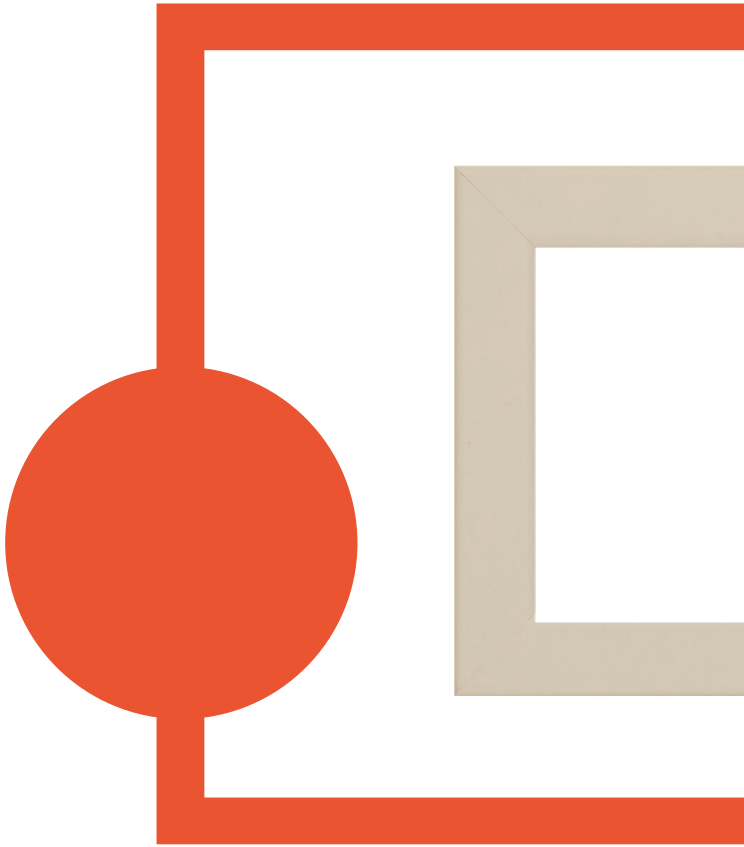
PaperFlow

2021 캡스톤 디자인 8조

20191632 윤상건 (팀장)
20160729 김태영
20181656 이민종



01 Problem

시장 조사를 통한 문제점 확인

설문 조사를 통한 문제점 확인

기존 문제점

02 Solution

PaperFlow

핵심 서비스 소개

기대 효과

03 Development

데이터셋

개발 기술

역할

협업 툴

멘토링

추가 개발 요소

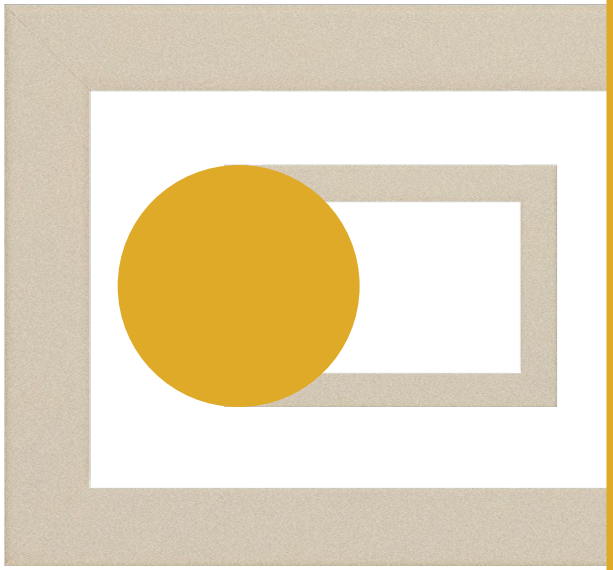
04 Appendix

논문흐름 그래프 생성 알고리즘

Use Case Diagram

배포 시스템 구조

01 Problem



시장 조사를 통한 문제점 확인

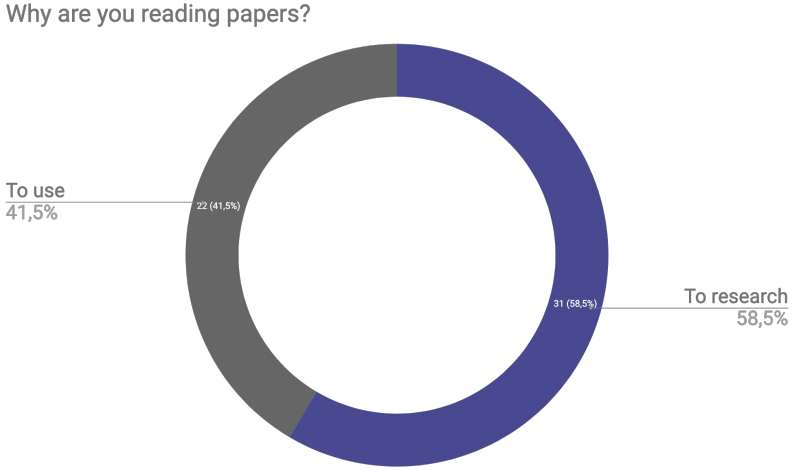
대표적인 논문 검색 서비스인    를 조사함.

