프로젝트 진행 상황 보고서

Carbot팀 안성규, 박상범, 김세희, 문벼리

Contents

5주차 리뷰

데이터 전처리

특수 데이터(표, 수식 등)는 Markdown,
*수식이 정확하게 인식되지 않음 -> 발전 필요
전체적인 틀과 텍스트는 JSONL 형식으로 처리

전문과 개정안 병합

데이터 베이스에 저장해놓고 불러오는 방식 선정

RAG 기반 LLM

저렴하고 성능도 좋은 GPT-4o mini 고려

Pipelines from PDF to JSONL

문서 원본 (PDF)　　마크 다운 텍스트

본문 텍스트

JSON

테이블 텍스트

수식 텍스트
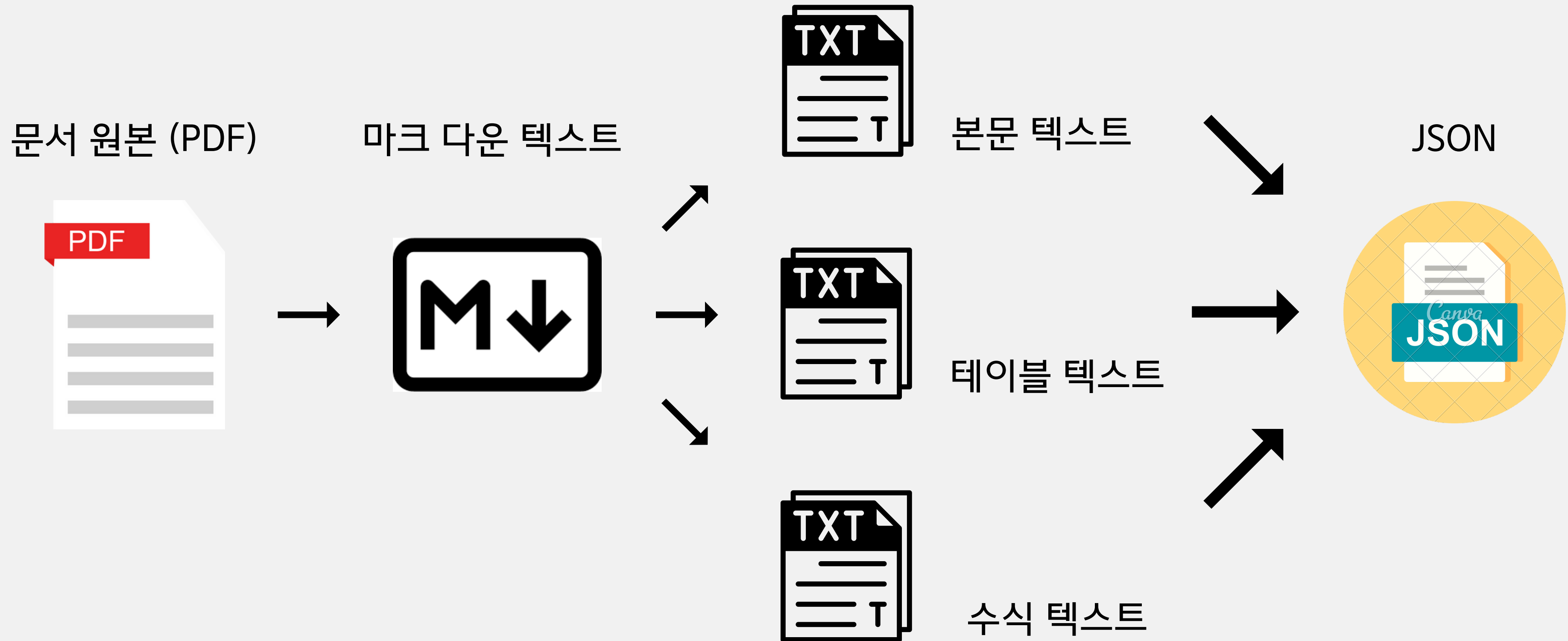
Pipelines from PDF to JSONL- Table part

