OSS 유사도 및 라이선스 분석 플랫폼에 관한 연구

김기환[1], 윤성철[2], 김수현[3], 이임영[4]
[1]순천향대학교 컴퓨터소프트웨어공학과 학부생
[2]순천향대학교 소프트웨어융합학과 석사과정
[3,4]순천향대학교 컴퓨터소프트웨어공학과 교수
20184004@sch.ac.kr ,ysc1346@sch.ac.kr, kimsh@sch.ac.kr, imylee@sch.ac.kr

A Study on Platform for OSS Similarity and License Analysis

Ki-Hwan Kim[1], SeongCheol Yoon[2], Su-hyun Kim[3], Im-Yeong Lee[4]
[1,3,4]Dept. of Software Computer Software Engineering, Soonchunhyang University
[2]Dept. of Software Convergence, Soonchunhyang University

요 약

소프트웨어를 개발하는 과정에서 소스 코드를 직접 작성하면 높은 비용과 시간이 필요하다. 이를 해결하기 위해 OSS를 활용해 개발 비용 절감 및 소요 시간 단축 등 다양한 이점을 가지게 된다. 그러나 수많은 기업은 무분별한 OSS 사용으로 인해 개발 중인 소프트웨어에 적용되는 OSS의 라이선스를 정확히 파악하지 못한다. 그로 인해 라이선스 위반 및 충돌로 인한 저작권 문제로 법적 분쟁과 상용화된 소프트웨어 제품을 전부 리콜해야 하는 등의 피해가 발생한다. 하지만 국내에는 이러한 문제를 체계적으로 분석하고 예방하기 위한 점검 도구가 부족하다. 본 논문은 앞서 언급된 문제를 해결하고자, 높은 접근성을 바탕으로 OSS의 정보를 효과적으로 분석하는 플랫폼을 구현하였다. 사용자가 소스 코드의 분석을 요청하면, 플랫폼에 등록된 OSS 프로젝트 중 가장 높은 유사도를 보이는 프로젝트의 메타데이터, 유사도 분석 결과, 라이선스 정보를 제공한다. 이를 통해 사용자들은 자신들이 사용 중인 소스 코드에 적용된 OSS의 세부 구성 요소를 편리하게 분석하고 조회할 수 있다.

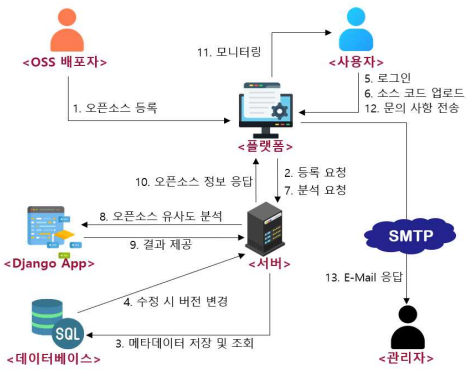
1. 서 론

소프트웨어 개발 시 모든 코드를 직접 구현하면 상당한 개발 비용과 시간이 소모된다. 이를 해결하고자 OSS(Open Source Software)를 사용하기 시작했다. OSS는 무료이자 유연한 커스터마이징이 가능한 소프트웨어이다. 따라서 OSS를 효율적으로 활용하면 위 문제를 해결할 수 있다[1]. 그러나 많은 기업에서는 OSS 라이선스를 정확히 인식하지 못한다. 만약 OSS로 개발한 소프트웨어를 라이선스를 고려하지 않고 배포하는 경우, 제품을 리콜해야 하는 문제가 발생한다[2].

본 논문은 위의 문제를 해결하기 위하여, 소스 코드 내 OSS 사용유무, 라이선스 종류를 분석하는 플랫폼을 구현한다. 프로젝트를 플랫폼에 등록하고, 등록된 프로젝트와 사용자가 요청한 소스 코드 유사도를 분석해 다양한 메타데이터와 유사도 등을 제공한다. 이를 통해 사용자들은 소스 코드의 OSS 구성 요소를 효율적으로 분석할 수 있다. 따라서 본 논문은 높은 접근성으로 OSS 정보 및 라이선스를 분석해 안전한 OSS 활용을 지원한다.

2. OSS 유사도 및 라이선스 분석 플랫폼

본 장에서는 OSS 유사도 및 라이선스 분석 플랫폼의 제안방식과 및 구현 과정에 대해 설명한다. 플랫폼은 HTML, CSS, JavaScript로 프론트엔드를 구현한다. 백엔드는 Python으로 구현하고 uWSGI(unified Web Server Gateway Interface)를 통해 웹 애플리케이션과 웹 서버간의 통신이 가능하다.



