

프로젝트 기획서

SKN04-오정연

1. 프로젝트 개요

- A. 프로젝트 명: Paper Map (논문 지도, 종이 지도)
- NLP 논문을 읽는 사람에게 필요한 학습 보조 도구
 - 학생, 대학원생, 연구자 등 논문을 읽는 사람을 타겟으로 함
- B. 제작기간: 2024년 12월 16일 ~ 2024년 1월 31일

2. 프로젝트 배경

- A. 최근 동향
- 학술 정보의 폭증
 - 최근 딥러닝, 자연어 처리 (NLP) 분야의 급격한 발전으로 매년 발표되는 학술 논문 수가 기하급수적으로 증가하고 있음
 - 예를 들어, NLP 분야 주요 학회 (ACL, EMNLP, NAACL 등)에서의 게재되는 논문 수 매년 증가

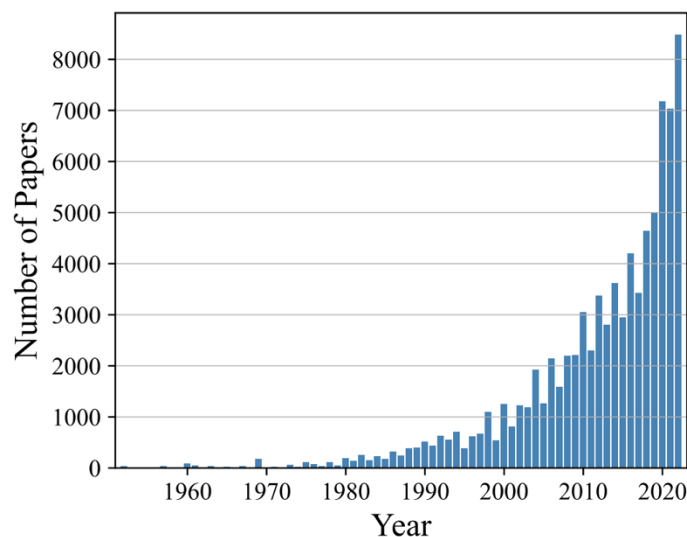


그림 1: ACL Anthology에 수록된 연도별 논문 수 분포

- 문헌 검색 시간 증가

- 관련 연구의 폭발적 증가로 인해, 연구자들의 문헌 검색 부담 증가
- 2019년 "Trust in Research" 보고서에 따르면 연구자들은 2011년에 비해 10% 적게 논문을 읽지만, 문헌을 검색하는 데는 11% 더 많은 시간을 소비하는 것으로 나타남
- 급격히 늘어나는 논문량과 이를 효율적으로 정리·탐색하는 도구 필요성 증가

B. 유사 서비스 및 한계점

- ScienceON - '논문타임라인 서비스'

논문의 인용/피인용 관계를 분석해 주요 키워드를 시간순으로 보여줌

- 한계점

- 1) 주로 국내 논문 중심으로 구성되어 해외 논문 제공 규모가 제한적임
- 2) 총 2,223개의 논문 중 실제 타임라인을 열람 가능한 논문은 34개에 불과하며, 나머지 논문 접근 시 오류가 발생하는 문제가 있음



ScienceON 논문타임라인 서비스 ©한국과학기술정보연구원

- Connected Papers

논문들이 서로 어떻게 연관되어 있는지 관련 논문들을 시각화해줌

- 한계점

- 1) 키워드 중심의 검색 방식에 의존하는 경우가 많아, 특정 모델이나 기술의 기초를 마련한 주요 논문을 바로 찾는 데 어려움이 있음
- 2) 검색한 키워드와 관련된 논문의 연결성은 보여주지만 흐름은 볼 수 없음