Step 1. 테이블 전처리
수식 전처리 + Markdown 형식화

Step 2. Json화
각 규제별로 딕션너리 만들어 json 파일 생성
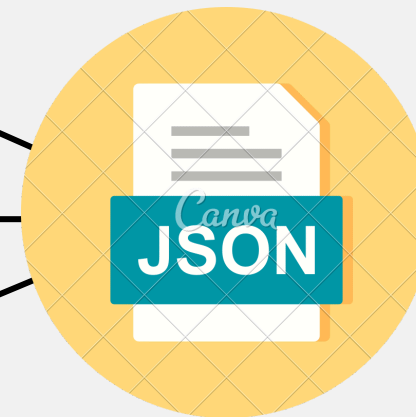
Step 3. 전달
생성된 파일 json원본 파일에 결합
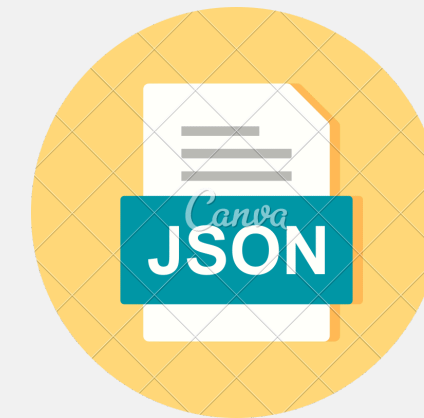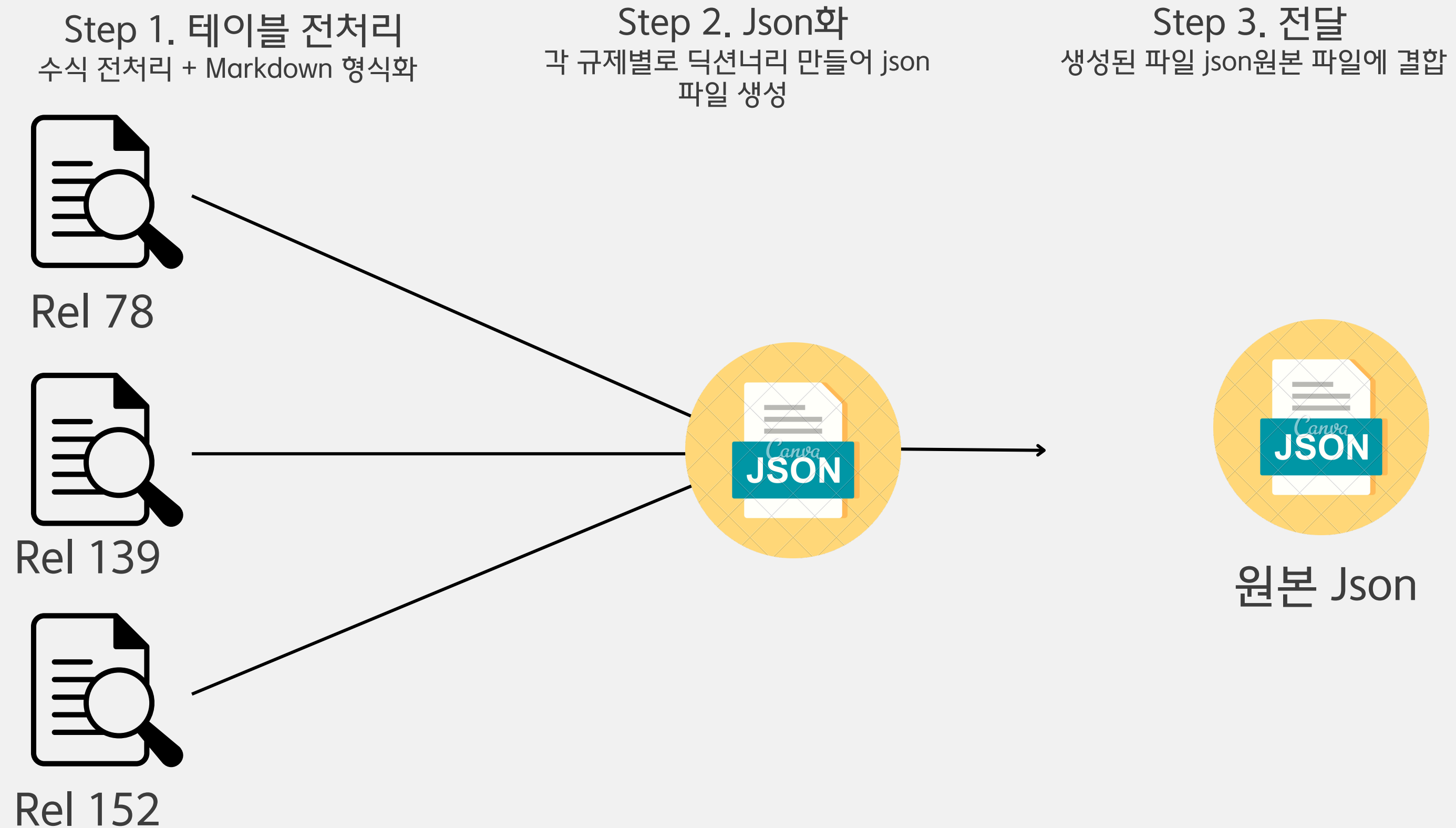
Rel 78

Rel 139

Rel 152

JSON

원본 Json

Json파일 예시

Pipelines from PDF to JSONL- Text part

문서 원본 (PDF)   마크 다운 텍스트

→

pymupdf4llm라이브러리 사용

RAG/LLM 특화 라이브러리

5. Specifications

5.1. Brake system requirements

5.1.1. Each vehicle shall meet each of the tests specified for a vehicle of its category and for those brake features on the vehicle.

5.1.2. Service brake system control operation

Vehicles shall have configurations that enable a rider to actuate the service brake system control while seated in the normal driving position and with both hands on the steering control.