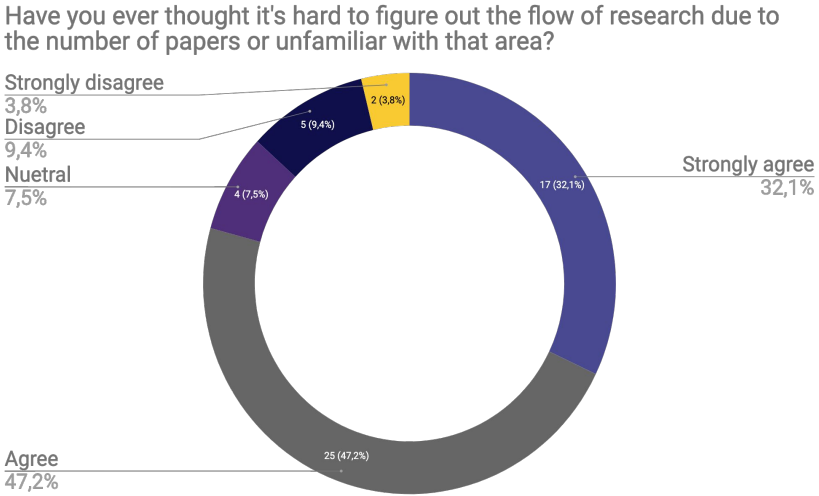
- 기존 논문 검색 플랫폼은 **특정 논문**에 대한 정보 전달에 집중하고 있다.

그렇기 때문에

- **여러 논문 간** 정보 전달 요소가 존재하지 않아 흐름 파악 어렵다.
- 특정 논문과 **연관된 핵심 논문**을 찾기 어렵다.

설문조사를 통한 문제점 확인





설문 조사를 Reddit, Facebook Group에서 직접 진행

20년 1월 19일 ~ 20년 1월 23일 설문 진행 / 총 53개의 응답

Reddit Machine Learning 채널: https://bit.ly/37Mcbbv / Facebook Groups: https://bit.ly/2UerC8d

기존 문제점

기존 분야의 **최근 연구 흐름**을 따라잡기 힘들다.

수많은 논문 중, 어떤 논문이 **핵심 논문**인지, 어떤 논문이 관심사와 가까운지 **파악**하기 힘들다.

※ 2019년, arxiv.org에서 CS분야 논문만 46,514건 (모든 분야에 대해서는 약 16만건)

