

Lexi Convert 공식 사용자 가이드

Lexi Convert에 오신 것을 환영합니다!

Lexi Convert는 EPUB, PDF, HTML과 같은 복잡한 문서들을 분석하여, 개발자나 데이터 분석가, AI가 즉시 활용할 수 있는 ****구조화된 데이터(JSON)****나, 사람이 읽기 편한 ****일반 문서(Markdown, Text)****로 변환해주는 강력하고 편리한 도구입니다.

이 가이드를 통해 Lexi Convert의 모든 기능을 쉽고 완벽하게 마스터해보세요!

🌟 1. 시작하기: 설치 및 첫 실행

Lexi Convert는 여러분의 편의를 위해, 프로그램에 필요한 라이브러리가 설치되어 있는지 스스로 확인합니다.

- `main.py`를 실행하면, 필요한 라이브러리가 없는 경우 아래와 같은 팝업창이 나타납니다.
- 망설이지 말고 `[예(Y)]` 버튼을 눌러 설치를 진행해주세요.
- 설치가 완료되면 "프로그램을 재시작해주세요"라는 안내창이 뜹니다. 확인을 누른 후, 프로그램을 완전히 껐다가 다시 실행하면 모든 준비가 끝납니다.

💡 전문가를 위한 팁

만약 터미널 환경에서 직접 설치하고 싶다면, 아래 명령어를 입력하세요.

Bash

```
pip install ebooklib beautifulsoup4 "pymupdf<1.24.0" pillow pandas lxml
```

호환성을 위해 PyMuPDF는 1.24.0 미만 버전 설치를 권장합니다.

🖥️ 2. 화면 구성: 한눈에 파악하기

프로그램은 크게 세 개의 탭과 하단의 공통 기능 영역으로 나뉩니다.

① 파일 바꾸기 (핵심 기능)

문서를 변환하는 대부분의 작업이 이루어지는 메인 공간입니다. 변환할 파일을 고르고, 저장할 위치와 형식을 지정합니다.

② 바꾸기 설정 (고급 옵션)

변환 방식을 더 세밀하게 조정하고 싶을 때 사용합니다. 예를 들어, 여러 파일을 하나의 파일로 합치거나, 텍스트를 특정 글자 수로 강제 분할할 수 있습니다.

③ 파일 합치기 (보조 기능)

문서 변환과 별개로, 폴더 안에 흩어져 있는 여러 개의 코드 파일(`.py` , `.c` 등)이나 텍스트 파일들을 하나의 파일로 간단하게 합쳐주는 유틸리티입니다.

④ 하단 공통 영역

- **실행 버튼:** [바꾸기 시작] , [합치기 시작] 등 실제 작업을 명령하는 버튼입니다.
- **진행 바:** 작업이 얼마나 진행되었는지 시각적으로 보여줍니다. 대용량 파일 변환 시 유용합니다.
- **로그 창: 가장 중요한 창입니다!** 프로그램이 현재 어떤 작업을 하고 있는지, 오류가 발생했는지 등 모든 과정을 실시간으로 알려줍니다. **"프로그램이 멈춘 것 같을 땐 가장 먼저 로그 창을 확인하세요."**

3. 핵심 기능: 문서 변환 완벽 가이드 (파일 바꾸기 탭)

가장 많이 사용하게 될 문서 변환 기능입니다. 아래 순서대로 따라 해보세요.

1단계: 변환할 파일 선택하기 (2가지 방법)

1. 하나씩 또는 여러 개 파일 직접 선택 (기본)

- [파일 선택] 버튼을 누르세요.
- 열리는 파일 탐색기에서 변환하고 싶은 `EPUB` , `PDF` , `HTML` 파일을 선택합니다.
- **꿀팁:** `Ctrl` 키를 누른 채로 클릭하면 여러 파일을 동시에, `Shift` 키를 누르면 연속된 파일들을 한 번에 선택할 수 있습니다.

