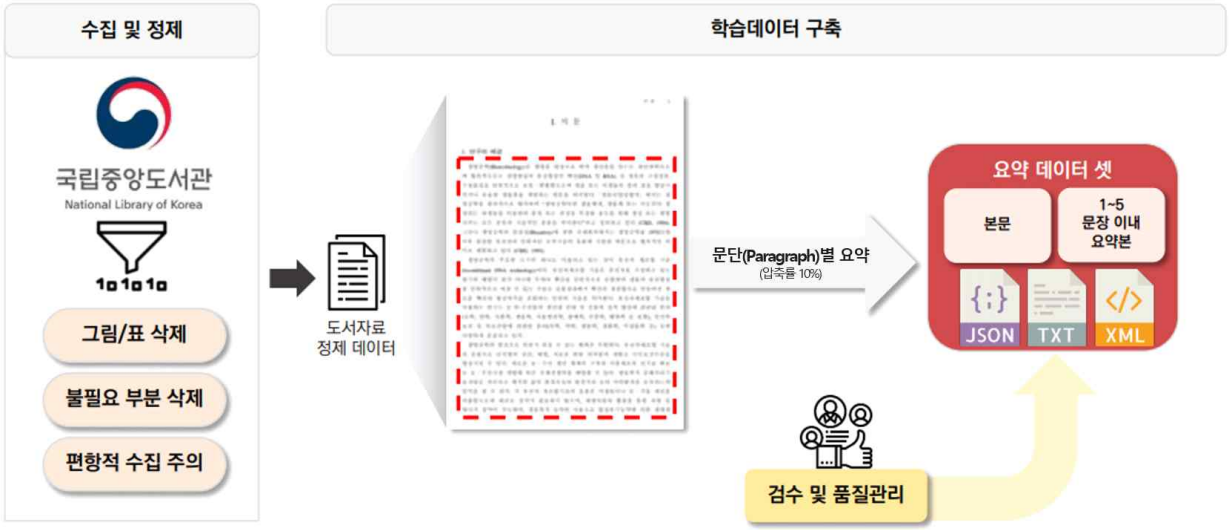

크라우드 작업자는 해당 프로젝트를 선택하고, 프로젝트에 진입한 뒤, 작업 대상 문서를 다운로드 받은뒤, 사이트내에 링크된 '도서 정제 가이드'에 맞게 정제 작업을 수행한다.




## 2.4 어노테이션/라벨링

### 2.4.1 어노테이션/라벨링 절차

문단 단위로 정제한 원문 데이터를 구축툴에 배포하고 작업자에게 구축툴 권한 및 작업을 할당하여 요약 데이터를 구축한다. 작업자는 1문장 이상 5문장 이내로 생성 요약문을 작성하여 데이터셋을 구축하며, 관리자는 수시로 구축 현황을 모니터링하며 주기적으로 작업물 검수 후 피드백을 제공한다. 데이터는 검수 및 품질 관리를 거쳐 최종 생성요약 학습데이터로 등록한다. 결과물은 범용적인 데이터 포맷인 JSON으로 도출하여 누구나 쉽게 사용 가능한 형태로 제공한다.




## 2.4.2 어노테이션/라벨링 기준

생성요약 학습용 데이터는 원문 1개당 1개의 문단으로 이루어진 요약문 작성하며, 각 요약문은 원문의 길이 및 내용에 따라 1문장 이상 5문장 이내로 원문 대비 10-30% 압축률로 작성한다. (예: 원문이 1,000자인 경우 100-300자로 작성)
요약문을 작성하는 순서는 아래와 같다.