02

Solution

PaperFlow

특정 논문이나 **연구 분야**에서의 **동향**에 대해

쉽고 빠르게 파악할 수 있는 자료를 제공하는 웹 서비스

핵심 서비스

기본 정보 열람

논문 흐름

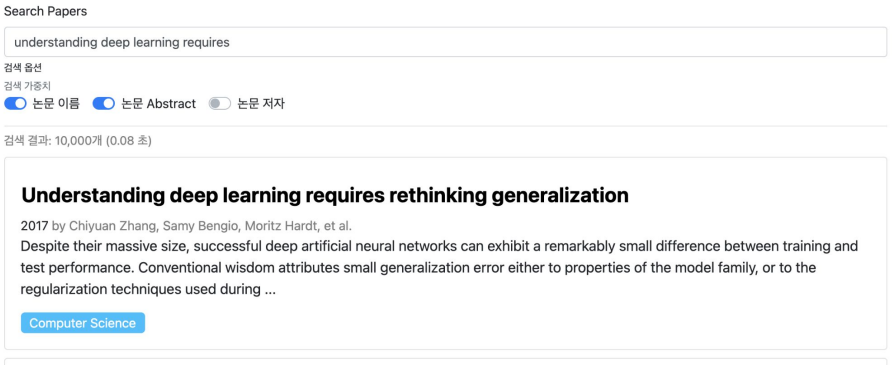
연구 키워드 시각화

연구 통계

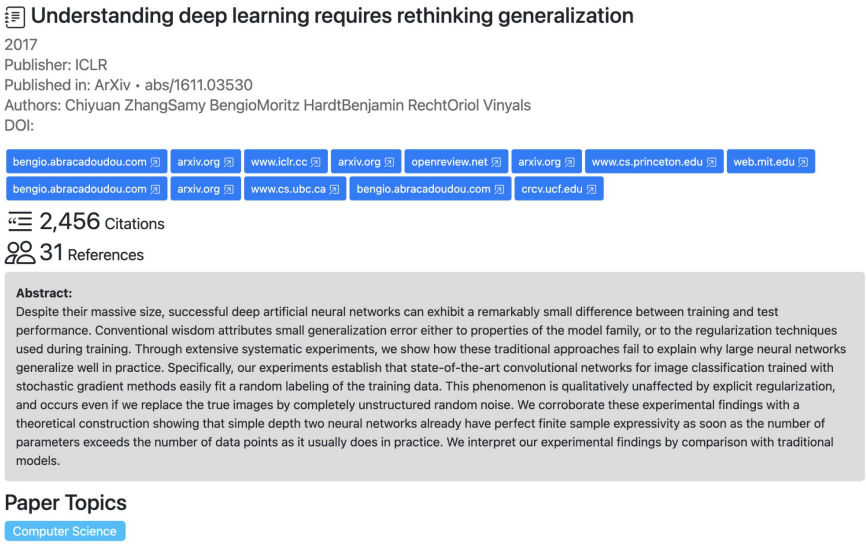
뉴스레터

핵심 서비스 1 - 기본 정보 열람

- **논문 검색 기능** 제공
- 검색한 논문에 대한 **기본 정보 제공**



〈 논문 검색 기능 〉

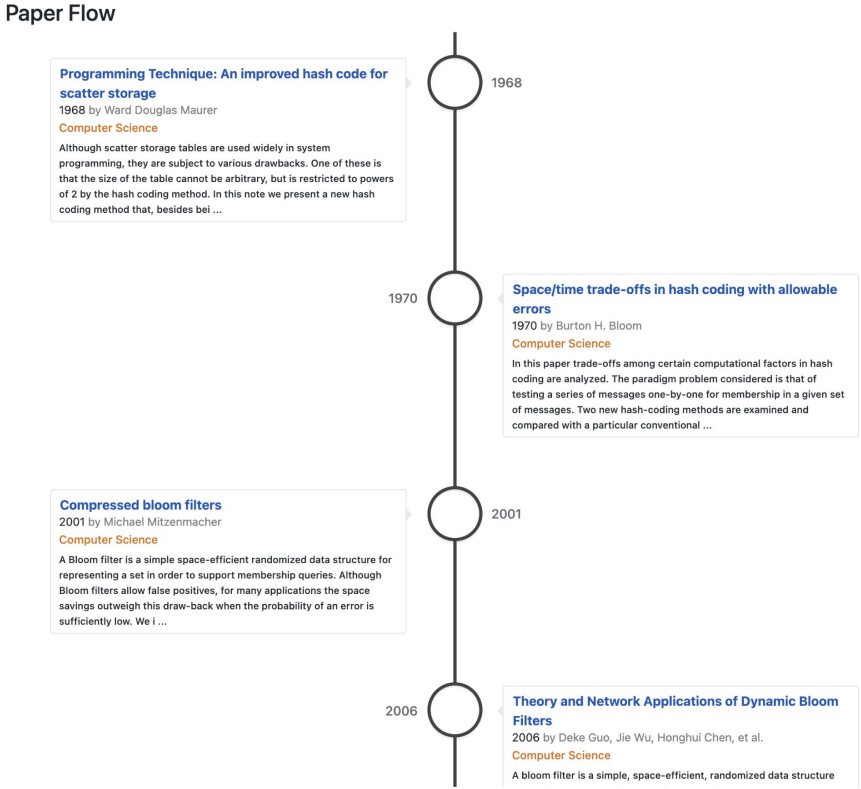


〈 논문 기본 정보 〉

핵심 서비스 2 - 논문 흐름