5.1.3. Secondary brake system control operation

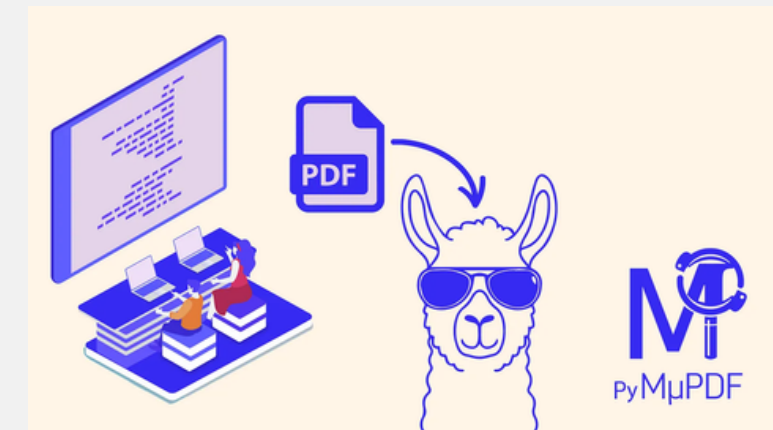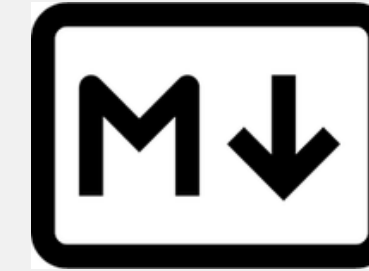Vehicles shall have configurations that enable a rider to actuate the secondary brake system control while seated in the normal driving position and with at least one hand on the steering control.

5.1.4. Parking brake system

If a parking brake system is fitted, it shall hold the vehicle stationary on the slope prescribed in paragraph 1.1.4. of Annex 3.

The parking brake system shall:

(a) Have a control which is separate from the service brake system controls; and

(b) Be held in the locked position by solely mechanical means.

