

Eungseo Kim

Researcher / Data Engineer
ngseo@s2w.inc
+82) 01046543090
<https://bento.me/ngseo>

EDUCATION

2020.03. ~ Ajou University / Cyber Security

WORK EXPERIENCES

- Data Engineer / Researcher at Knowledge Engineering, S2W inc., Korea (Nov. 2021 ~ Present) / Data engineering, Software Engineering, Graph mining

SKILLS AND TECHNIQUES

- Data pipeline, Crawling, Ontology, Graph data science, Statistical analysis
- Elasticsearch 7, RabbitMQ, MongoDB, Lucene, MySQL, MinIO
- Threat Intelligence. OSINT, Secure coding, Kubernetes, Docker
- Spring boot, FastAPI
- Angular
- Python, Java, Typescript
- Jira, Gitlab, Notion, Confluence

*숙련도 순으로 정렬됨 (Proficient -> Beginner).

RESEARCH INTERESTS

- Information retrieval
- Semantic search
- Graph mining
- Cyber threat intelligence

PROJECTS

36억 규모의 과기정통부-경찰청 사이버 안보 침해대응 플랫폼

관련 링크: <https://www.boannews.com/media/view.asp?idx=108567>

기술 스택: Java 17, Python 3, Spring boot, MongoDB, Lucene, Typescript, Angular, MinIO, MySQL

- 2차년도 기술 기획부터 R&D 전과정에 참여하였으며 하반기에는 유일한 실무자로 기존 업무에 더해 프론트엔드 및 백엔드 인수인계 후 신규 피쳐 개발 및 VoC 대응에 대한 유지보수 관리함.
- 데이터 크롤링, ETL 파이프라인 및 Lucene 기반 Rule-based 탐지엔진 단독 개발
- 프론트엔드 신규 피쳐로 탐지 분석 결과에 대한 유향 그래프 시각화 컴포넌트 및 Admin UI 개발
- 백엔드 신규 피쳐로 RBAC, 그래프 기반 데이터 모델 반영, 사내 데이터레이크 연동 및 보안 문제 패치 4건 등의 성과

핵심 제품의 Graph DB 색인 파이프라인 성능 11배 개선

기술 스택: Python3

- 2만줄이 넘는 코드베이스에서 bottleneck을 Cprofile로 진단
- 사내 Graph DB 특성상 Transaction을 보장하지 않고, 그래프 노드 및 엣지들 간의 데이터 의존성 문제로 인하여 많은 부분에서 순차적으로 데이터를 처리하고 있었음
- 그래프의 특성을 고려하여 일부 병렬화가 가능한 포인트들을 찾아 전체 성능 10배 이상 개선

IPv4 전대역 전포트 스캐너

기술 스택: Python 3, RabbitMQ, Elasticsearch 7

- 1Gbps 망을 대부분 사용하는 스캐너 및 스캔 결과에 대한 내결함성 있는 데이터 파이프라인 개발
- 초당 100MB 이상을 저장하는 Elasticsearch 서버 세팅 및 정책 유지보수
- 스캐너가 네트워크 트래픽을 대부분 점유하는 상태에서의 RabbitMQ 운영

사내 ETL 및 Pull-based CDC 파이프라인 프레임워크 신규 개발 및 유지보수

기술 스택: Python 3

- 사내 크롤러 일부 23종 및 파이프라인 일부 8종에 적용되어 실환경에 도입된 상태

핵심 제품 데이터 Export 파이프라인 성능 50배 개선

기술 스택: Java 11, RabbitMQ, Elasticsearch, Amazon S3, Redis

- 메모리를 많이 차지하는 애플리케이션임을 고려하여, jmap으로 힙덤프하여 진단.
- XSSFWorkbook은 파일 전체를 자바 객체로 변환하여 처리하기 때문에 대용량 파일에 대해서 잦은 append 작업이 있을 때 큰 비효율을 보였음
- SXSSFWorkbook으로 구현체를 바꾸어 메모리 사용량을 대폭 줄여 성능 개선

TAXII 프로토콜 구현체 개발

관련 링크: <https://docs.oasis-open.org/cti/taxii/v2.1/os/taxii-v2.1-os.html>

기술 스택: Java 17, Spring boot

- 관련 링크에 해당하는 프로토콜 스펙을 맞춘 서버 애플리케이션 구현
- 오픈소스로 공개되어 있는 애플리케이션을 사용하기에는 SSPL-1.0 라이선스적 한계가 있어 직접 구현

핵심 제품의 Elasticsearch 기반 자체 GraphDB 개발 및 유지보수

기술 스택: Python 3, Elasticsearch 7

- 유지보수, 모니터링 및 ad-hoc한 데이터 요청들에 대한 대응

FaaS (Function as a Service) 형태의 데이터 파이프라인 플랫폼 개발

기술 스택: Python 3, FastAPI, Flutter Web

- 업로드한 데이터를 처리하는데 있어서 Python Edge Function만 배포하여 처리하는 파이프라인 플랫폼 개발
- 악성 Python 스크립트를 대비하여 코드 정적분석을 AST 기반의 화이트리스트 필터링 기법을 사용하여 보안성 강화