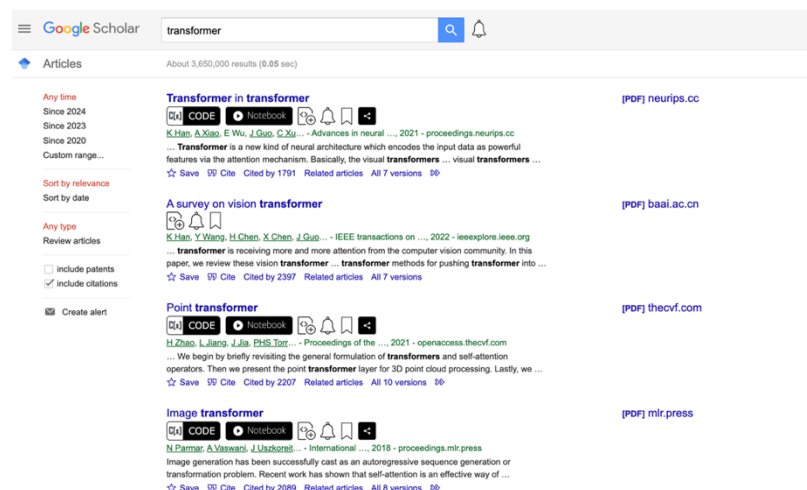


- 그 외의 논문 검색 사이트 (Google Scholar, Semantic Scholar, dblp)

학술 검색 전용 사이트들로 다양한 학술자료를 검색할 수 있음

- 한계점

- 1) 키워드 중심의 검색 방식에 의존하는 경우가 많아, 특정 모델이나 기술의 기초를 마련한 주요 논문을 바로 찾는 데 어려움이 있음
- 예를 들어, Transformer로 검색하는 경우 해당 모델의 시초 논문인 “Attention is All You Need” 논문 접근하기 어려움



C. 문제 정의

- 키워드 중심 탐색 한계
 - 기존 키워드 기반 검색 방식으로는 특정 모델이나 기술의 최초 제안을 담은 핵심 논문을 효과적으로 찾기 어려움
- 연구 흐름 파악의 어려움
 - 특정 연구 주제에 대한 논문 기반 학습 시, 일반적인 논문 검색 서비스를 통해서 해당 분야가 어떻게 전개되고 발전되어 왔는지 전반적인 흐름을 파악하기 어려움
 - 체계적, 시각적 형태로 정리된 정보 부재 시, 논문의 인용 목록을 일일이 확인하고 인용 관계를 추적해야 함

3. 해결방안

A. 의미적 검색

- 텍스트의 의미, 맥락, 주제적 유사성에 기반한 검색 기법을 도입함으로써, 사용자는 입력한 키워드가 문서 내에 직접 언급되지 않더라도 해당 개념과 의미적으로 유사한 논문을 효과적으로 탐색

B. 키워드 혹은 문제 중심의 “논문 흐름 지도” 제공

- 특정 연구 문제나 핵심 키워드를 검색 시, 해당 주제와 관련된 주요 논문들을 시간 순 정렬하여 시각화 함

C. 브라우저 확장 프로그램을 통한 접근성 강화

- 확장 프로그램을 통해, 별도의 웹페이지 이동 없이 기존에 활용하던 검색 엔진이나 플랫폼 상에서도 즉시 논문 흐름 정보를 확인할 수 있음

4. 구현 목표

A. 문서 임베딩

- 논문 본문 중 Abstract, Background, Previous Work 등과 같이 연구의 주요 내용과 선행 연구를 알 수 있는 부분을 임베딩하여 벡터로 변환 후 저장

B. 의미적 질의 처리 및 유사도 기반 검색

- 사용자 질의를 임베딩한 후, 벡터 데이터베이스 내 문서 벡터와의 유사도를 측정해 연관성이 높은 논문 반환

C. 연도별 정렬 및 흐름 시각화

- 연관된 논문을 연도별로 정리하고 시각화하여, 특정 문제나 주제에 대한 연구 흐름을 직관적으로 제시

5. 기술스택

- 데이터베이스: PostgreSQL, S3
- 프론트엔드: HTML, CSS, Javascript
- 백엔드: Django
- 서비스배포: AWS, Chrome 개발자 대시보드