| 단계 | 내용 |
|---|---|
| ① 원문 | 생명공학기술, 특히 유전자 재조합 기술에 의하여 창출되는 LMOs 및 그 제품에 의한 인간의 건강 및 자연생태계의 비가역적 파괴를 사전에 방지할 수 있는 추진 방향으로 실험실에서의 안전한 생명공학기술의 적용과 LMOs 및 그 제품의 안전한 환경도입단계로 구분하여 설정할 수 있다. 실험실에서 안전한 생명공학기술의 적용을 위해서는 유전자재조합 기술에 의하여 LMOs가 창출되고 연구되는 실험실에서의 안전성이 유지되어야한다. LMOs의 환경도입단계는 창출된 LMOs 및 그 제품의 연구과정으로서 LMOs의 환경적응능력과 연구자가 원하는 LMOs의 능력발현을 검정할 수 있는 환경도입실험단계, LMOs 및 그 제품의 국내·외간 교역단계와 운송단계로 구분할 수 있는데, 생명공학기술의 안전성은 각 단계별로 구분하여 안전한 생명공학기술의 실험, LMOs 및 그 제품의교역 및 운송체계를 유지함으로써 달성할 수 있다.<br>본 연구의 목적은 생명공학기술의 응용에 대한 자연생태계의 안전성 유지체계에서 LMOs 및 그 제품에 의한 자연생태계의 비가역적 파괴를 사전에 방지할 수 있도록 LMOs 및 그 제품의 환경도입 이전에 LMOs 및 그 제품이 우리나라의 자연 생태계에미칠 수 있는 영향을 사전에 파악할 수 있는 LMOs 및 그 제품의 환경영향 평가방법 및 평가(기술)지침, 이를 체계적으로 시행할 수 있는 평가제도를 제안하는 데 있다.<br>특히, LMOs 및 그 제품의 환경영향 평가방법은 UN기관 및 선진외국에서 연구된 생명공학안정성에 대한 평가방법이 체계적으로 분석되어 우리나라의 현실에 적용할 수 있도록 제시하는 데 있다. 즉, LMOs 및 그 제품에 의한 우리나라 생태계의 비가역적인 훼손방지 및 이에 따른 국민들의 불안감을 해소하고 대외적으로 우리나라 생명공학산업의 국제경쟁력 확보와 세계시장으로 진출할 수 있도록 양면을 고려하여 현실적으로 적용할 수있는 LMOs 및 그 제품 환경영향 평가방법과 평가지침을 제안하는 데 있다. 또한, 제시된 LMOs 및 그 제품의 환경영향 평가방법과 평가지침을 체계적으로 시행할 수 있는LMOs 및 그 제품의 환경영향 평가제도를 제안하는 데 있다. |

