

2024년 9월 주주총회



LangGraph Parser

```
{'table': '/Users/teddy/Dev/github/teddynote-parser/data/디지털정부혁신 추진계획/143.png', 'text': "- 공공부문 자기정보 다운로드 서비스 개시 (2020년 포털 구축) - 행안부\n- 마이데이터 포털 구축: 국민이 자기정보 검색 및 다운로드 가능\n- 클라우드 기반 '자기정보 저장소' 제공\n- '나의 데이터' 및 '꾸러미 데이터' 다운로드 제공 (은행대출, 건강검진, 신용회복 등)\n- 데이터 3법 입법 이후 자기정보 제3자 전송권 검토\n- 스마트폰 기반 디지털 신분증 도입 (기존 신분증과 병행 발급)\n- 국민이 고지서를 원하는 온라인 채널로 수신 및 간편납부 가능\n- 2018년 종이고지서 건수: 국세 1.1M, 지방세 19.5M, 4대보험 12.5M, 자동차검사안내 23M\n- 디지털 고지·수납 공통기반 구축 (2021년)\n- 종이증명서 감축 시 연간 5천억원 사회적 비용절감 기대\n- 2018년 전국 지자체 종이고지서 제작·발송비용: 1,000억원 규모", 'page': 12, 'id': 143, 'language': 'Korean'},
```

Text

- `text`: page 단위의 전체 Text (Raw Text)(`page_content`):
 - `RecursiveCharacterTextSplit(500, 50)`
- `text_summary`: page 단위의 텍스트 요약(`page_content`)

Table

- `'tables'`: table 을 crop 한 이미지.
`'/Users/teddy/Dev/github/teddynote-parser/data/디지털정부혁신 추진계획/83.png'`: (metadata: table)

- 'table_summary': 테이블 요약 (page_content)
- 'table_markdown': 테이블 마크다운 (metadata: markdown)

