

Web as DB RAG

2025.06.19

최현영

Contents

01	배경
02	비교
03	구조
04	질문 확장
05	웹 서칭
06	Retrieval
07	Chunk DB
08	샘플
09	결론

01 배경

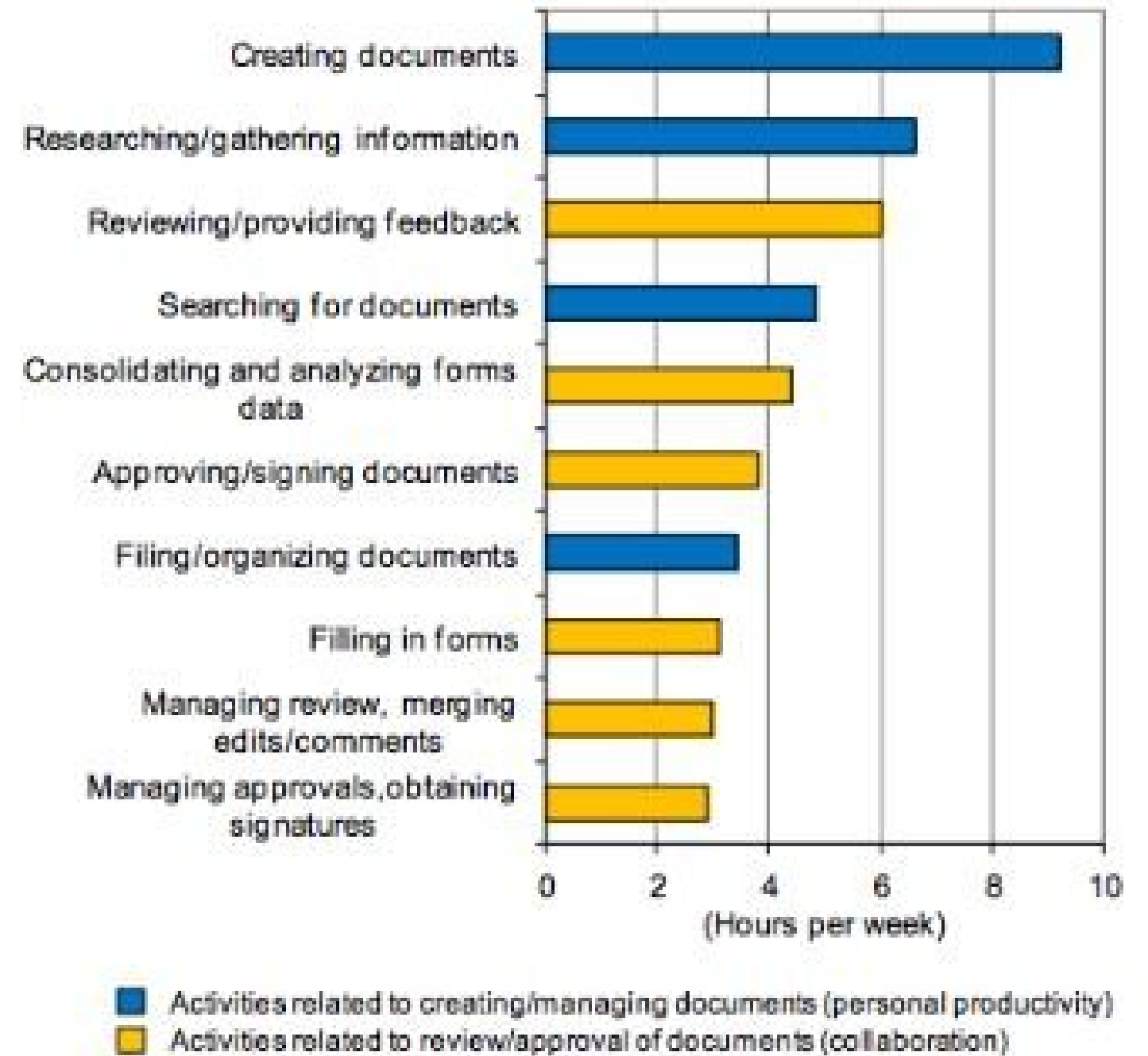
주제 소개

LLM의 한계인 최신성 부족과 환각 문제를 해결하기 위해 RAG가 등장했습니다. 웹이라는 거대한 비정형 DB에서 정보를 능동적으로 검색하고, 이를 RAG 컨텍스트로 연결하는 방식을 설계했습니다.