[그림 1] 전체 시나리오

로그인과 회원가입은 Django의 유저 인증 시스템을 적용했다. 플랫폼에 등록되는 OSS의 메타데이터는 SQLite로 구축하여 DB(DataBase)에 저장한다. 소스 코드의 OSS의 유사도 분석은 TF-IDF(Term Frequency-Inverse Document Frequency)와 Cosine Similarity 알고리즘을 적용했다. 요청 및 문의 사항의 메일 통신은 SMTP(Simple Mail Transfer Protocol)를 활용했다. [그림 1]은 이를 적용한 시나리오를 나타내며, 플랫폼 세부 동작 과정들은 아래와 같다.

1. OSS 등록 과정

- ***Step 1***: 이름, 소스 코드, 라이선스, README를 입력한다. 그리고 개발한 폼에 작성 후, 서버에 OSS 등록을 요청한다.
- ***Step 2:*** 라이선스 파일과 README를 마크다운 파일로 생성하여 소스 코드와 함께 동적 생성한 디렉토리에 추가하여 서버에 저장한다.

2. OSS 데이터 저장 및 조회 과정

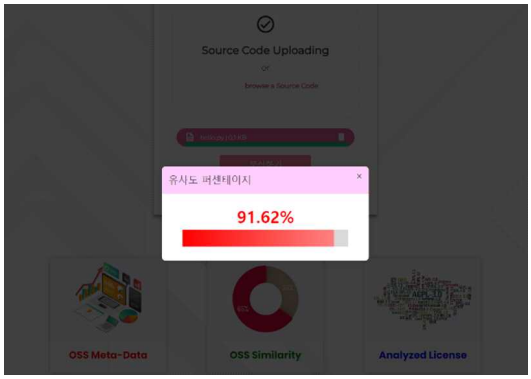
- ***Step 1:*** 등록 폼 요소들을 OSS 메타데이터 속성으로 구성한다. 그리고 배포자, 배포 날짜, 버전을 1.0으로 구성하여 속성에 동적 추가하고 메타데이터를 저장된다.
- ***Step 2:*** 동일한 OSS를 재등록할 경우 버전의 부 번호를 자동으로 +0.1씩 더하여 등록이 가능하도록 구성했다. 등록된 OSS의 리스트를 기반으로 검색창에 OSS의 이름 또는 배포자를 검색하여 일치하는 내용을 조회한다.
- ***Step 3:*** 이름을 클릭하여 모든 메타데이터가 표시된 각각의 URL로 이동하여 세부 정보를 조회할 수 있다.

3. OSS 분석 과정

- ***Step 1:*** 소스 파일 또는 압축 파일을 폼에 업로드한다. 버튼 클릭 후 3초 동안 로딩되며 [그림 2]와 같이 유사도 분석을 수행한다.
- ***Step 2:*** 메타데이터, 유사도, 라이선스 정보 및 충돌 여부를 하단의 카드를 클릭 후 모달(Modal) 창에서 모니터링한다. 라이선스 준수 사항은 모달 창 안의 버튼을 통해 URL 이동 후, 카드를 클릭한 뒤 라이트박스(LightBox)에서 확인한다.

4. 사용자 문의 사항 송수신 과정

- ***Step 1:*** 제목, 내용의 폼에서 플랫폼 요청 및 문의 사항을 입력한다.
- ***Step 2:*** 관리자가 자신의 네이버(Naver) 메일에서 조회할 수 있다.



[그림 2] OSS 분석 결과 모니터링

4. 결론

본 제안방식은 OSS를 사용하기 전에 사전 분석을 통해 주요 정보를 파악함으로써, 이전에 발생했던 문제들을 해결한다. 기존분석 도구는 솔루션 이용 방법도 복잡하고 로열티가 부과되어 모든 사용자가 활용하기에 어려움이 있었다. 본 제안방식에서는 높은 접근성을 기반으로 개발에 사용하는 소스 코드의 OSS가 적용하는 기본 정보, 라이선스 및 버전 별 수정 사항 등을 확인할 수 있다. 이를 통해 개인 및 기업은 SW 개발 과정에서 저작권 문제로부터 안전하게 OSS를 활용할 수 있다.

향후 본 제안방식을 활용하여 분석한 OSS를 정밀하게 스캐닝하고 취약점을 분석하는 소프트웨어 공급망 공격에 대한 보안성을 강화하기 위한 연구 또한 필요할 것으로 사료된다.

Acknowledgments

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2023년도 SW저작권 생태계 조성 기술개발 사업(과제명 : 클라우드 서비스 활용 구축 형태별 대규모 소프트웨어 라이선스 검증 기술개발, 과제번호 : RS-2023-00224818, 기여율: 50%)과 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2022H1D8A3038040)

참 고 문 헌

[1] 정보통신산업진흥원, “2020년 오픈소스 SW 시장 동향 조사 보고서”, 2020.

[2] 이동건, 서영석, “오픈소스 라이선스 양립성 위반 식별 기법 연구”, Vol. 7, No. 12, pp. 451-460, 2018.