Pipelines from PDF to JSONL

마크 다운 텍스트

본문 텍스트

테이블 텍스트

수식 텍스트
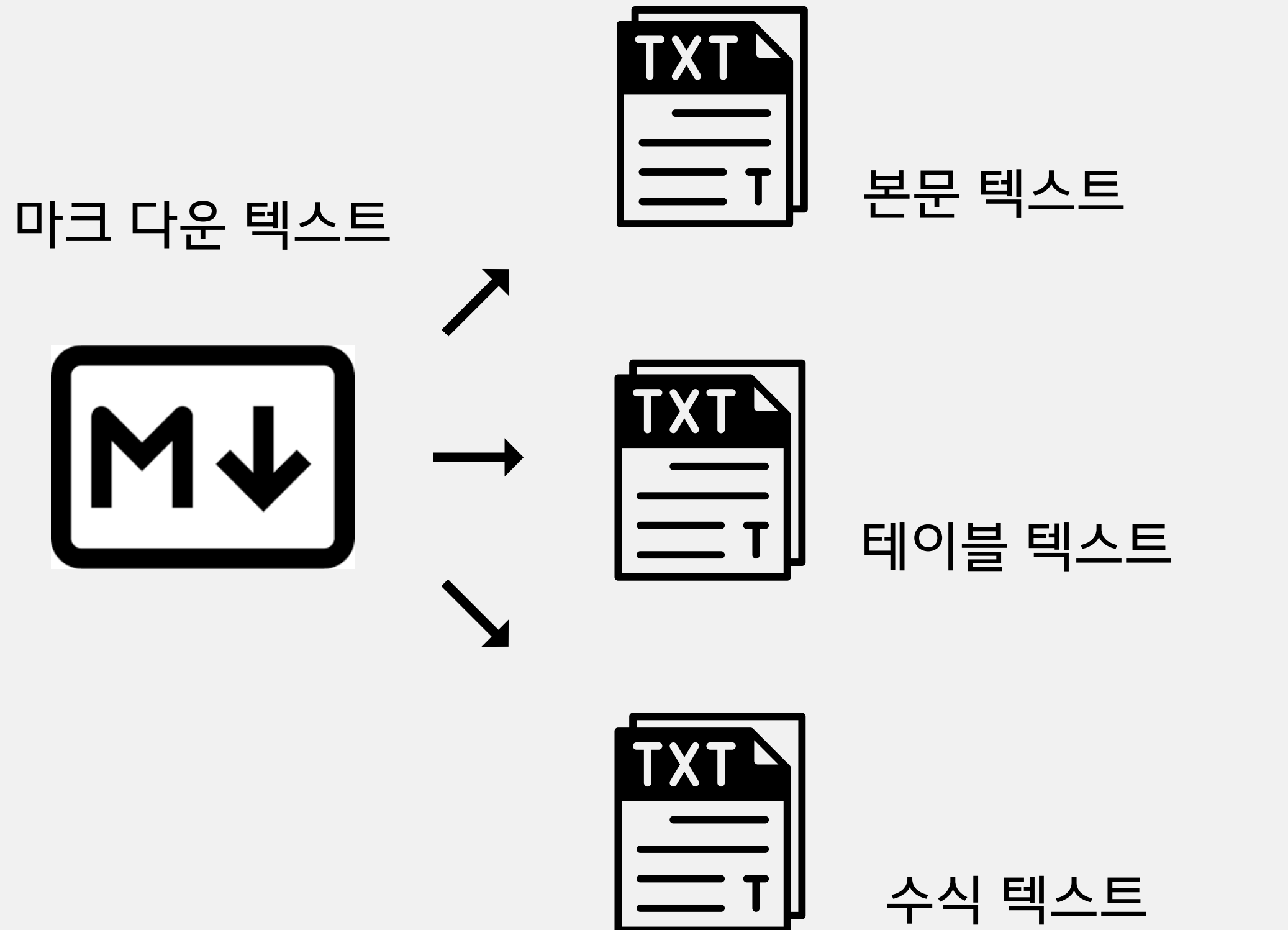
- 머릿말 및 쪽번호 제거

- 수식과 테이블 추충

- 행 병합

Pipelines from PDF to JSONL
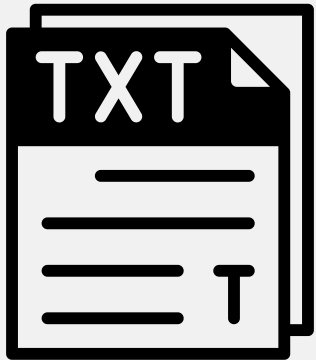
본문 텍스트

1. Alternative method for the determination of peak braking coefficient (PBC)
1.1. General:
(a) The test is to establish a PBC for the vehicle type when being brakedon the test surfaces described in Annex 3, paragraphs 1. 1.1. and 1.1.2.
(b) The test comprises a number of stops with varying brake control forces. Both wheels shall be braked simultaneously up to the point reached befor
(c) The maximum vehicle deceleration rate is the highest value recorded during all the test stops.
(d) The Peak Braking Coefficient (PBC) is calculated from the test stop that generates the maximum vehicle deceleration rate, as follows: [Equation 1]
(e) The value of PBC shall be rounded to two decimal places.
[Equation 1]
1.2. Vehicle condition:
(a) The test is applicable to all vehicle categories.
(b) The anti-lock system shall be either disconnected or inoperative (ABSfunction disabled), between 40 km/h and 20 km/h.
(c) Lightly loaded.

TAG

Pipelines from PDF to JSONL

테이블 텍스트

Table 3:
The test series comprises the following individual tests, which may be carried out in any order:

|ABS Tests|Paragraph|
|a. Stops on a high friction surface - as specified in paragraph 1.1.1. |9.3.
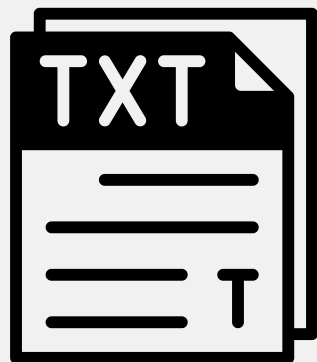b. Stops on a low friction surface - as specified in paragraph 1.1.2. |9.4.
c. Wheel lock checks on high and low friction surfaces. |9.5.
d. Wheel lock check - high to low friction surface transition.| 9.6.
e. Wheel lock check - low to high friction surface transition. |9.7.
f. Stops with an ABS electrical failure.| 9.8.|
[End of Table]

Table 4:
Vehicle Category	STOPPING DISTANCE(S) (Where V is the specified test speed in km/h and S is the required stopping distance in metres)		MFDD
Single brake system			
L 1	S ≤ 0.1 V + 0.0143 V^2		≥ 2.7 m/s^2
L2 and L6	S ≤ 0.1 V + 0.0143 V^2		≥ 2.7 m/s^2
L 3	S ≤ 0.1 V + 0.0133 V^2		≥ 2.9 m/s^2
L 4	S ≤ 0.1 V + 0.0105 V^2		≥ 3.6 m/s^2
Vehicles with CBS or SSBS			
ALL	S ≤ 0.1 V + 0.0154 V^2		≥ 2.5 m/s^2

Pipelines from PDF to JSONL

TXT
본문 텍스트

TXT
테이블 텍스트

TXT
수식 텍스트
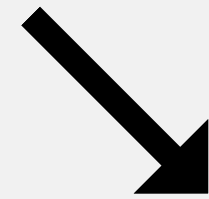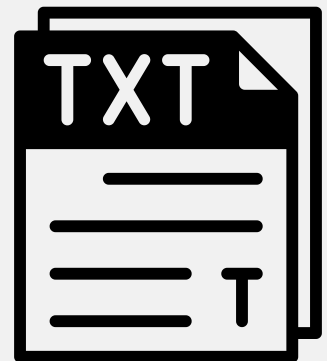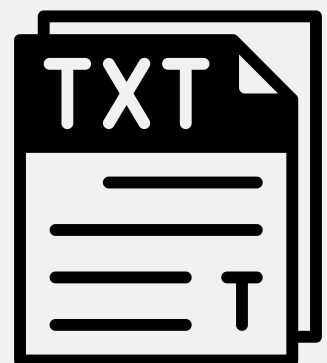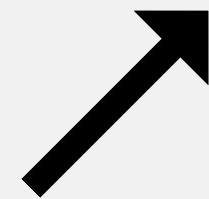
JSON

- 구조 체계에 따라 dict 객체 생성

  - Chapter N.
  - Article N.N.
  - Paragraph (N.N.N.)
  - Sub-Paragraph (N.N.N.N.)