문제 제기

1. GPT web browsing과 무엇이 다른가?
→ 검색어 생성·출처 제어가 불가능하며, 결과 재활용이 어렵다.
2. Web을 DB로 사용할 때 문제점은?
→ 정보의 신뢰성, 적절성 확보가 어렵고 검색 비용이 크다.

Q. Thinking about your typical workweek, how many hours would you say you spend on the following document-related activities?



02 비교

our service

검색어 확장 및 출처 제어 기능

유저에 맞는 context DB로 재사용

유저 보유 포트폴리오, 관심사, 검색기록 반영

추론 시간 상대적으로 긴 편

VS

GPT web browsing

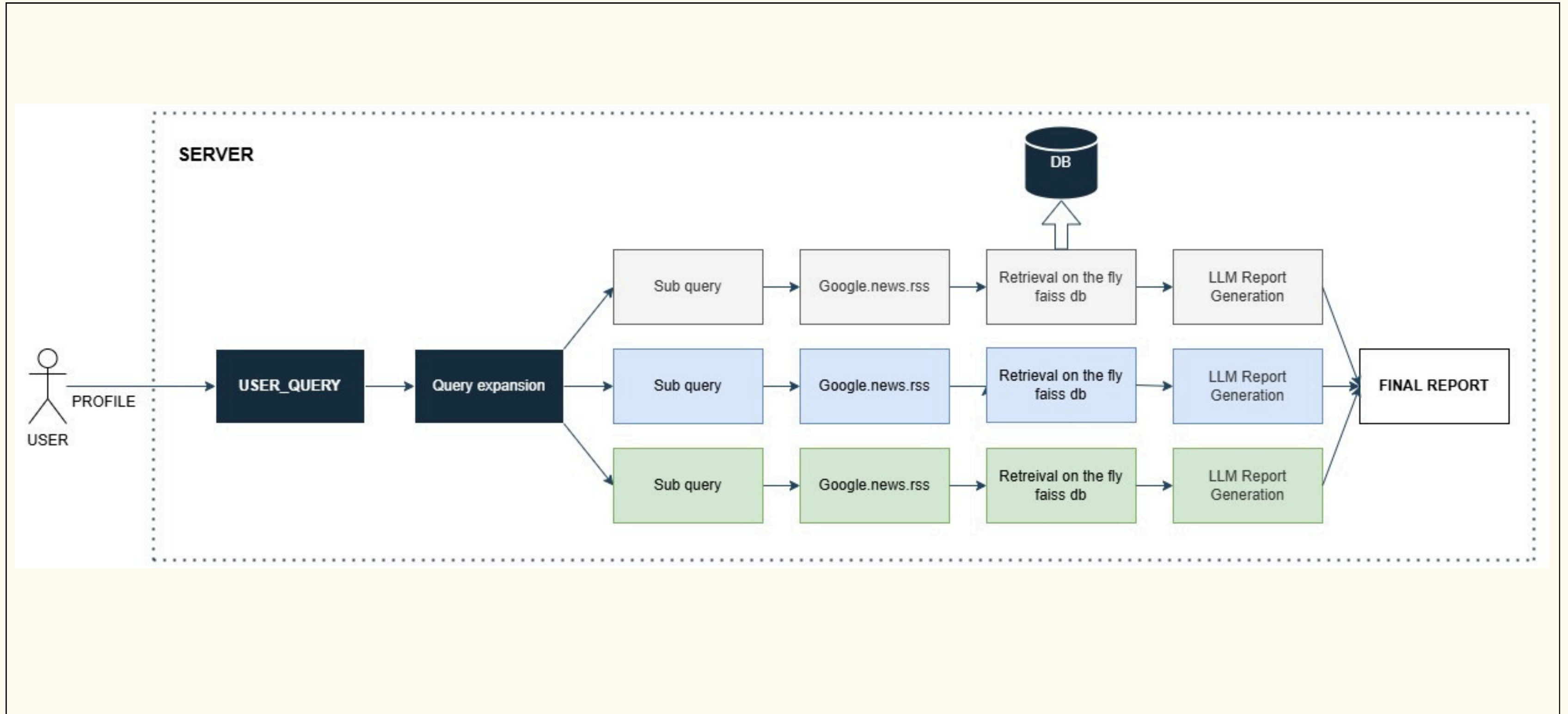
검색어 생성 로직 및 검색어 출처 제어 불가

context로 사용된 chunk 재사용 불가

사용자 프로파일 반영 제한적

추론 시간 상대적으로 짧음

03 구조





검토 전략

1. 구체적인 시간 표현 강제 여부
2. 사용자 프로파일 반영 강도
3. 서로 다른 관점의 질문으로 확장 여부

고려 사항

1. 질문이 웹 서칭에 적합한 형태로 정제되는가?
2. 최신성이 있는 검색 결과가 나오는가?
3. 세부질문이 사용자의 의도와 프로파일을 반영하여 잘 확장되는가?
4. 세부질문 출력이 원하는 형태로 제어되는가?

선택 전략

1. 구체적인 시간 표현 제약
2. 프롬프트 수준의 제어
3. few-shot 예제
4. 현재 시점 프롬프트에 명시

METRIC

1. 생성보고서 GPT 정성평가
2. 특정 case 위주로 최신정보 포함 여부(ex. 2025 1분기 실적, 체리자동차 사업 수주)
3. 직접 검색한 결과와 비교



검토 전략

1. 구체적인 시간 표현 강제 여부
2. 기사의 출처 제어
3. metadata 관리

고려 사항

1. 최신성이 있는 검색 결과가 나오는가?
2. 검색 결과가 지나치게 중복되거나 연관성이 떨어지는 않는가?
3. 검색 시간비용을 고려했을 때 적절한 검색어 개수 및 내용을 읽어올 기사 개수는 적절한가?
4. metadata 관리가 용이한가?

선택 전략

1. RSS URL 파라미터 설정