| | |
|---|---|
| ② 주요문장 선별 | 생명공학기술, 특히 유전자 재조합 기술에 의하여 창출되는 LMOs 및 그 제품에 의한 인간의 건강 및 자연생태계의 비가역적 파괴를 사전에 방지할 수 있는 추진 방향으로 실험실에서의 안전한 생명공학기술의 적용과 LMOs 및 그 제품의 안전한 환경도입단계로 구분하여 설정할 수 있다. ~~실험실에서 안전한 생명공학기술의 적용을 위해서는 유전자재조합 기술에 의하여 LMOs가 창출되고 연구되는 실험실에서의 안전성이 유지되어야한다.~~ LMOs의 환경도입단계는 창출된 LMOs 및 그 제품의 연구과정으로서 LMOs의 환경적응능력과 연구자가 원하는 LMOs의 능력발현을 검정할 수 있는 환경도입실험단계, LMOs 및 그 제품의 국내·외간 교역단계와 운송단계로 구분할 수 있는데, ~~생명공학기술의 안전성은 각 단계별로 구분하여 안전한 생명공학기술의 실험, LMOs 및 그 제품의교역 및 운송체계를 유지함으로써 달성할 수 있다.~~<br>본 연구의 목적은 생명공학기술의 응용에 대한 자연생태계의 안전성 유지체계에서 LMOs 및 그 제품에 의한 자연생태계의 비가역적 파괴를 사전에 방지할 수 있도록 LMOs 및 그 제품의 환경도입 이전에 LMOs 및 그 제품이 우리나라의 자연 생태계에 미칠 수 있는 영향을 사전에 파악할 수 있는 LMOs 및 그 제품의 환경 영향 평가방법 및 평가(기술)지침, 이를 체계적으로 시행할 수 있는 평가제도를 제안하는 데 있다.<br>특히, LMOs 및 그 제품의 환경영향 평가방법은 UN기관 및 선진외국에서 연구된 생명공학안정성에 대한 평가방법이 체계적으로 분석되어 우리나라의 현실에 적용할 수 있도록 제시하는 데 있다. ~~즉, LMOs 및 그 제품에 의한 우리나라 생태계의 비가역적인 훼손방지 및 이에 따른 국민들의 불안감을 해소하고 대외적으로 우리나라 생명공학산업의 국제경쟁력 확보와 세계시장으로 진출할 수 있도록 양면을 고려하여 현실적으로 적용할 수있는 LMOs 및 그 제품 환경영향 평가방법과 평가지침을 제안하는 데 있다. 또한, 제시된 LMOs 및 그 제품의 환경영향 평가방법과 평가지침을 체계적으로 시행할 수 있는LMOs 및 그 제품의 환경영향 평가제도를 제안하는 데 있다.~~ |
| ③ 불필요한 수식어구 제거 | ~~생명공학기술, 특히~~ 유전자 재조합 기술에 의하여 창출되는 ~~LMOs 및 그 제품에 의한 인간의 건강 및 자연생태계의~~ 비가역적 파괴를 사전에 방지할 수 있는 추진 방향으로 실험실에서의 안전한 ~~생명공학~~기술~~의~~ 적용과 ~~LMOs 및 그~~ 제품의 안전한 환경도입단계로 구분하여 설정할 수 있다. ~~LMOs의~~ 환경도입단계는 ~~창출된 LMOs 및 그 제품의 연구과정으로서 LMOs의 환경적응능력과 연구자가 원하는 LMOs의 능력발현을 검정할 수 있는~~ 환경도입실험단계, ~~LMOs 및 그 제품의~~ 국내·외간 교역단계~~와~~ 운송단계로 구분할 수 있는 데,<br>본 연구의 목적은 ~~생명공학기술의 응용에 대한~~ 자연생태계~~의~~ 안전성 유지체계에서 ~~LMOs 및 그 제품에 의한 자연생태계의 비가역적 파괴를 사전에 방지할 수 있도록 LMOs 및 그 제품의 환경도입 이전에 LMOs 및 그 제품이 우리나라의 자연 생태계에미칠 수 있는 영향을 사전에 파악할 수 있는 LMOs 및 그~~ 제품의 환경영향 평가방법 ~~및~~ 평가~~(기술)~~지침, ~~아를 체계적으로 시행할 수 있는~~ 평가제도를 제안하는 데 있다.<br>특히; ~~LMOs 및 그 제품의~~ 환경영향 평가방법은 UN기관 및 선진외국에서 연구된 생명공학안정성에 대한 평가방법이 체계적으로 분석되어 우리나라~~의~~ 현실에 적용할 수 있도록 제시하는 데 있다. |
| ④ 간략한 표현으로 축약 및 유사 표현으로 변경 | 유전자 재조합 기술에 의하여 ~~창출되는 비가역적 파괴를 사전에 방지할 수 있는~~*건전한* 추진 방향으로 실험실에서의 안전한 기술 적용과 제품의 안전한 환경도입단계로 ~~구분하여 설정할 수 있다.~~*구분할 수 있다.* 환경도입단계는 환경도입실험단계, 국내·외간 교역단계, 운송단계로 ~~구분할 수 있는 데,~~*나누어진다.*<br>본 연구의 목적은 자연생태계 안전성 유지체계에서 제품의 환경 영향 평가방법, 평가지침, 평가제도를 제안하는 데 있다.<br>특히 환경영향 평가방법은 ~~UN기관 및 선진외국에서 연구된 생명공학안정성에 대한 평가방법이 체계적으로 분석되어~~*외부 평가방법 분석 후* 우리나라 현실에 ~~적용할 수 있도록 제시하는 데~~*적용하는 데* 있다. |
| ⑤ 자연스럽게 연결되도록 수정 | 유전자 재조합 기술~~에 의하여~~*의* 건전한 추진 방향~~으로~~*은* 실험실에서의 안전한 기술 적용과 제품의 안전한 환경도입단계로 구분할 수 ~~있다.~~*있으며,* 환경도입단계는 *다시* 환경도입실험단계, 국내외간 교역단계, 운송단계로 나누어진다. 본 연구~~의 목적은~~*는* 자연생태계 안전성 유지체계~~에서~~*를 위한* 제품의 환경 영향 평가방법, 평가지침, 평가제도를 제안~~하는 데 있다.~~*을 목적으로 한다.* 특히 환경영향 평가방법은 외부 평가방법 분석 후 우리나라 현실에 적용하는 데 *목적이* 있다. |
| ⑥ 요약문 완성 | 유전자 재조합 기술의 건전한 추진 방향은 실험실에서의 안전한 기술 적용과 제품의 안전한 환경도입단계로 구분할 수 있으며, 환경도입단계는 다시 환경도입실험단계, 국내외간 교역단계, 운송단계로 나누어진다. 본 연구는 자연생태계 안정성 유지체계를 위한 제품의 환경 영향 평가방법, 평가지침, 평가제도 제안을 목적으로 한다. 특히 환경영향 평가방법은 외부 평가방법 분석 후 우리나라 현실에 적용하는 데 목적이 있다. |