- 지정된 테그를 인식하여 특수한 데이터를 입력

- dict 객체를 JSON으로 변환

Pipelines from PDF to JSONL

Function recursive_dictLoader(chapter_line, ...):

1. Chapter: abc
1.1. Article: 123
1.2. Article: ~
1.3. Article: ~

2. Chapter: ~
2.1. Article: ~
2.2. Article: ~

상위 조항부터 탐색 진행

조건에 맞는 행을 찾으면
행의 내용을 바탕으로 dict 객체 생성

Chapter 1 : { description : "abc"}

Pipelines from PDF to JSONL

Function recursive_dictLoader(chapter_line, ...):

1. Chapter: abc
1.1. Article: 123
1.2. Article: ~
1.3. Article: ~

2. Chapter: ~
2.1. Article: ~
2.2. Article: ~

상위 항목에 해당하는 행을 기준으로 하위 항목에 대한 탐색 진행

Chapter 1 : { description : "abc"}

Pipelines from PDF to JSONL

JSON

Function recursive_dictLoader(chapter_line, ...):

    1. Chapter: abc
1.1. Article: 123
1.2. Article: ~
1.3. Article: ~

    2. Chapter: ~
2.1. Article: ~
2.2. Article: ~

하위 항목에 해당하는 행을 찾으면 이를 다시 상위 항목으로 지정하여 함수를 재호출 (객체 생성 – 탐색)

Chapter 1 : { description : "abc"
                    Article 1.1. : {"123"}
            }

Pipelines from PDF to JSONL

JSON

```
'2.5.': {'Description': ['Burnishing procedure',
                         'The vehicle brakes are burnished prior to '
                         'evaluating performance. Thisprocedure may be '
                         'completed by the manufacturer:'],
         'Item': ['(a) Vehicle lightly loaded;',
                  '(b) Engine disconnected;',
                  ['(c) Test speed:',
                   ['(i) Initial speed: 50 km/h or 0.8 Vmax, whichever is '
                    'lower;',
                    '(ii) Final speed = 5 to 10 km/h;']],
                  ['(d) Brake application:',