2. publisher metadata 추출 및 필터링 로직 추가
3. 검색 지역, 검색 언어 분기 추가

METRIC

1. 특정 case 위주로 Retrieval 결과 비교
2. 직접 검색한 결과와 비교

06 Retrieval



검토 전략

1. embedding model 결정 및 score metric 결정(L2 vs cosine sim.)
2. MMR 도입 여부
3. reranker 도입 여부 및 모델 결정
4. 시간 정보 활용 여부

고려 사항

1. Retrieval 결과가 최신성, 적절성이 있는가?
2. 정보가 중복되어 정보밀도가 떨어지지 않는가?
3. 보고서 생성 LLM을 고려할 때 context 개수와 chunking size는 어느정도가 적절한가?
4. cross-encoder가 time reasoning이 가능한가?

선택 전략

1. faiss 기반 cosine similarity - intfloat/multilingual-e5-base
2. 다양성 확보를 위해 MMR 알고리즘 사용
3. BAAI/bge-reranker-v2-m3 사용
4. rerank 단계에서 시간정보 prefix로 추가

METRIC

1. 조건 분기별로 특정 case로 정성평가
2. 소모 시간 및 retrieval 결과 비교

07 Chunk DB



검토 전략

1. chunk DB의 사용 목적 및 확장성
2. web searching 시간 비용 감소
3. chunk DB 내 데이터 최신성 관리

고려 사항

1. web searching 비용 감소를 위해 DB를 활용할 때
가능성과 장단점은 무엇인가?
2. 서비스 데이터가 모여 chunkDB를 구성할 때 확장가
능한 서비스는 없는가?
3. chunkDB의 품질을 어떻게 관리할 것인가?

선택 전략

1. 보고서 생성에 활용한 chunk 재사용
2. web search 필요성 판단 로직 추가 및
테스트 : 반감기 기반 지수감쇠 scaling
3. timestamp metadata 생성 → DB에서
시간 조건 기반 필터링 용이

PROFILE

- LLM prompt context로 활용
 1. 보유 포트폴리오
 2. 관심사
 3. 이전 쿼리 기록
 4. 선호 보고서 양식

08 샘플

main_query : 요즘 LG에너지솔루션 상황 어때?