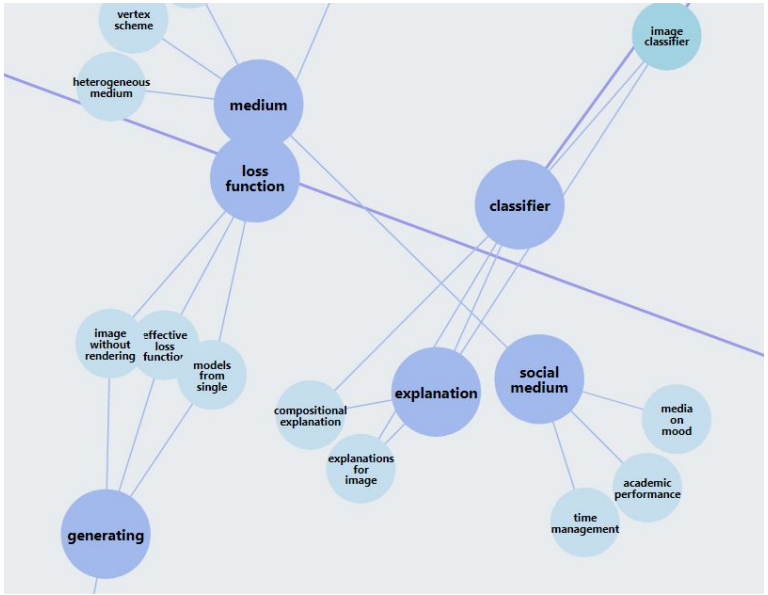
- 각 논문 별로 관련 **핵심 과거 논문** 및 **최신 논문**을 시간 순으로 제공
- 연구의 **진행 흐름** 파악을 도움
- 레퍼런스 기반 더 자세한 원리는 Appendix 참고
- A/B 테스트의 형태로 성능 개선

더 자세한 원리는 Appendix 참고



핵심 서비스 3 - 연구 키워드 시각화

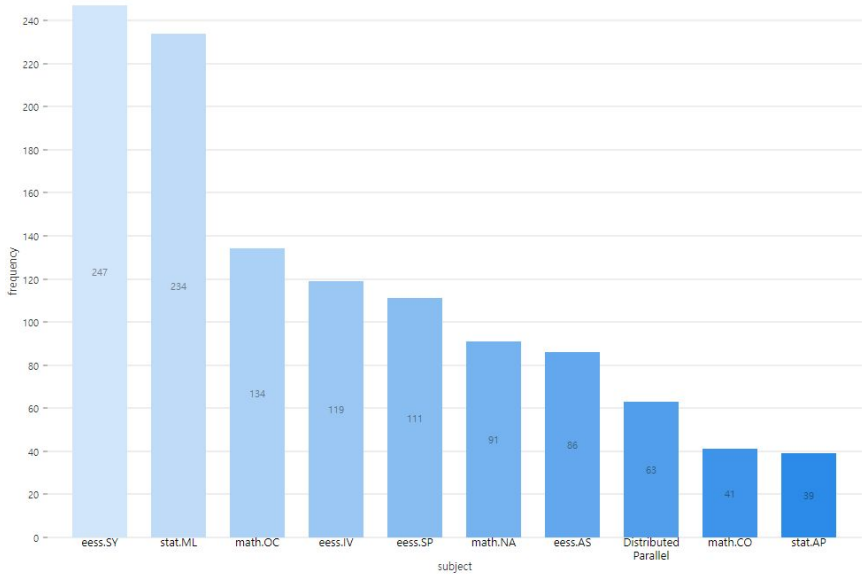
- 논문에서 등장한 **키워드**들의 **관련성**을 네트워크 형태로 **시각화** (월 단위)
- **생소한 분야**에서 **새로운 연구 분야**를 **탐색**하고 **발견**할 수 있도록 도움