'② 주요문장 선별' 시 주요문장과 부가문장을 구분하는 대한 객관적인 기준은 없으나 아래와 같은 경향성을 보인다. 다만 모든 구축 대상 문단에 해당하는 내용은 아니므로 반드시 작업자가 문단 전체 내용을 파악한 뒤 주요문장을 선별해야 하도록 주의를 기한다.

**※ 주요/부연문장의 경향성**

| | |
|---|---|
| 주요문장 | - 결론, 근거 등 핵심논조를 설명하는 문장: 따라서, 그러므로, 왜냐하면 등으로 시작<br>- 주로 첫 번째 또는 마지막 문장 |
| 부가문장 | - 사례를 설명하는 문장: 예를 들어, 일례로, 구체적으로는 등으로 시작<br>- 의문문: ~은 왜 그런 것일까? ~은 무엇일까? 등 |

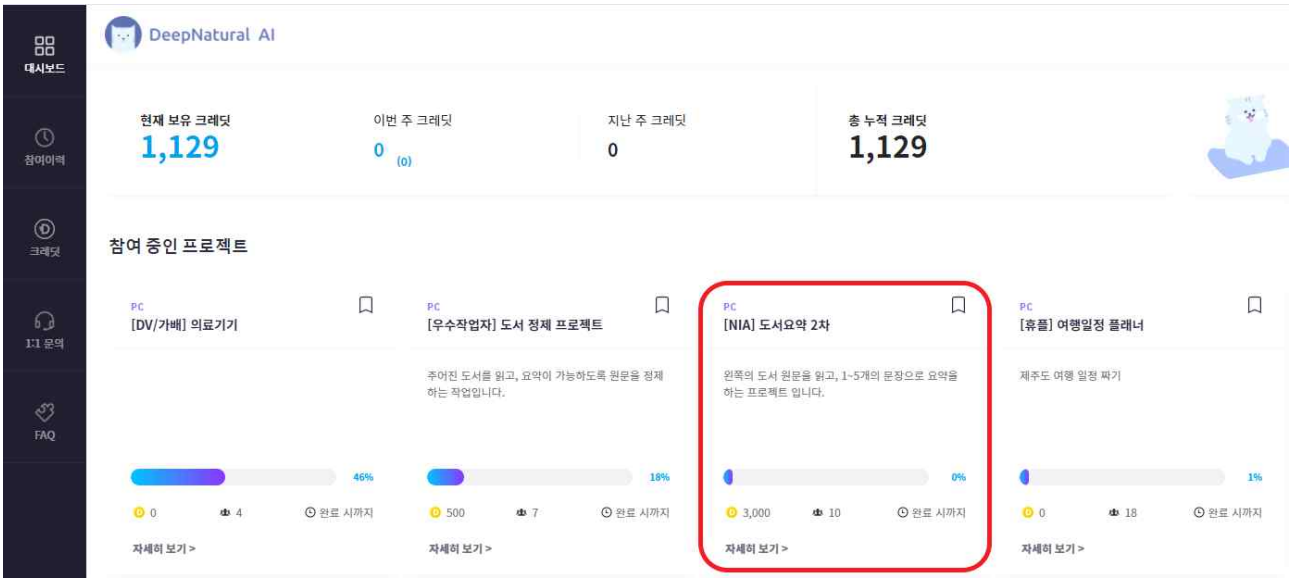
'④ 간략한 표현으로 축약 및 유사 표현으로 변경' 시 문장부호는 본문에 있는 부호를 그대로 사용한다. 또한 본문의 모든 단어를 작위적으로 변경할 필요는 없으며, 원문보다 더 간단하고 이해하기 쉬운 수준의 문장이 되도록 작성한다.