                   ['(i) Each service brake system control actuated '
                    'separately;']],
                  ['(e) Vehicle deceleration:',
                   ['(i) Single front brake system only: 3.0-3.5 m/s^2 for '
                    'vehicle categories L 3 and L 4; 1.5-2.0 m/s2 for vehicle '
                    'categories L 1 and L 2;',
                    '(ii) Single rear brake system only: 1.5 -2.0 m/s^2;',
                    '(iii) CBS or split service brake system: 3.5 -4.0 '
                    'm/s^2;']],
                  '(f) Number of decelerations: 100 per brake system;',
                  '(g) Initial brake temperature before each brake '
                  'application ≤ 100 °C;',
                  '(h) For the first stop, accelerate the vehicle to the '
                  'initial speed and then actuate the brake control under the '
                  'conditions specified until the final speed is reached. '
                  'Then reaccelerate to the initial speed and maintain that '
                  'speed until the brake temperature falls to the specified '
                  'initial value. When these conditions are met, reapply the '
                  'brake as specified. Repeat this procedure for the number '
                  'of specified decelerations. After burnishing, adjust the '
                  "brakes in accordance with the manufacturer's "
                  'recommendations.']},
'Chapter': '2',
'Title': 'Preparation'}
```
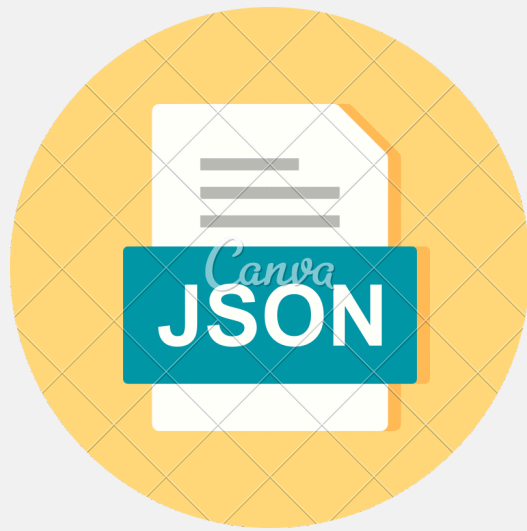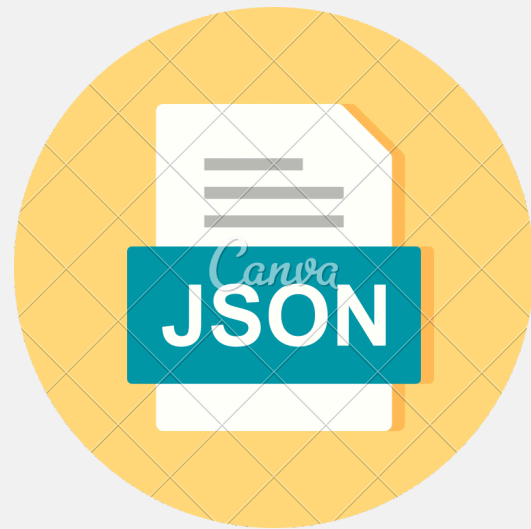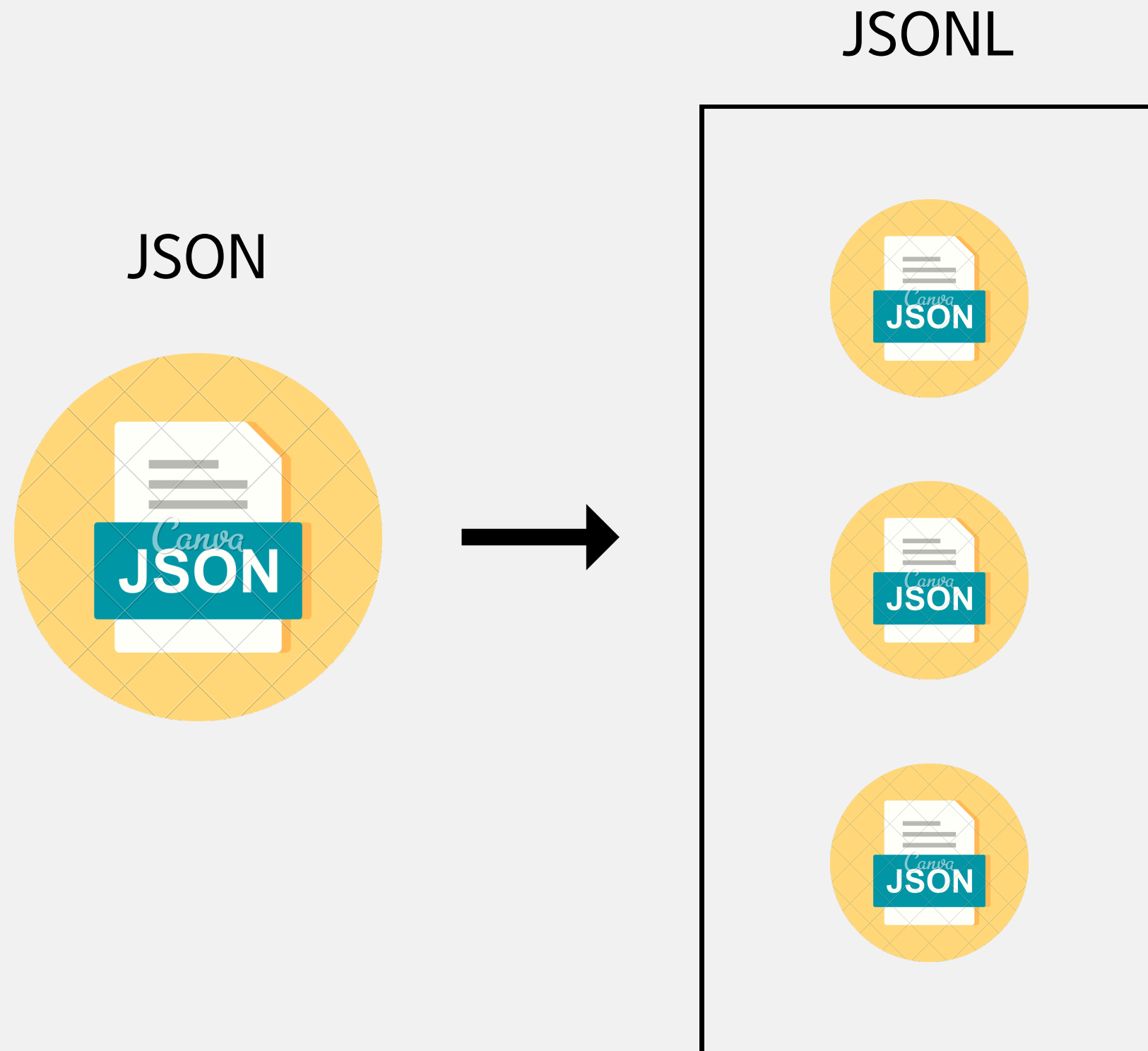
Pipelines from PDF to JSONL