< 키워드 그래프 예시 >

핵심 서비스 4 - 연구 통계

- 월 단위로 투고된 논문들의 정보에 대해 **통계** 및 **시각화**
- 최근 **연구 동향 파악**에 도움



핵심 서비스 5 - 뉴스레터

- **통계** 및 **이슈**에 대해 소개하는 **뉴스레터**를 이메일을 통해 구독 및 발송

기대 효과

비전문 분야라도 **관심 논문**에 대해 과거부터 현재까지의 **연구 흐름**을 쉽게 파악

시각적 도구로 **특정 분야, 특정 기간**에 진행되었던 연구의 동향을 쉽게 파악

비전문 분야라도 **키워드**를 이용하여 최신 관련 분야를 쉽게 발견 및 탐색

03

Development

데이터셋

1. Semantic Scholar
   - 14,116,789 개
   - 전체 CS 분야 논문

수많은 논문 정보 →

- 기본 논문 정보
- 논문 흐름

2. ArXiv
   - 214,623 개
   - 2017 ~ 2021.04 CS 분야 논문

월별 데이터 →

- 연구 키워드 시각화
- 연구 통계
- 뉴스레터

개발 기술

Front End

웹개발

UI/시각화

배포

더 자세한 구조는 Appendix 참고

Docker 및 AWS를 이용한 배포

Back End

API 서버

데이터 분석

DB

김태영

백엔드
DB 구축
데이터 분석

윤상건

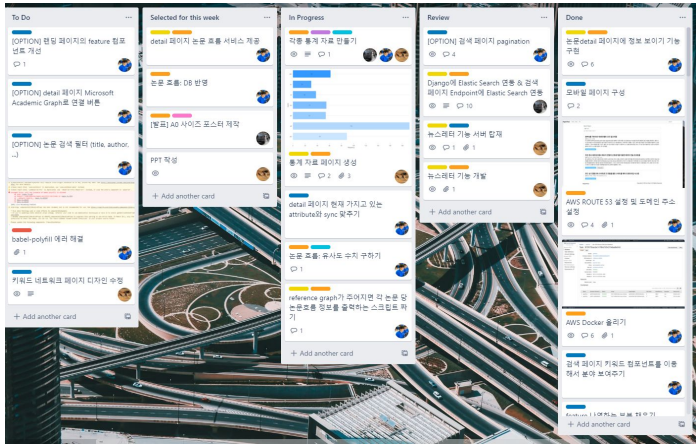
풀스택
AWS 배포
데이터 분석

이민종

풀스택
데이터 시각화
데이터 분석

협업 툴

Trello



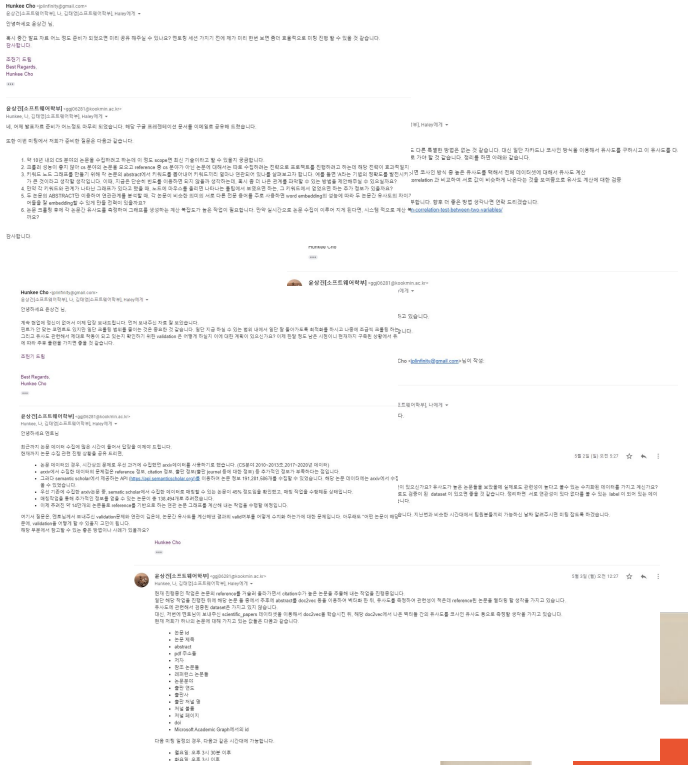
GitHub

Google Drive

DISCORD

멘토링

- 조헌기 멘토님
- 40회 이상의 메일을 주고받으며, 여러 도움을 받을 수 있었음
  - 프로젝트 스코프에 대한 의견 등을 질문 드리고 도움을 받음.
  - 논문에 친숙하신 멘토님께서 소비자 입장에서의 경험 공유를 통해 서비스에 대한 피드백 반영
  - 참고할만한 타 서비스 추천 (Microsoft Academic)
  - 기술적인 조언
    - Ex) 도메인에 맞는 워드 임베딩 학습을 위한 데이터셋 추천