### *2.4.3* 어노테이션/라벨링 조직

어노테이션/라벨링을 위한 조직은 소수의 내부 작업자와 다수의 크라우드 작업자로 구성되며, 요약을 위한 크라우드 작업자는 정해진 내부 규정(예:국문학 전공자 등)에 따라 선발하여 진행한다. 소수의 내부 작업자는 요약을 위한 작업물과 크라우드 작업자에 대한 관리의 업무를 진행한다.

### *2.4.4* 어노테이션/라벨링 도구

크라우드 작업자가 프로젝트에 참여할 경우 아래와 같은 웹기반 크라우드 작업 플랫폼에서 작업을 한다.




크라우드 작업자는 해당 프로젝트를 선택하고, 프로젝트에 진입한 뒤, 사이트내에 링크된 '도서 요약 가이드'에 맞게 왼쪽에 주어진 문장을 오른쪽에 요약문으로 요약/작성하는 작업을 수행한다.


## 2.5 검수

### 2.5.1 검수 절차

데이터의 정확성 확보 및 빠른 검수결과 피드백을 통한 효율적인 데이터셋 구축을 위해 단계별 품질관리체계를 확립하고, 리스크에 대한 사전 대응 방안 및 중점 관리 방안을 수립한다.

효율적인 품질관리를 위해 아래와 같이 여러 역할에 거쳐 다단계 품질검수를 실시하며, 작업자와 별도로 검수자가 구축된 학습데이터를 검수하여 데이터가 고품질을 유지하도록 한다. 각 단계별로는 분량 확인과 함께 검수 지침서를 기반으로 세부적인 검수를 진행하며, 검수자가 데이터를 검수 완료하게 되면 생성요약 학습데이터로 최종 등록한다.

각 단계별 검수 관련 내용은 아래와 같다.

- 1단계 : 작업자에 의한 검증
  예) 작업자 작업 중 자체 검증 방법으로 최종 제출전 자체적으로 가이드라인을 재확인 하고 요약된 문장 제출
- 2단계 : 전문 검수자에 의한 검증
  예) 검증을 담당하는 전문 기관의 조언을 받아서 작업자 1차 검수 본에 대한 샘플링 검증
- 3단계 : 프로토타입을 통한 기계적 자체검증 실시
  예) 2단계 요약 과정에서 나온 학습데이터 셋 중 일부를 테스트 데이터 셋으로 활용하여 기계적 자체 검증
- 4단계 : 외부기관 품질인증
  예) 최종 완성된 도서 요약 데이터 셋을 TTA에서 제시하는 품질 인증 기준에 맞춘 검증

### 2.5.2 검수 기준

데이터 검수는 아래와 같이 형태, 연관성, 일관성을 고려하여 부적합 여부를 판단하며, 부적합한 경우 피드백과 함께 수정을 요청한다.

형태 측면에서는 아래와 같은 경우 부적합한 요약문으로 판단한다.

- 오탈자가 있는 경우
  예) 유전다 재조합 기술의 건전한 …
- 문장 부호 및 기호를 잘못 기입한 경우
  예) 국내외간 교역단계, 운송단계로 나누어진다,
- 문장이 비문이거나 미완성인 경우
  예) 특히 환경영향 평가방법은 (…) 우리나라 현실에 적용하는 있다.

