데이터 조회 프로그램

SKN Family AI Camp 2기 : 최종 프로젝트 2조



| 프로젝트 주제 | 경제 지식이 없는 아동 및 청소년을 대상으로 경제 지식을 쉽게 설명하는 Al 인플루언서 블로그 관리 Tool |
|-----------------|--|
| 원본 데이터 (텍스트) | [한국은행 제공 경제교육용 PDF 문서] - <u>알기 쉬운 경제 이야기 (일반인용)</u> - <u>경제금융용어 700선</u> |
| | 알기 쉬운 경제 이야기는 경제 교양서처럼 서술되었고, 경제금융용어 700선은 사전처럼 서술되었습니다. 따라서 각각의 데이터는 형식 및 용도 면에서 큰 차이가 있으므로 데이터 고유 특성을 고려하여 전처리 방식을 결정해야 합니다. |
| | [알기 쉬운 경제 이야기 (일반인용)] |
| 데이터 전처리 | 1. 날개 부분에 붙은 '용어 설명'과 '페이지 번호' 제거된 부분을 이미지 형태로 추출합니다. |
| 과정 | 2. OCR을 사용하여 이미지를 텍스트로 추출합니다. |
| | |
| | [경제금융용어 700선] |
| | [0.41000.1 100.5] |
| | 1. pdf 라이브러리를 사용하여 텍스트 추출합니다. |
| | |
| | 1. pdf 라이브러리를 사용하여 텍스트 추출합니다. |
| | 1. pdf 라이브러리를 사용하여 텍스트 추출합니다. 2. 페이지 번호 및 사전 인덱스(한글 자음) 삭제합니다. |
| | 1. pdf 라이브러리를 사용하여 텍스트 추출합니다. 2. 페이지 번호 및 사전 인덱스(한글 자음) 삭제합니다. [Vector DB] |
| DB 사용 용도 | 1. pdf 라이브러리를 사용하여 텍스트 추출합니다. 2. 페이지 번호 및 사전 인덱스(한글 자음) 삭제합니다. [Vector DB] AI 인플루언서 블로그 포스팅 작성 시, 제공 정보의 정확도를 높이기 위해 RAG를 사용합니다. |
| DB 사용 용도 | 1. pdf 라이브러리를 사용하여 텍스트 추출합니다. 2. 페이지 번호 및 사전 인덱스(한글 자음) 삭제합니다. [Vector DB] Al 인플루언서 블로그 포스팅 작성 시, 제공 정보의 정확도를 높이기 위해 RAG를 사용합니다. 이를 위해 전처리가 끝난 각각의 pdf 파일을 서로 다른 Vector DB에 넣습니다. |
| DB 사용 용도 | 1. pdf 라이브러리를 사용하여 텍스트 추출합니다. 2. 페이지 번호 및 사전 인덱스(한글 자음) 삭제합니다. [Vector DB] Al 인플루언서 블로그 포스팅 작성 시, 제공 정보의 정확도를 높이기 위해 RAG를 사용합니다. 이를 위해 전처리가 끝난 각각의 pdf 파일을 서로 다른 Vector DB에 넣습니다. |
| DB 사용 용도 | 1. pdf 라이브러리를 사용하여 텍스트 추출합니다. 2. 페이지 번호 및 사전 인덱스(한글 자음) 삭제합니다. [Vector DB] Al 인플루언서 블로그 포스팅 작성 시, 제공 정보의 정확도를 높이기 위해 RAG를 사용합니다. 이를 위해 전처리가 끝난 각각의 pdf 파일을 서로 다른 Vector DB에 넣습니다. 두 종류의 블로그 글 스타일(일반 경제 교양 포스팅, 오늘의 경제 단어 포스팅)을 고려하여 위와 같은 방향으로 접근했습니다. |

Vector DB

구조

01c4613f-fb75-4514-9805-e06b4dae62c9 하기 때문에 정책수립자가 당초 생각했던 결과가 나타나 지 않는 01f9c3ae-b58d-4a6c-827a-ddc2ef3469a2 사람을 뜻하는 것이 아니고 매월 15일이 속한 1주일 동안에 수입 02440f51-ca21-4f79-8ad8-d12e5bcb1d12 볼 것 없이 가격과 생산량을 결정할 수 있기 때문입니다. 기업이 027e1a16-334f-4f05-9e39-082ad0258f85 나 현실에서는 어느 누구도 고속도로나 강에 대해 소유권을 가지 4 032f7a68-b749-4d30-a651-de037976266e 차가 확대되고 있다면 함후 경제가 좋아질 것이라고 시장이 기대 6 04515a77-3419-4b75-8350-3835f14dd6ce 음을 나타냅니다. 이러한 물가지수는 상품별로 중요한 정도에 따 054e9293-b14d-42b4-849f-c59685934fb3 준다면 아무 이득도 가져다주지 못할 것이기 때문입니다. 경쟁기 05ea1712-7669-4bda-8ae3-2254ba7ceb93 때문에 체감물가와 지수물가 사이에 차이가 나기도 합니다. 이 9 06e0344d-6d37-48f8-9488-41153b129f2b 이처럼 한 나라의 경제에서 중요한 위치를 차지하는 임금은 어떻 10 09765a1b-de9d-4b30-96e5-c76811209eca 얼굴 환경문제는 사람들이 문명사회를 이루어 주변의 환경을 이 11 097d496d-c77c-48b8-8287-5d3845c27242 로 생산한 재화와 서비스의 가치'란 각 생산단계에서 추가된 가치 12 0a3ef704-32c7-461a-9146-c259ff2045e7 다. 우선 경제안정을 위해 정부가 어느 정도 개입할 것인지에 대한 14 0afb6116-574a-42f7-b474-2db840dd8b08 과학기술자나 의사, 변호사와 같은 소위 지식노동자들의 지적 노 15 0c8e6d23-370b-4fbe-a59c-f87361ebe144 의한 결과라고 비유했습니다. 일반적으로 경제성장은 자본과 5 0cecd687-b1f0-4252-a1d2-1cf065a79f87 것입니다. 지리산 속의 반달가슴곰,왜 볼뺑할까? 우리 주위에 17 Oda6c5f1-4058-43a5-afc1-9273107ab639 높은 것이 일반적입니다. 공급량은 목점에 비해 많지만 결정시절 18 1080d91e-3b04-4e7b-9551-47f2d270eee5 정보가 불완전한 상황에 서는 여러 가지 방법으로 필요한 정보를 19 14b2a59b-d228-45d3-ab21-e102aafd7192 불수 있습니다. 쉬운 예로 국내 경제에서 '갑' 이라는 사람이 상다 20 15b3cf0b-3db6-4f75-b9c0-4b8c82e3f9a4 . 이 중에서도 가장 중요한 목표는 불가안정이라고 할 수 있습니!

자주 바뀌는 성질의 데이터는 아니라고 판단하여 RDB에 저장하기로 결정했습니다.

< 알기 쉬운 경제 이야기 (일반인용) Vector DB >



< 경제금융용어 700선 Vector DB >

Django 내부 RDB 구조