sub_query1 : LG에너지솔루션 주가 동향

sub_query2 : 배터리 산업 트렌드

sub_query3 : LG에너지솔루션 사업 확장

1. 최신시점의 정보
2. 구체적인 키워드
3. 깊은 추론이라 볼 수 없음, 사실 위주의 나열
4. LG엔솔은 profile 정보에 관련점이 크게 없기 때문에 큰 반영되지 않음.

```
user_profile = {
  "user_id": "abc123",
  "portfolio": ["삼성전자", "네이버", "두산로보틱스"],
  "interests": ["AI 반도체", "핀테크", "로봇"],
  "query_history": ["네이버 실적", "삼성전자 하락 이유"],
  "alert_schedule": "매일 18:00",
  "user_preference" : ""[제목]: 보고서 제목
[핵심 요약]:\n
- 이 세부 조사항목에 대한 요약된 결론을 한두 문장으로 정리하세요.

[배경 및 주요 사실]:\n
- 관련된 배경 정보, 수치, 시기 등을 요약하세요.

[세부 분석]:\n
- 원인, 영향, 관점 차이 등 심화 내용을 서술하세요.
- 다양한 관점이 존재한다면 관점의 출처와 함께 기술하세요.

[향후 전망 또는 시사점]:\n
- 이 이슈가 앞으로 어떤 영향을 미칠 수 있는지 서술하세요.
- 관련 기업이나 산업에 대한 함의도 포함할 수 있습니다.""
}
```

08 샘플

[제목]: LG에너지솔루션 주가 동향 및 향후 전망 분석

[핵심 요약]:

- LG에너지솔루션의 주가는 최근 약세를 보이고 있으며, 이는 미국 인플레이션 감축법(IRA) 세액공제 조기 폐지 우려 및 북미 전기차 시장 불확실성 때문으로 분석된다. 그럼에도 불구하고 북미 에너지 저장장치(ESS) 시장의 성장 가능성이 긍정적인 전망을 제시하고 있다.

[배경 및 주요 사실]:

- LG에너지솔루션의 주가는 지난 16일 이후 공모가 30만원을 하회하며 28만원 수준까지 하락했다. 이는 상장 후 최저가 경신이다.
- 미국 공화당 지도부가 IRA 세액공제를 2028년까지 폐지할 것이라는 소식이 주요한 주가 하락 요인으로 작용했다.
- 올 1분기 매출 6조2650억원, 영업이익 3747억원을 기록했으나, 세액공제 혜택 없이는 적자 상태이다.
- 북미 ESS 시장을 통해 새로운 성장 동력을 확보하고자 하며, LFP 배터리 셀 양산을 최근 시작했다.

[세부 분석]:

- 미국 IRA의 조기 폐지 가능성은 LG에너지솔루션을 포함한 국내 이차전지 기업들의 수익성에 부정적 영향을 미칠 가능성이 크다. 세액공제 폐지로 인한 세제 혜택이 감소되면 영업이익에 직접적인 부정적 영향을 줄 수 있다. 이는 외국인 및 기관 투자자들의 매도세로 이어질 수 있다.
- 반면, LG에너지솔루션은 북미 ESS 시장으로 눈을 돌려, 배터리 수요의 새로운 장을 개척하고 있다. 신재생 에너지 및 AI 데이터센터 수요 증가에 대비해 리튬인산철(LFP) 배터리 셀을 양산하기 시작했다. ESS 시장의 성장세는 중장기적인 관점에서 기업 실적에 긍정적인 영향을 미칠 것으로 기대된다.

[향후 전망 또는 시사점]:

- 단기적으로는 미국과 EU의 정책 불확실성으로 인해 주가 변동성이 지속될 것으로 보인다. 이는 특히 IRA의 조기 폐지 가능성이 현실화될 경우 더욱 심화될 것이다.
- 그러나 중장기적으로 봤을 때 북미 ESS 시장의 확대에 의해 LG에너지솔루션이 성장 모멘텀을 회복할 가능성이 높다. 특히 AI 및 데이터센터 시장 성장에 따른 추가 수혜가 예상된다.
- 투자자들은 시장의 단기적인 불확실성 속에서 장기적인 성장 잠재력을 고려한 투자 전략이 필요하다. LG에너지솔루션은 에너지 저장장치로 포트폴리오 다변화를 추진하고 있어 장기적인 시점에서 긍정적인 성과를 낼 가능성이 있다.

[제목]: LG에너지솔루션 및 배터리 산업 트렌드 분석

[핵심 요약]:

- LG에너지솔루션은 리튬망간리치(LMR) 배터리 기술 개발에 집중하고 있다. 이는 니켈과 코발트 비용을 낮추고 망간 비율을 높여 가격경쟁력을 확보한 신제품으로, 미국 시장을 타겟으로 하고 있다.