- 본문의 문장을 축약, 변형하지 않고 본문 내용을 그대로 복사, 붙여넣기한 경우

예)

| | |
|---|---|
| 원문 | 최근 정세 변동의 의미와 평가 북핵 문제는 지난 25 년 동안 다양한 철학과 접근법 그리고 전략의 동원에도 풀리지 않았던 난제 중 난제이다. |
| 부적합 | 최근 정세 변동의 의미와 평가 북핵 문제는 지난 25 년 동안 다양한 철학과 접근법 그리고 전략의 동원에도 풀리지 않았던 난제 중 난제이다. |
| 적합 | 북핵 문제는 지난 25 년 동안 다양한 노력에도 해결되지 않은 난제로 |
| 적합 | 북핵 문제는 지난 25 년 동안 다양한 철학과 접근법 그리고 전략의 동원에도 풀리지 않았던 난제이다. |

- 길이가 원문의 30%를 넘는 경우(원문과 요약문의 길이를 비교하여 자동 필터링)
  예) 700자 원문에 대해 300자 요약문 작성

연관성 측면에서는 아래와 같은 경우 부적합한 요약문으로 판단한다.

- 원문의 핵심 내용이 누락된 경우

예)

| | |
|---|---|
| 원문 | 20세기를 지나면서 생활수준이 향상되고 수명이 연장되면서 만성질환은 증가하였지만 전염병은 감소하고 있는 추세이다. |
| 부적합 | 20세기 이후 생활수준 향상과 수명 연장과 더불어 만성질환은 증가했다. |
| 적합 | 20세기 이후 생활수준 향상과 수명 연장과 더불어 만성질환은 증가한 반면 전염병은 감소 추세이다. |

- 핵심이 아닌 부차적인 내용을 포함한 경우 (e.g. 사례, 수치, 약자 등)

예)

| | |
|---|---|
| 원문 | 먼저, 만성 알코올 섭취 동물모델을 만들기 위해 7주령 된 수컷 쥐(C57/BL6J)에 단백질 18%, 지방 35%와 함께 탄수화물이 47%가 든 대조군 액상식이(Lieber-Decarliregularliquiddiet;Dyets)와 47% 탄수화물대신에 11% 탄수 화물 및 36%에탄올이 든 에탄올 식이를 쥐에 8주간 섭취하게 하였다. |
| 부적합 | 만성 알코올 섭취 동물모델을 위해 7주령 된 쥐(C57/BL6J)에 단백질 18%, 지방 35%와 함께 탄수화물이 47%가 든 대조군 액상식이(Lieber-Decarliregularliquiddiet;Dyets)와 47% 탄수화물 대신 11% 탄수화물, 36% 에탄올이 함유된 식이를 쥐에 8주간 섭취하게 하였다. |
| 적합 | 만성 알코올 섭취 동물모델을 위해 7주령 된 쥐에 단백질 18%, 지방 35%와 함께 탄수화물이 47%가 든 대조군 액상식이와 47% 탄수화물 대신 11% 탄수 화물, 36% 에탄올이 함유된 식이를 쥐에 8주간 섭취하게 하였다. |

일관성 측면에서는 아래와 같은 경우 부적합한 요약문으로 판단한다.

- 원문과 명확히 다른 내용

예)

| | |
|---|---|
| 원문 | 개인별 예방접종 일정안내 정보서비스를 제공함으로써 중복접종을 방지하고 적기 접종과 완전 접종을 보장할 수 있다. |
| 부적합 | 개인별 예방접종 일정안내를 제공하여 중복접종, 적기 접종과 완전 접종을 보장한다. |
| 적합 | 개인별 예방접종 일정안내를 제공하여 중복접종 방지, 적기 접종과 완전 접종을 보장한다. |