JSON

JSONL

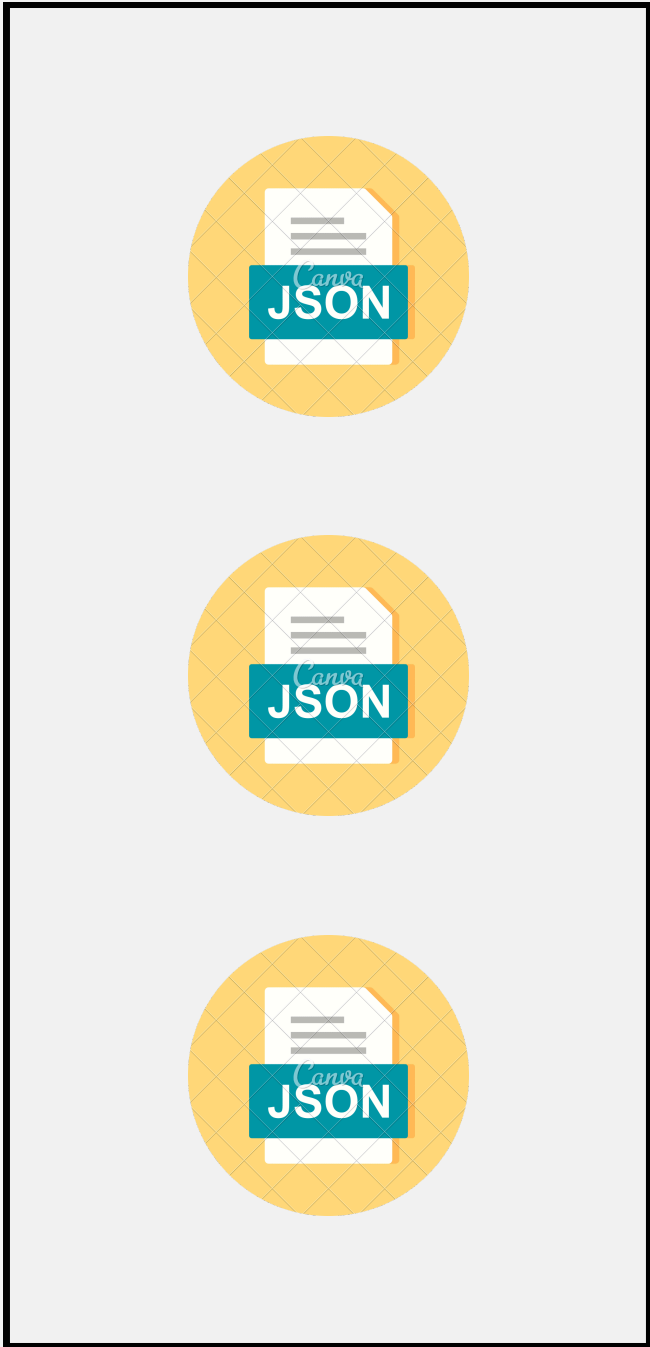- 지정한 청크 및 오버랩 사이즈를 바탕으로 JSON 분할

- 분할된 JSON 파일을 엮어서 JSONL 파일을 생성하여 RAG 프로세스에 전달

Pipelines from PDF to JSONL

JSONL

청크 사이즈: 500tk   오버랩 사이즈: 100tk

1. Chapter
1.1. Article: 250
1.2. Article: 120

370 tk

2. Chapter
2.1. Article: 150
2.2. Article: 160

510 tk

2.3. Article: 190
2.4. Article: 300

490 tk

데이터 수집

PDF

MarkDown

JSONL

RAGAS 평가방식 사용

질문과 답변 사이에서 사용된 토큰, api비용에 대한 데이터 수집

gpt-3.5-turbo
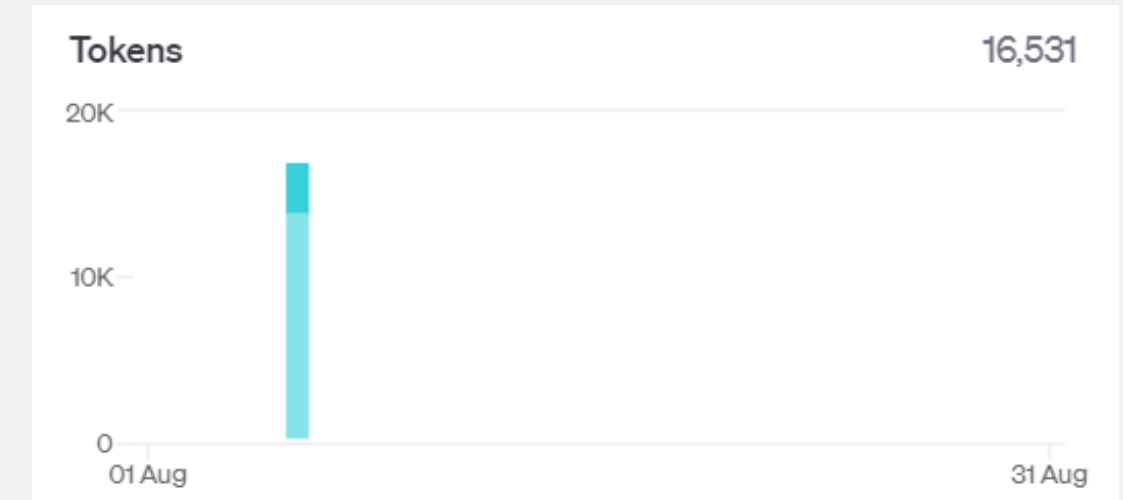

context 19,773
generate 2,744

0.01$

gpt-4o-mini

context 12,152
generate 3,341

< 0.01$

gpt-4-turbo

context 13,491
generate 3,040

0.07$

gpt-3.5-turbo

질문: "What is the purpose of UN Regulation No. 78?"
답변: "UN Regulation No. 78 establishes uniform provisions concerning the approval of vehicles of categories L1, L2, L3, L4, and L5 with regard to braking."
근거 점수: 5
관련성 점수: 5
독립성 점수: 5

gpt-4o-mini

질문: "What is the position of the spray nozzles on the brake testing equipment according to the given context?"
답변: "The spray nozzles are positioned two thirds of the distance from the outer circumference of the rotating drum to the wheel hub centre."
근거 점수: 5
관련성 점수: 5
독립성 점수: 5