[배경 및 주요 사실]:

- LG에너지솔루션은 전 세계 배터리 산업 변화를 따라잡기 위해 리튬망간리치(LMR) 배터리의 2028년 양산을 계획하고 있다. LMR은 망간 비율을 60~65%로 높여 기존 배터리 대비 비용을 절감하고 가격경쟁력을 높이는 기술이다.
- 글로벌 배터리 시장은 소용이온 배터리를 중심으로 변화하고 있다. 중국 CATL이 소용이온 배터리 양산을 앞둔 것은 배터리 가격을 크게 낮출 가능성이 있다.

[세부 분석]:

- LG에너지솔루션은 중국 시장의 새로운 트렌드인 소용이온 배터리 대응책으로 미국 시장에 적합하다는 평가를 받고 있는 LMR 배터리를 통해 대응하고 있다. 니켈과 코발트의 낮은 비율로 인해 원재료 비용을 절감하고, 가격 경쟁력을 강화했다.
- 한국 시장에서는 현재 중국 CATL의 소용이온 배터리가 주목받고 있다. 소용이온 배터리는 나트륨을 주원료로 사용해, 리튬이온 배터리에 비해 원재료 비용이 훨씬 낮고 공급망 리스크가 적은 것이 장점이다.

[향후 전망 또는 시사점]:

- 향후 몇 년간 LG에너지솔루션은 LMR 배터리를 활용한 미국 시장 진입을 가속화할 것으로 보인다. 이 기술은 리튬 기반 기술과의 경쟁에서 차별화할 수 있는 가능성을 제공할 것이다.
- 한국 배터리 업체들이 소용이온 배터리와 같은 차세대 기술에 대한 개발을 강화해야 할 필요성이 크다. 소용이온 기술이 상용화되면 가격경쟁력이 크게 향상될 가능성이 있기 때문이다. 또한, 이는 EV 시장의 확대와 함께 배터리 수요 증가에 기여할 수 있다.
- 배터리 및 전기차 산업의 변화는 LG에너지솔루션이 지속가능성과 가격경쟁력을 지닌 새로운 기술 개발을 계속 추진해야 함을 시사한다. 경쟁이 심화되고 있는 글로벌 배터리 시장에서 생존을 위해서는 기술 혁신이 필수적이다.

08 샘플

[제목]: LG에너지솔루션의 사업 확장 및 혁신 동향

[핵심 요약]:

- LG에너지솔루션은 사업 포트폴리오 확장을 통해 에너지 솔루션 분야에서 차별화된 가치를 실현하고 있다. 다양한 산업 분야와의 협력과 기술 혁신은 전기차 시장에 대한 의존도를 낮추고 안정적인 수익 기반을 확보하는 데 기여하고 있다.

[배경 및 주요 사실]:

- LG에너지솔루션은 프랑스 메탈 재활용 기업 DBG와 합작법인을 설립하고 배터리 리사이클링 사업에 진출했다. 이는 유럽의 배터리 재활용 규제에 대한 선제적 대응을 가능하게 한다.
- 미국 태양광 전기차 스타트업 애펜타 모터스에 원통형 배터리를 독점 공급하고, 두산밥캣과 협력하여 소형 건설 장비 배터리 팩 개발을 추진하고 있다.
- LG에너지솔루션은 연구개발(R&D) 비용을 계속하여 증가시키고 있으며, 2025년 1분기에는 전년 대비 21% 증가한 3074억 원을 투자했다.

[세부 분석]:

- LG에너지솔루션의 신사업 확장은 글로벌 시장에서의 경쟁력을 높이는 핵심 전략이다. 특히, 배터리 리사이클링과 소형 건설 장비용 배터리 개발은 에너지 시장의 지속가능성과 유연성을 높인다.
- 다양한 산업과의 협력을 통해 단순 배터리 공급을 넘어, 기술 파트너십을 구축하고 있다. 두산밥캣과의 협력은 북미시장 점유율 확대에 중요한 역할을 할 것으로 보이며, 이는 LG에너지솔루션의 기술적 응용능력을 보여주는 사례다.
- LG에너지솔루션의 46시리즈 배터리는 안정성과 효율성을 인정받으며 미국 리비안을 포함한 여러 전기차 제조사의 주요 선택지가 되고 있다. 이는 고성능 배터리 시장에서의 기술적 우위를 강화하는 결과로 이어진다.