- 실제 내용을 파악하기 어려울 정도로 상이한 용어 및 표현을 사용한 경우

예)

| | |
|---|---|
| 원문 | 효과적인 자폐치료를 위해 가져야 할 로봇의 주요 역할은 아동과 상호작용하기 위한 |
| 부적합 | 효과적인 정신질환 치료를 위한 기계의 주요 역할은 아동과 상호작용하기 위한 |
| 적합 | 효과적 자폐 치료를 위한 로봇의 주된 역할은 아동과 소통하기 위한 |

### 2.5.3 검수 조직

검수 조직에 대한 구성은 컨소시엄내 협의를 통해서 구성하되, 검수 절차의 2단계에 해당하는 '전문가 검수 집단'을 포함하여 구성한다. 아울러, 검수 조직은 필히 2인 1조의 형식으로 검수 진행하여 학습데이터의 완성도를 높인다.

### 2.5.4 검수 도구

검수자는 검수 프로젝트를 선택하고 프로젝트에 진입한 뒤, 사이트 내에 링크된 '도서 정제 가이드'와 '도서 요약 가이드'를 기반으로 검수 작업을 수행한다. 다만, 수행 중 검수 기준이 불명확 할 경우, 전문가 검수 집단과의 협의를 통해 불명확한 내용에 대한 검수 기준을 확정하고, 이를 가이드에 추가 하여 작업의 혼선을 방지한다.


### 2.5.5 기타 품질관리 활동

우수한 품질의 데이터를 기간 안에 확보하기 위해 사업 추진일정 및 마일스톤, 데이터셋 구축 및 검수 수량, 구축 데이터셋의 난이도 및 정확도 등 평가 지표를 선정하여 수시로 요소별로 성과를 측정한다. 평가 결과를 종합하여 개선 사항을 도출하고, 구축된 데이터 품질 및 성과를 상시 모니터링하여 이슈에 즉시 대처할 수 있도록 한다.

또한 한국정보통신기술협회(TTA)와 연계하여 안정성, 유일성, 유효성, 일관성, 정확성 등의 품질기준에 따라 데이터의 품질검토 기준을 정의하고 품질 인증을 요청함으로써 품질 검증의 객관성을 확보한다. 이를 통해 데이터의 패턴 및 체크비트, 허용범위, 분류 표준 정의 값 및 유효값 등 주요 항목에 대한 품질 인증을 획득한다. 필요한 검증 환경 및 도구, 데이터 제공 및 수반하는 일체의 비용 등을 제공하는 등 검증 프로세스가 원활히 수행되도록 적극적으로 지원한다. 필요시 학습 데이터 제작 툴 내에서 구축된 데이터의 건수를 확인할 수 있는 기능을 활용하여, 대시보드를 구축, 데이터의 수량이나 내용 등을 상시 모니터링 한다.

## 2.6 활용

### 2.6.1 활용 모델

#### 2.6.1.1 모델 학습

생성 요약(Abstractive Summarization)은 텍스트 요약의 최신 기술로 주어진 텍스트의 내용을 기계가 이해한 내용을 바탕으로 압축하여 새로운 텍스트로 만드는 것을 의미하며 딥러닝에 기반한 자연어 이해와 생성 기술이 필요하다. 이를 구현하기 위해 본 과제에서는 현재 자연어처리 분야에서 최고의 성능을 발휘하고 있는 Transfer learning 기반의 요약 모델을 구축한다. Transfer learning 모델은 1차적으로 대규모 말뭉치를 이용하여 언어모델 기반의 학습이 이루어진 Pre-trained Language Model(PLM)에 각 목적에 맞는 데이터로 추가 학습을 함으로써, 적은 양의 학습데이터와 학습 시간에도 불구하고 뛰어난 성능을 발휘한다. 본 과제에서는 BERT 등의 최신의 pre-trained language model(PLM)을 활용하여 요약 모델을 구축한다.

구축된 데이터는 학습용(Training), 검증용(Validation), 평가용(Test)으로 분할하며 그 비율은 각각 80%, 10%, 10%로 구성하여 모델 학습 및 평가에 활용한다. 모델의 성능 평가는 ROUGE와 METEOR와 같은 재현율, 정밀도 등을 측정하는 메트릭을 활용하여 수행한다. 지속적인 테스트를 통해 요약 모델의 성능을 개선하고, 최적의 도서자료 요약 모델로 구축한다.