추가 개발 요소

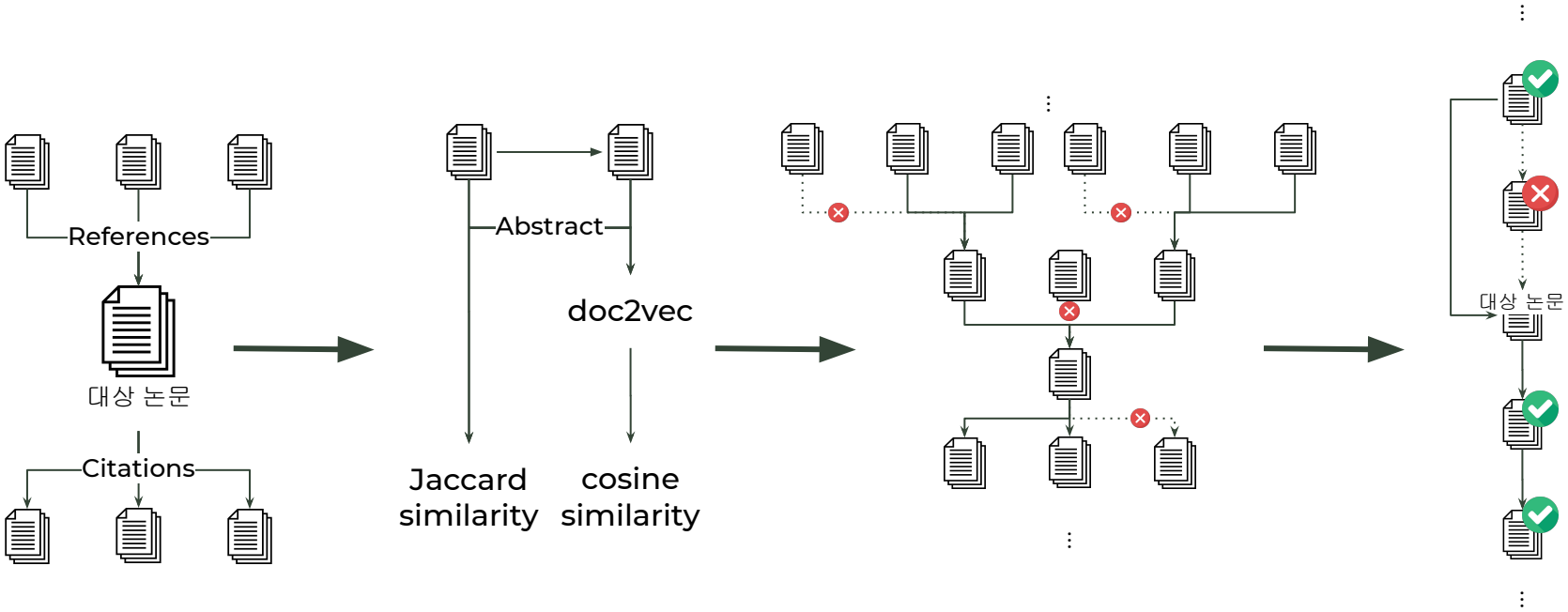
- 분야 확장
  - Computer Science 외 분야까지 포함
- 추가 기능
  - 로그인 기능
  - 로그인을 통한 개인 맞춤 기능
  - 논문 요약 서비스 제공
- 통계 고도화
- 논문 흐름 서비스 고도화
  - 사용자 경험에 따라 서비스 성능 개선