[향후 전망 또는 시사점]:

- LG에너지솔루션의 기술 및 사업 다변화 노력은 지속 가능한 성장과 글로벌 시장 점유율 확대에 기여할 것으로 예상된다. 특히 환경 규제 강화에 선제적으로 대응하는 리사이클링 사업은 향후 중요한 수익원으로 발전할 수 있다.
- LG에너지솔루션은 북미 및 유럽의 전기차 배터리 수요 증가에 대응하여, 기존 생산 인프라를 최적화하고 새로운 시장 기회를 효과적으로 활용할 것이다.
- 다양한 산업과의 협력은 로봇 및 AI 분야로의 확장을 가능하게 하여, 향후 글로벌 자동화 시장에서도 중요한 역할을 할 가능성이 크다. 이는 사용자의 관심사와 관련된 로봇 분야로의 잠재적인 시장 진출을 의미한다.

LG에너지솔루션의 지속적인 사업 확장과 기술 혁신 노력은 시장의 다양한 변화에 대응하는 능력을 강화하며, 전기차 외의 다양한 에너지 애플리케이션에서의 성장을 지속적으로 지원할 것이다.

08 샘플

주가 및 밸류에이션

- 현재 주가: 약 **292,000원** (6월 19일 기준) tradingview.com +6
- 최근 큰 폭의 하락세(-12 ~ 22%)를 보였으나, 주간 기준으로 소폭 상승(-0.17%)
- 시가총액: 약 **68조 원** 수준 marketscreener.com +7
- PER(2025년 기준) 약 137배, EV/Sales 약 3.2배 marketscreener.com
- 애널리스트 평균 12개월 목표주가는 약 **397,000원**으로 현 주가 대비 약 +36% 상승 여력 stockopedia.com +5

1분기 실적 및 전망

- 매출: **6조 2650억원** (+2.2% YoY)
- 영업이익: **3747억원** (+138% YoY) newswire.co.kr
- 순손실: 약 **1460억원** 기록, 전년 대비 적자폭 확대
- 2분기 매출은 미·중 금리·관세 불확실성 영향으로 다소 둔화 예상, 이에 따라 투자 규모도 **2025년 30% 축소** 계획

전략 및 기술 동향

- **LG·GM**이 미국에서 LMR(리튬망간리치) 저비용 차세대 배터리 생산 계획(2028년부터) reuters.com
- 미국과 프랑스에서 ESS용 LFP 배터리 대량 생산 개시 lgensol.com
- 최근 중국 **체리(Chery)** 완성차向 6년간 8GWh 공급 계약 체결 newsroomindex.com +1

1.수치 정보

2.체리 자동차 사업 수주



Future work

1. web search 단계에서 소규모 기업의 경우 시간 제약으로 인해 검색이 잘 수행되지 않음 → fallback 구현
2. 보고서 작성에 사용된 chunk를 저장하여 보고서 보강 및 사용자 상호작용 기반 질의응답 인터페이스 구현
3. deep reasoning 기능 추가 : 질문의 폭을 확장시키는 방향이 아닌 step-by-step으로 깊이를 확장시키는 추론단계, adversarial한 agent 사용하여 서로의 논리를 보강하는 식으로 추론 진행
4. 웹 검색을 비동기 요청 기반으로 병렬 처리 / 서비스 환경에서 웹 검색 traffic 관리
5. user profile 정보 효과적인 활용 방안

문제점

- query_expand 단계가 보고서 품질을 좌우함.
- 단일 키워드로 제약을 줄 경우 검색품질은 올라가지만 데이터 검색의 폭은 상대적으로 감소
- 따라서, 단일 키워드의 경우 1~2일 전의 키워드와 직접적인 연관이 부족하지만 유저 쿼리 기준에서 중요한 정보가 포함 안되는 경우가 발생
- query_expand 결과의 재현성이 부족함.