**※ 데이터셋의 구성**

| 데이터셋 구분 | 용도 | 건수(비율) |
|---|---|---|
| 학습용(Training) | 요약 모델 도출을 위한 기계학습에 활용하는 데이터셋 | 16만건 (80%) |
| 검증용(Validation) | 요약 모델의 성능 향상을 위한 알고리즘, 파라미터 조정 등에 활용하는 데이터셋 | 2만건 (10%) |
| 평가용(Test) | 도출된 요약 모델의 최종 성능 평가에 활용하는 데이터셋 | 2만건 (10%) |

#### 2.6.1.2 서비스 활용 시나리오

생성요약 기술은 비정형 문서의 내용 자체를 쉽게 이해할 수 있는 텍스트로 간략히 표현해 주는 기술이므로 매일 접하게 되는 많은 양의 문서 내용을 빠르게 습득하고 일상생활이나 업무에 활용할 수 있도록 도움을 줄 수 있다. 이용자들의 정보 습득의 접점이 PC 환경에서 스마트폰을 넘어 스마트 스피커, 챗봇, 자동차, 가상현실 기기(VR devices) 등으로 이동함에 따라 문서에 대한 검색보다는 실제적인 정보를 빠르게 습득할 수 있는 생성 요약 서비스의 필요성이 대두되고 있다. 일례로 인공지능 기술 기반의 생성 요약 기술은 뉴스 기사 요약, 기업의 KMS(Knowledge Management System), 인터넷 정보 탐색 서비스, 서적 관련 업무 등 다양한 분야에 적용하여 활용할 수 있다.

본 사업에서는 저작권이 해소된 자료 위주로 원시데이터를 수급한 관계로 도서가 정책 자료에 한정되어 있다는 한계점이 있으나, 해당 데이터로 학습한 모델로도 모든 문서에 대한 요약이 가능하므로 다른 분야의 서비스에 적용하는 것도 가능하다. 또한 추후 일반 도서(인문, 자기계발, 역사/문화 등) 데이터 수급이 가능하다면 유사한 방식으로 데이터 구축 후 이에 더 적합한 서비스도 제공할 수 있을 것으로 예상된다.

### 2.6.2 데이터 제공

구축된 생성요약 데이터는 AI 오픈이노베이션 허브에 공개하여 민간 영역의 자원과 융·복합을 유도하고, 이를 통해 신규 비즈니스 창출을 독려하고자 한다. 단 학습데이터를 쉽게 이용할 수 있도록 효율적인 정보 전달 형태는 AI 오픈이노베이션 허브와 협의 후 가장 적절한 형태로 데이터를 제공한다. 정보 제공 시에는 소개 페이지를 통해 구축 내용과 데이터베이스 구조 등을 함께 공개하여 활용에 참고할 수 있도록 하며, 현재 텍스트 데이터 분류 기준에 도서 자료 요약 데이터셋에 적합한 범주가 없으므로 데이터셋 등록시 새로운 기준으로 등록되기를 기대한다.

이와 함께 한국정보화진흥원에서 운영하는 공공데이터 포털(https://data.go.kr)에도 데이터를 공개함으로써 사용자들이 공공데이터를 원할 경우, 간편하게 회원 가입 후 파일 다운로드 혹은 오픈 API 형태로 자유롭게 활용할 수 있도록 한다.

단, 저작권 등의 이슈를 사전에 해결하기 위해 학습데이터 공개방법에 준용하여 이용조건 등의 동의 및 서약서 업로드 후 이용 가능하도록 조치한다.

### 2.6.3 데이터 유지보수

AI 허브에 공개된 학습 데이터 중 오류가 발견/보고된 데이터에 대해서는 해당 작업에 대한 1) 히스토리 확인 2) 오류내용 확인 3)작업 가이드 확인을 통하여 데이터를 보완하여 재공유 한다. 또한 보완하는 과정에서 본 가이드라인의 변경이 필요할 경우에는 추가적으로 가이드라인을 변경하여 공지한다.