THANKS

20160729 김태영
20191632 윤상건
20181656 이민종

04 Appendix

논문흐름 그래프 생성 알고리즘



참조를 통한
Directed Acyclic Graph 생성

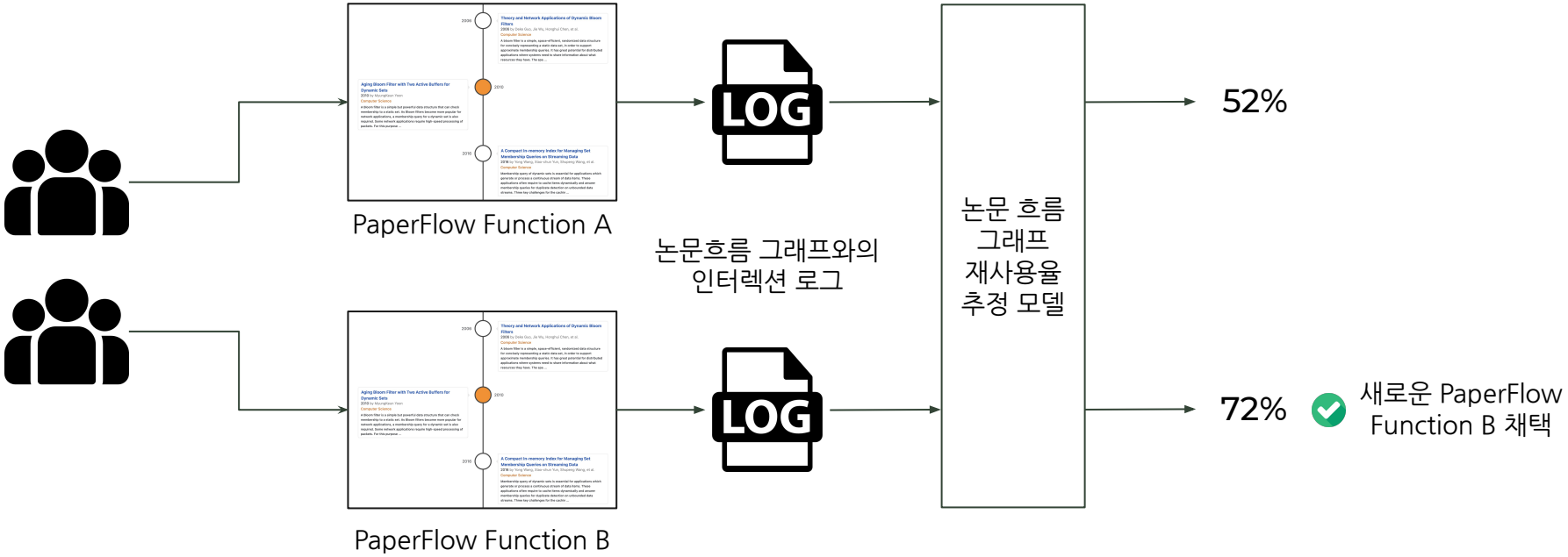
간선 간 유사도 계산

논문간 유사도 기반
Pruning

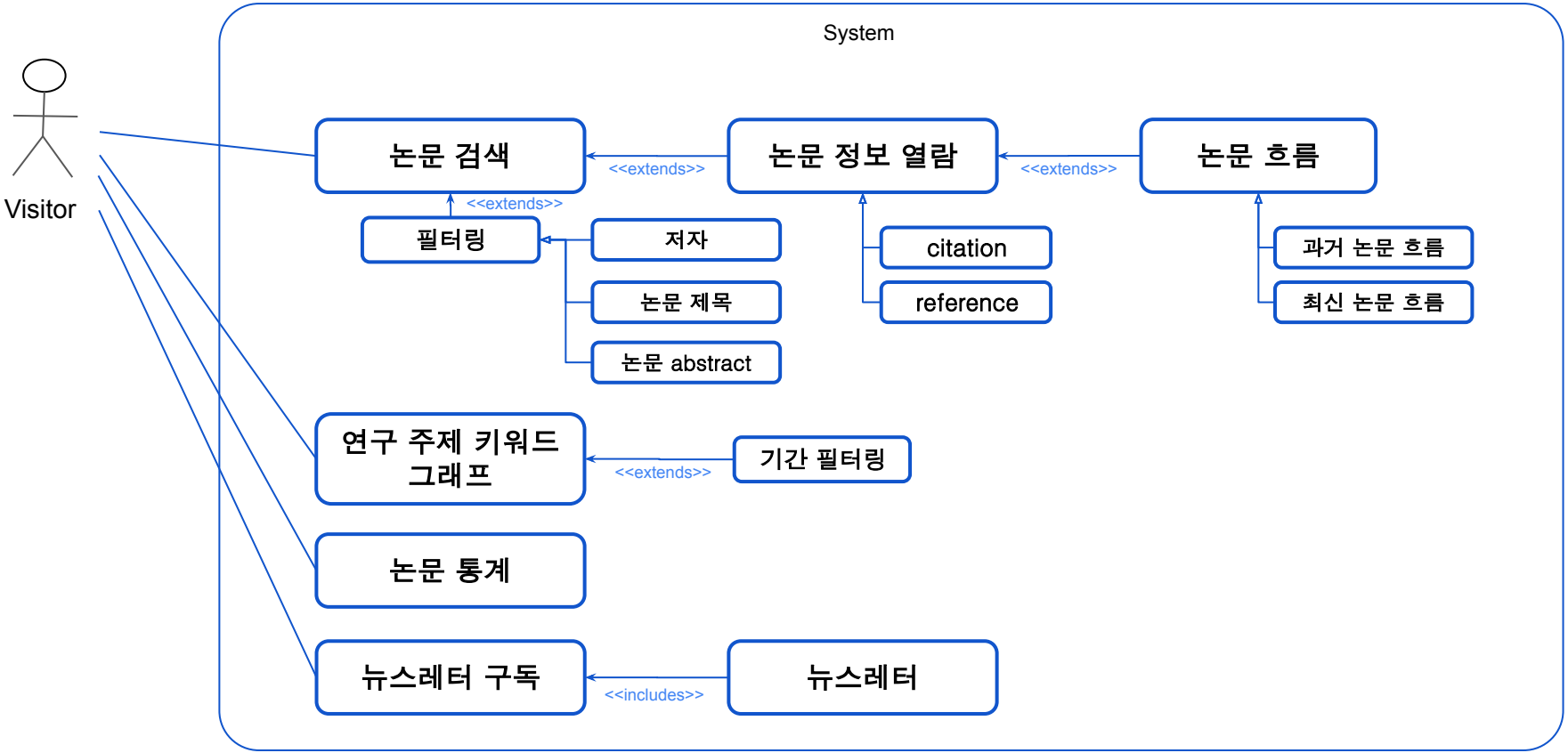
위상정렬 및 Citation 수
기반 상위 논문 필터링

논문흐름 그래프 평가 및 개선 전략

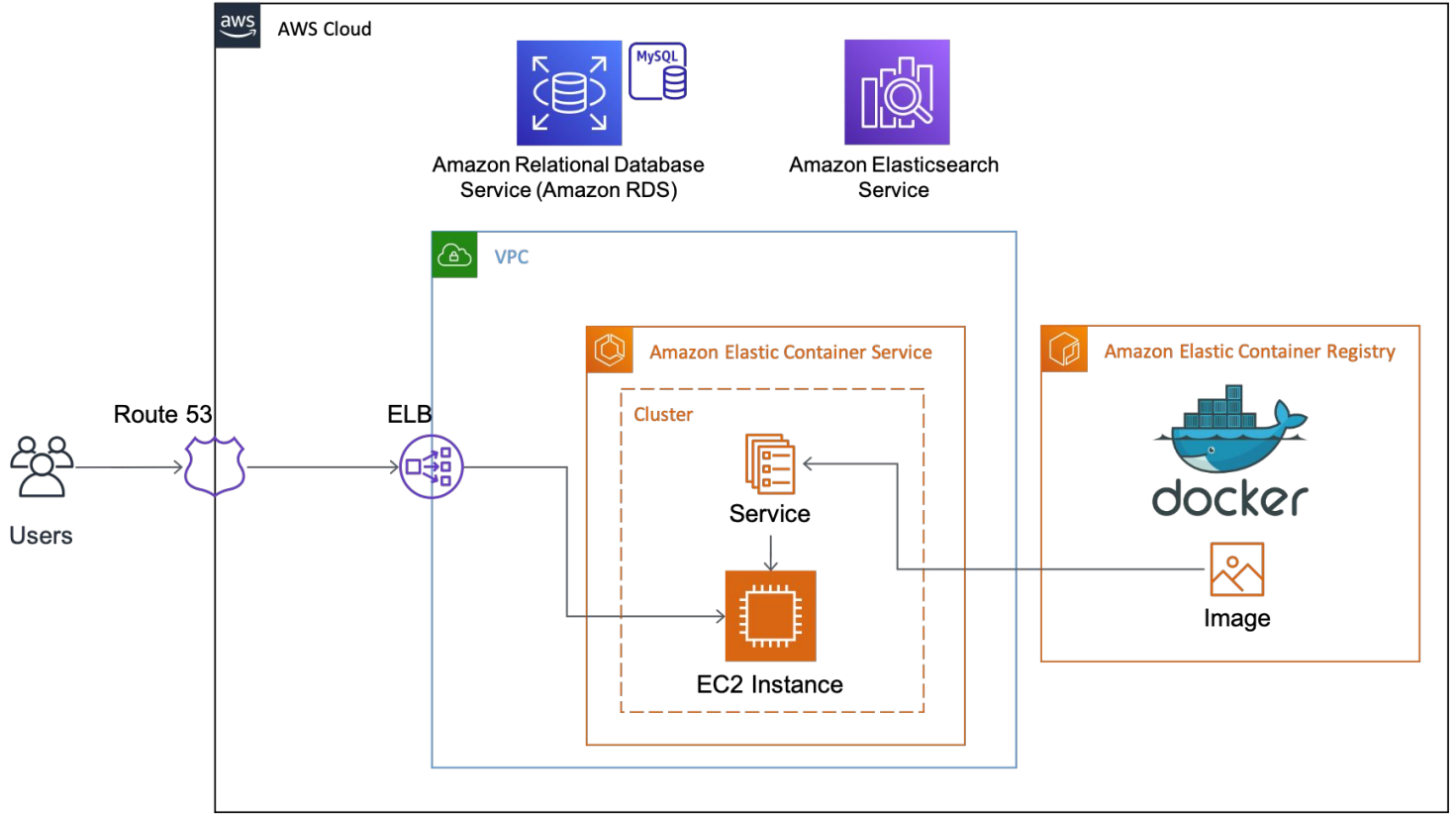
- 결국 논문 흐름 그래프의 성능은 논문간 유사도 함수에 달려있음
- A/B 테스트 형태를 통해 **유사도 함수**를 개선하면 논문 흐름 그래프를 유저 입장에서 잘 보여주는 **최적의 논문흐름 함수**로 수렴할 것



Use Case Diagram



배포 시스템 구조


AWS Cloud
MySQL
Amazon Relational Database Service (Amazon RDS)
Amazon Elasticsearch Service
VPC
Amazon Elastic Container Service
Cluster
Service
EC2 Instance
Amazon Elastic Container Registry
docker
Image
Route 53
ELB
Users