Image

- 'image_summary': 이미지 요약(page_content)
- 'images': 이미지 경로. '/Users/teddy/Dev/github/teddynote-parser/data/디지털정부혁신 추진계획/79.png' (metadata: image)

```
{'id': None, 'metadata': {'type': 'text', 'page': '1',
'source': '/Users/teddy/Dev/github/teddynote-parser/data/디지털정부혁신 추진계획.pdf', 'summary': '- 개요: 문서의 구조 및
내용 소개\n- 디지털 정부혁신 추진계획:\n - 우선 추진과제:\n
1. 선제적 · 통합적 대국민 서비스 혁신\n 2. 공공부문 마이데이터
활성화\n 3. 시민참여를 위한 플랫폼 고도화\n 4. 현장중심 협업
을 지원하는 스마트 업무환경 구현\n 5. 클라우드와 디지털서비스
이용 활성화\n 6. 개방형 데이터 · 서비스 생태계 구축\n - 중장기
범정부 디지털 전환 로드맵 수립\n- 추진체계 및 일정: 4페이지에
상세 내용 포함\n- [붙임]: 디지털 정부혁신 우선 추진과제(상세)
8페이지'}, 'page_content': '순 서\n I . 개요
```

```
. . . . .
. . . . .
```

```
. . . 1\n II . 디지털 정부혁신 추진계획
```

```
. . . . . 2\n 1. 우선 추
진과제
```

```
. . . . .
. . . . .
```

```
. . . . 2\n ① 선제적 · 통합적 대국민 서비스 혁신\n ② 공공부문
마이데이터 활성화\n ③ 시민참여를 위한 플랫폼 고도화\n ④ 현장
중심 협업을 지원하는 스마트 업무환경 구현\n ⑤ 클라우드와 디지
털서비스 이용 활성화\n ⑥ 개방형 데이터 · 서비스 생태계 구축
```

```
\n2. 중장기 범정부 디지털 전환 로드맵 수립
. . . . . 4\nIII. 추진체계 및 일정
. . . . .
. . . . . 4\n[붙임] 디지털 정부혁신 우선 추진과제(상
세) . . . . . 8', 'type': 'Document'}
```

```
{'id': None, 'metadata': {'type': 'image', 'image':
'/Users/teddy/Dev/github/teddynote-parser/data/디지털정부혁신
추진계획/79.png', 'page': '6', 'source':
'/Users/teddy/Dev/github/teddynote-parser/data/디지털정부혁신
추진계획.pdf', 'id': '79'}, 'page_content':
'<image>\n<title>\n디지털 정부혁신 추진전략
\n</title>\n<summary>\n01 문서는 디지털 정부혁신 추진전략에
대한 내용을 담고 있으며, 주요 목표는 정부 서비스의 디지털화
및 효율성 증대이다. 핵심 기술로는 AI, 빅데이터, 클라우드 컴퓨
팅이 포함되며, 추진 기간은 2023년부터 2025년까지이다. 주요 성
과 지표로는 서비스 이용자 수, 처리 속도 개선, 비용 절감이 있
다.\n</summary>\n<entities> \n- 비전: 디지털로 여는 좋은 세상
\n- 추진 원칙: 최종 이용자의 관점, 공공서비스 수준 향상, 혁신
적인 방법\n- 우선 추진 과제: 대국민 서비스 혁신, 마이데이터
활성화, 시민참여 플랫폼 고도화\n- 핵심 기술: AI, 빅데이터, 클
라우드 컴퓨팅\n- 추진 기간: 2023년 - 2025년\n- 성과 지표: 서
비스 이용자 수, 처리 속도 개선, 비용 절감
\n</entities>\n\n</image>', 'type': 'Document'}
```

```
{'id': None,
'metadata': {'type': 'table',
'table': '/Users/teddy/Dev/github/teddynote-parser/data/디
지털정부혁신 추진계획/83.png',
'markdown': '| 추진일정 | 2019 | 2020 | 2021 | 2022 | 관계
부처 |\n|-----|-----|-----|-----|-----|-----
```

```
|\\n|          | 상   | 하   | 상   | 하   |          |',
  'page': '7',
  'source': '/Users/teddy/Dev/github/teddynote-parser/data/디지털정부혁신 추진계획.pdf',
  'id': '83'},
  'page_content': '<table>\\n<title>\\n주요 과제별 추진일정 검토 및 확정\\n</title>\\n<summary>\\n이 표는 2019년부터 2022년까지의 주요 과제별 추진일정을 연도별로 정리한 것입니다. 각 부처의 반영이 필요하며, 추진일정은 사업계획 구체화 및 연차별 예산 확보 상황에 따라 변동 가능성이 있습니다.\\n</summary>\\n<entities>\\n- 연도: 2019, 2020, 2021, 2022\\n- 추진일정: 상반기, 하반기\\n- 관계 부처: 각 부처\\n</entities>\\n<data_insights>\\n- 추진 일정은 연도별로 상반기와 하반기로 나뉘어 있으며, 각 연도마다 일정이 다를 수 있음.\\n- 각 부처의 반영 필요성이 강조 됨.\\n</data_insights>\\n\\n</table>',
  'type': 'Document'}
```

docs 구조

metadata

- type: "text", "page_summary", "table", "image"
- page: 페이지. 숫자 처리 끝났음(1페이지 부터 시작)
- source: 파일명

각 요소별 구조

type: "text"

- page_content: 500 Chunk 기준으로 나눈 청크가 기입
- metadata
 - summary: page 단위의 요약

- page: 페이지
- source: 파일명

type: "page_summary"

- page_content: 요약
- metadata
 - text: 원본 텍스트(1개 페이지에 대한 전체 텍스트)
 - source: 파일명
 - page: 페이지 번호

type: "image"

- page_content: 이미지 설명, entity, 요약 등
- metadata
 - image: 이미지 파일 경로
 - source: 파일명
 - page: 페이지 번호
 - id: 추출 element id

type: "table"

- page_content: 테이블 설명, entity, 인사이트 등
- metadata
 - table: 테이블 이미지 경로
 - markdown: 테이블을 마크다운으로 변경한 내용
 - source: 파일명
 - page: 페이지 번호
 - id: element id