gpt-4-turbo

질문: "What are the test conditions for conducting stops on a high friction surface in ABS tests?"
답변: "Initial brake temperature between 55°C and 100°C, test speed of 60 km/h or 0.9 Vmax, simultaneous actuation of both brake controls, and brake actuation force to ensure full cycling down to 10 km/h."
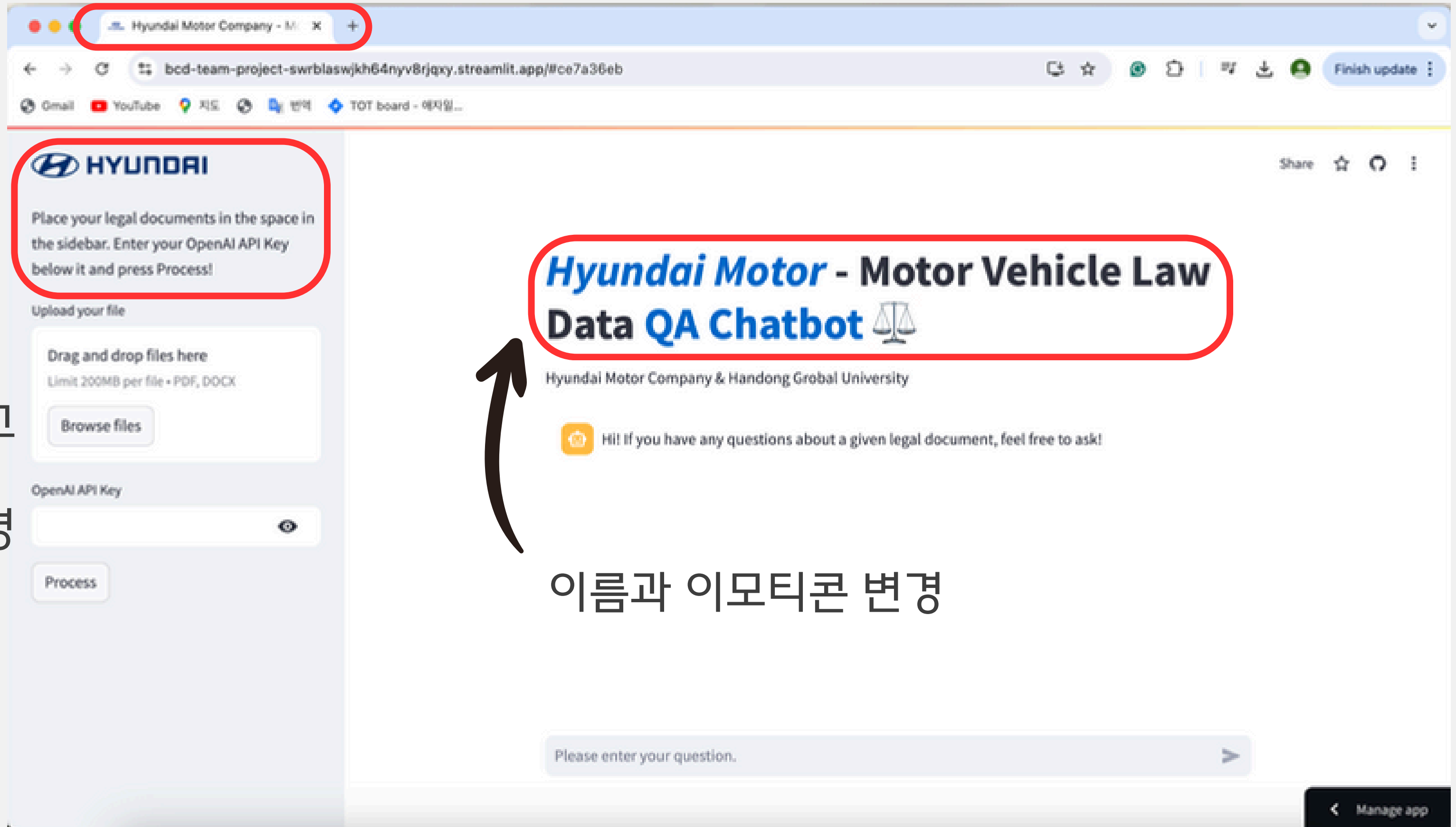근거 점수: 5
관련성 점수: 5
독립성 점수: 5

Streamlit 웹페이지 개선

웹페이지 대표 로고 선정 및 웹페이지 이름 변경

사이드바 로고
설정 및
웹페이지 설명

이름과 이모티콘 변경

참가할 수 있는 학회 일정 및 향후 계획

학회	작년 일정	데드라인	웹사이트	비고
한국데이터마이닝학회	2023년 11월 7일 (화) 09:00 ~ 18:00	초록 10/10/2023 초록 결과 10/17	한국데이터마이닝학회 링크	추계학술대회 학생논문경진대회
한국데이터정보과학회	2023년 11월 3일 (금) ~ 11월 4일(토)	10/15/2023	한국데이터정보과학회링크	추계학술논문발표회 대학원생논문발표대회
실용인공지능 학술대회	2023년 12월 7일 ~ 8일	초록 10월 27일(금) - 연장마감	실용인공지능링크	
한국정보기술학회	2023년 11월 23일 (목) ~ 25일(토)	논문제출: ~ 2023년 10월 31일(화)	한국정보기술학회링크	추계종합학술대회 및 대학생논문경진대회
한국인공지능학회	2023년 11월 23일 (목) ~25일(토)	논문제출 11/3/2023	한국인공지능학회	추계학술대회