6. 기대효과

A. 연구 생산성 향상

- "Trust in Research" 보고서 (Elsevier와 Sense about Science가 협력)
 - 연구자들은 주당 연구 논문을 검색하는데 4시간 이상을 소비
 - 2011년부터 2019년 사이, 연구자들이 읽는 논문의 수는 10% 감소했지만 논문을 찾는 데 소비하는 시간은 11% 증가
- 연구 논문 수의 급증과 논문 탐색 시간 단축의 필요성
 - 최근 관련 조사가 없지만, 2019년 이후 연구 논문의 수가 급격히 증가함에 따라 논문 검색에 소요되는 시간이 더욱 늘어날 것으로 예상
 - 국내 연구에 따르면, Open Access 논문 검색 시스템(Discover)을 활용하면 연구자들이 논문 한 편을 검색하고 이용하는 데 걸리는 시간이 평균 30.13분에서 5.64분 단축되어 약 18.7%의 시간 절감 효과를 보임
 - 이와 유사하게 제안하는 서비스는 사용자가 입력하는 주제나 문제에 따라 주요 논문의 흐름 및 논문 리스트를 제공하여 논문 검색에 소요되는 시간을 최소 20% 이상 절감할 수 있을 것으로 기대

B. AI 인력 부족 및 인재 양성 정책에 따른 서비스 유입 증가 전망

- AI 인력 부족 현황
 - SPRi 실태조사 결과에 따르면, 기업이 사업 수행 시 겪는 9가지 주요 애로사항 중 'AI 인력 부족'이 5점 척도 기준으로 평균 4.25점으로 가장 높은 점수를 기록하였으며, 이는 AI 전문 인력 공급이 수요에 현저히 미치지 못하고 있음을 시사함
- 인재 양성을 위한 기업 및 정부의 적극적 노력
 - 최근 AI 인재 양성을 위해 일부 특성화 고등학교의 교육과정이 변화
 - 전국 4년제 대학의 약 40%가 AI 전공을 설치해 운영
 - 석·박사급 AI 융합 인재를 육성하기 위한 정부 사업 진행

- AI 인재 양성 정책의 강화로 인해 AI 입문자의 수가 증가함에 따라, 분야별 연구 흐름과 핵심 논문을 직관적으로 파악할 수 있는 Paper Map 서비스에 대한 수요도 함께 증가할 것으로 기대됨

C. 시장 확장 가능성

- 글로벌 학술 연구 데이터베이스 시장 규모는 2024년에 2억 8,626만 달러(USD)로 추산되었으며, 2031년까지 4억 3,614만 달러(USD)에 이를 것으로 예상됨
- 학술 데이터베이스 시장의 지속적인 성장에 발맞춰, 연구자와 초심자 모두에게 효율성과 접근성을 제공하며 시장에서의 입지를 강화할 수 있음
- Google Scholar와 같은 익숙한 학술 검색 플랫폼과의 연계를 통해 사용자는 직관적이고 신속한 정보 접근이 가능해지며, 본 서비스는 차별화된 경쟁력을 바탕으로 시장 점유율을 확대하고, 시장 성장률(CAGR 6.20%)에 맞춰 동반 성장 가능성이 높음

참고 문헌

1. Schopf, Tim, Karim Arabi, and Florian Matthes. "Exploring the landscape of natural language processing research." *arXiv preprint arXiv:2307.10652* (2023).
2. Science Business, Growth in data and questions on quality are increasing researcher workload, finds new study from Elsevier and Sense about Science, <https://sciencebusiness.net/network-updates/growth-data-and-questions-quality-are-increasing-researcher-workload-finds-new> (2019-08-22)
3. 표순희, et al. "국가오픈액세스플랫폼 논문검색시스템의 수요와 경제적 이용가치 측정." *한국문헌정보학회지*/56.3 (2022): 313-334.
4. 이경탁 기자, 교육계도 AI 바람... 전국 대학 'AI 학과'만 76개, 조선일보, 2024년 1월 13일, <https://biz.chosun.com/it-science/ict/2024/01/13/HYKBBQOGGVGMRCCKXEJN5B5MAY/>
5. 안유리, 韓기업, AI 비즈니스 가장 큰 장벽은 "인력 부족", 이투데이, 2024년 06월 13일, <https://www.etoday.co.kr/news/view/2369122>
6. Proficient Market Insights, ACADEMIC RESEARCH DATABASES MARKET REPORT OVERVIEW, <https://www.proficientmarketinsights.com/market-reports/academic-research-databases-market-1719> (2024-12-05)