2. 폴더 전체 선택

- [폴더 선택] 라디오 버튼을 클릭하세요.
- [찾아보기] 버튼으로 문서들이 들어있는 폴더를 선택합니다.
- [폴더 내 모든 ... 파일 찾기] 버튼을 누르면, 해당 폴더는 물론 그 안의 모든 하위 폴더까지 살살이 뒤져 지원하는 형식의 파일을 모두 찾아 목록에 추가해줍니다.

2단계: 결과물 저장 위치 지정하기

- [출력 설정] 부분의 [찾아보기] 버튼을 눌러, 변환된 파일들을 저장할 폴더를 지정합니다.

3단계: 출력 포맷 결정하기

어떤 결과물이 필요한지에 따라 3가지 형식 중 하나를 선택하세요.

- 🏆 **JSON (강력 추천)**
 - 이것이 **Lexi Convert**의 존재 이유입니다. 문서의 구조(제목, 소제목, 문단, 표, 참고문헌 등)를 완벽하게 보존하는 **데이터** 형식입니다.
 - **용도:** 개발, 데이터 분석, AI 모델(GPT 등) 학습 데이터 생성에 가장 이상적입니다.

- 📄 마크다운 (.md)
 - 사람이 읽기 매우 편한 문서 형식입니다. `Notion`, `Obsidian`, `Github` 등에서 활용하기 좋습니다.
 - 표나 제목 등이 깔끔하게 정리되어 출력됩니다.
- 📄 텍스트 (.txt)
 - 오직 순수한 글자만 필요할 때 사용합니다. 모든 서식과 구조가 제거된 텍스트만 출력됩니다.

4단계: 옵션 이해하고 사용하기

- "문서 구조 무시하고 텍스트만 강제 분할" (주의! ⚠️)
 - 이 옵션을 체크하면, Lexi Convert의 가장 강력한 기능인 '구조 분석'을 포기하는 것과 같습니다.
 - 체크 시, 앞서 분석한 문서 구조를 모두 무시하고, 모든 내용을 하나의 긴 텍스트로 합친 뒤 단순히 글자 수에 맞춰 잘라냅니다.
 - 언제 사용하나요?: 구조가 없는 일반 소설책 같은 파일을 다룰 때만 사용하세요. 리포트나 웹페이지 같은 구조적인 문서에는 절대 체크하지 않는 것을 권장합니다. (기본값: 체크 해제)

5단계: 변환 시작!

- 모든 설정이 끝났다면, 힘차게 `[바꾸기 시작]` 버튼을 누르세요!
- 하단의 로그 창과 진행 바를 통해 작업 과정을 실시간으로 확인하실 수 있습니다.

💎 4. 산출물(JSON) 200% 활용하기

JSON으로 출력된 결과물은 다음과 같은 체계적인 구조를 가집니다.

- `"metadata"` : 파일명, 크기, 변환 일시 등 문서의 기본 정보가 담겨있습니다.
- `"sections"` : 문서의 가장 큰 주제들(`<h2>` 태그 기준)이 배열로 담겨있습니다. 각 `section` 은 다음과 같이 구성됩니다.
 - `"title"` : 대주제의 제목입니다.
 - `"content"` : 대주제에 직접 속한 내용(문단, 표)이 배열로 담깁니다.
 - `"subsections"` : 대주제 아래의 소주제들(`<h3>` 태그 기준)이 배열로 담깁니다. 각 `subsection` 은 다음과 같이 구성됩니다.
 - `"subtitle"` : 소주제의 제목입니다.
 - `"content"` : 소주제에 속한 내용(문단, 표)이 배열로 담깁니다.
- `"references"` : 문서 하단의 참고문헌 목록이 별도로 추출되어 담깁니다.

이 구조 덕분에, "Executive Summary 섹션의 Key Change Drivers 소주제에 있는 세 번째 문단"과 같이 원하는 정보에 정확하고 쉽게 접근할 수 있습니다.

궁금한 점이 있다면 언제나 다시 찾아주세요! Lexi Convert와 함께 즐거운 데이터 변환의 세계를 경험하시길 바랍니다.