

Quote Distortion Detection Service

인용문 왜곡 탐지 시스템 아키텍처

⑨ 사용자 행동 (Chrome Extension)

Content Script

- 기사 DOM 분석
- 직접 인용문 추출
- 주변 문장 span 수집

Output:
Direct Quote + Context Span

Panel UI

- 결과 시각화
- 원문보기 기능
- 왜곡 점수 표시

Display:
링크, 유사도, 원문, 왜곡점수

Direct Quote + 주변 문장 span 전달

Backend Server

FastAPI + WebSocket

- Direct Quote 수신
- 파이프라인 실행
- 실시간 상태 전송

Redis

- 세션 유지
- 문맥 저장
- 진행상황 캐싱

Monitoring

- Logging
- Performance Tracking

Original Quote Retrieval Pipeline

1. 키워드 추출 (KeyBERT)

Input: Direct Quote → Output: keywords[]

2. 개체명 인식 (KoELECTRA NER)

Input: Quote Text → Output: PER/ORG 엔티티

3. 인명 정규화; 위키에 쿼리 날림 (per_name.py)

Input: PER Entity → Output: Full Name

4. 번역 (Helsinki MT)

Input: 한국어 문장 → Output: 영어 번역

5. 검색 쿼리 생성 (generate_query.py)

Input: Keywords + Name → Output: Query String

Distortion Model

왜곡 분류 모델(KoBERT와 KPF, 테스트중)

- Semantic Reduction 왜곡
- Interpretive Extension 왜곡
- Lexical Coloring 왜곡

Explainability

- 왜곡 부분 highlight
- 설명 정보 포함

Final Output:
3-Axis 점수 + 근거

Quote Distortion Detection Diagram

6. Google Search API 호출

Input: Query String → Output: Search Results[]



7. Span 유사도 평가 & Alignment

Input: Results[] → Output: Best Match + Score

Method: SBERT, cosine 유사도

Backend Data Layer (Faster Retrieval)

Local Speech Database

- 대표적인 정치인 연설 녹취록 저장
- 예: Donald Trump speeches

Fields:

- id, speaker, date
- span, embedding vector
- source metadata

Type:
PostgreSQL / MongoDB

Vector Search Engine

- Semantic similarity search
- 빠른 span 매칭
- Pre-hit 체크 수행

Input/Output:

Input: span embedding
Output: candidate spans + similarity scoresTechnology:
FAISS / Elasticsearch

Redis Cache Layer-필요시 넣을 예정

- 매칭된 span 캐싱
- 외부 검색 결과 저장
- 반복 쿼리 가속화

Features:

- TTL-based eviction
- Frequently matched results

Acceleration:
10x faster for cached queries

Data Flow & Logic

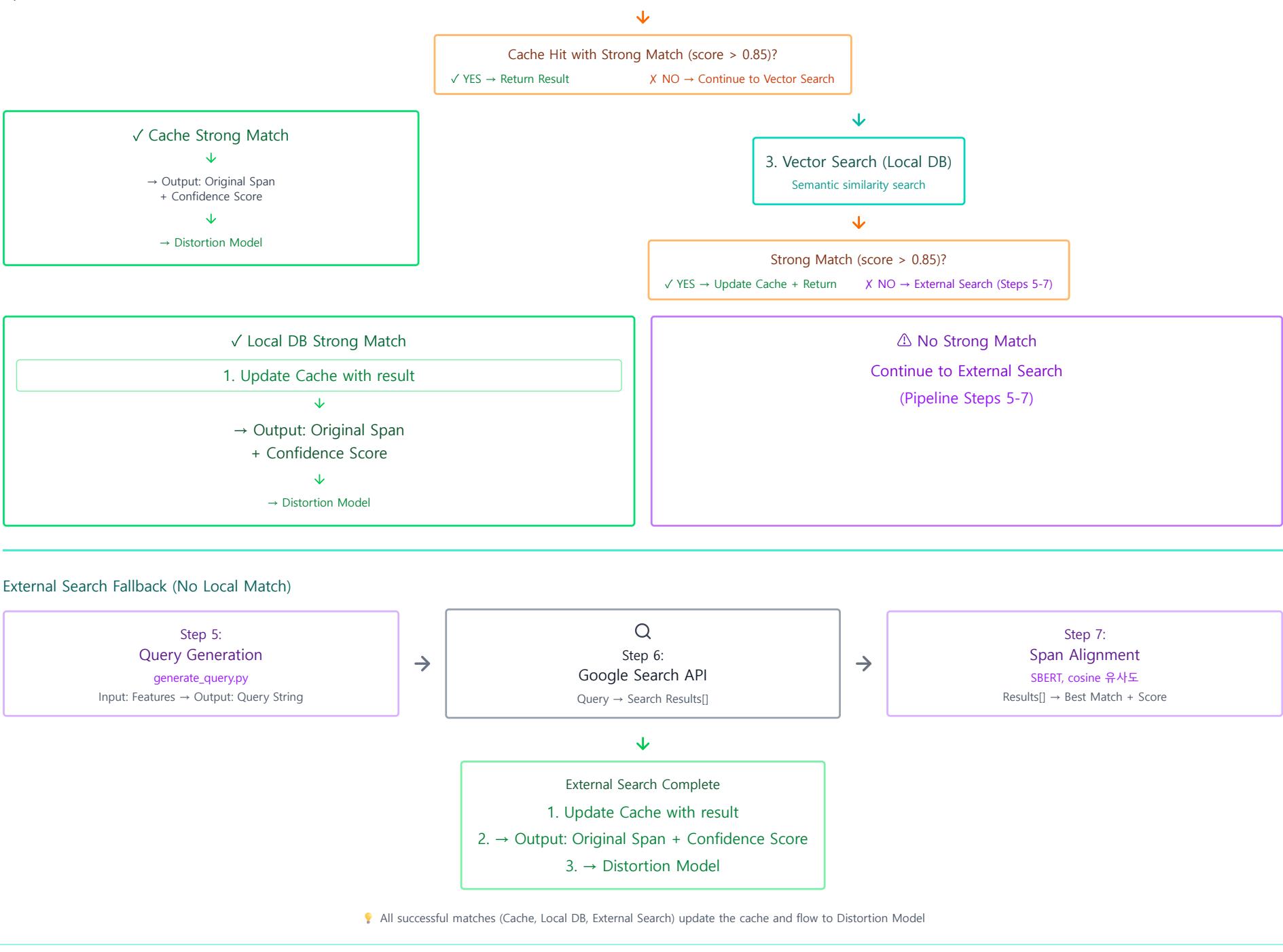
Preprocessing (Pipeline Steps 1-4):

1. Keywords
(KeyBERT)2. NER
(PER/ORG)3. Normalize
(Names/Dates)4. Translate
(Helsinki)

Output: Normalized Names + Dates + Keywords

1. Receive Extracted Features
Names, Dates, Keywords2. Check Cache
Query by extracted features

Quote Distortion Detection Diagram



 외부 의존성 (External APIs) Google Search API

검색 결과 수집

 Wikipedia

인명 정보 조회

 Helsinki MT

한영 번역 서비스

최종 결과 반환: 원문 링크, 유사도 점수, 원문 span, 직역, 왜곡 3점수

사용자 플로우

1. 사용자

기사 페이지 방문

2. 확장 기능

인용문 자동 감지

3. Panel 표시

분석 결과 확인

4. 원문 보기

Highlight Jump Link

범례 (Legend)

 Frontend Layer Backend Services Data Layer (Fast Retrieval) AI Pipeline External APIs