Reference

1. IDC's Information Worker Survey
2. https://www.proprofskb.com/blog/workforce-spend-much-time-searching-information/?utm_source=chatgpt.com

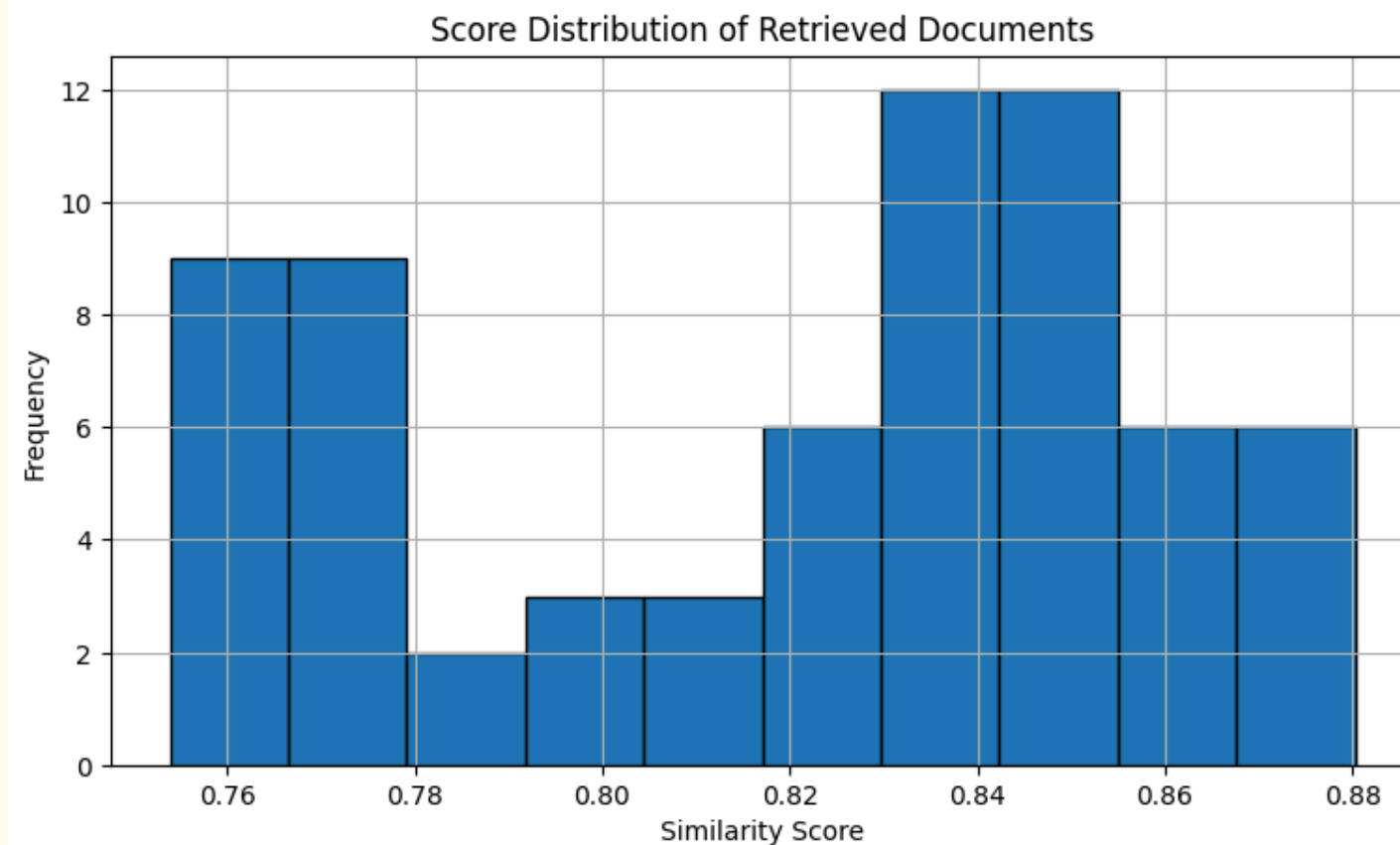


KPT

KEEP :

PROBLEM : metric이 설득력을 가지지 못한다.

TRY : 검색 품질 정량지표 도입, retrieval query case별 키워드 포함 여부



user_case1 : LG에너지솔루션, keyword : 캐즘, 체리자동차, LFP 배터리
user_case2: 삼성전자, keyword : AI 반도체, 폴더블폰, hbm4, 엔비디아
user_case3: 인텔리안테크, keyword : VSAT, 저궤도, 게이트웨이 안테나

Q&A