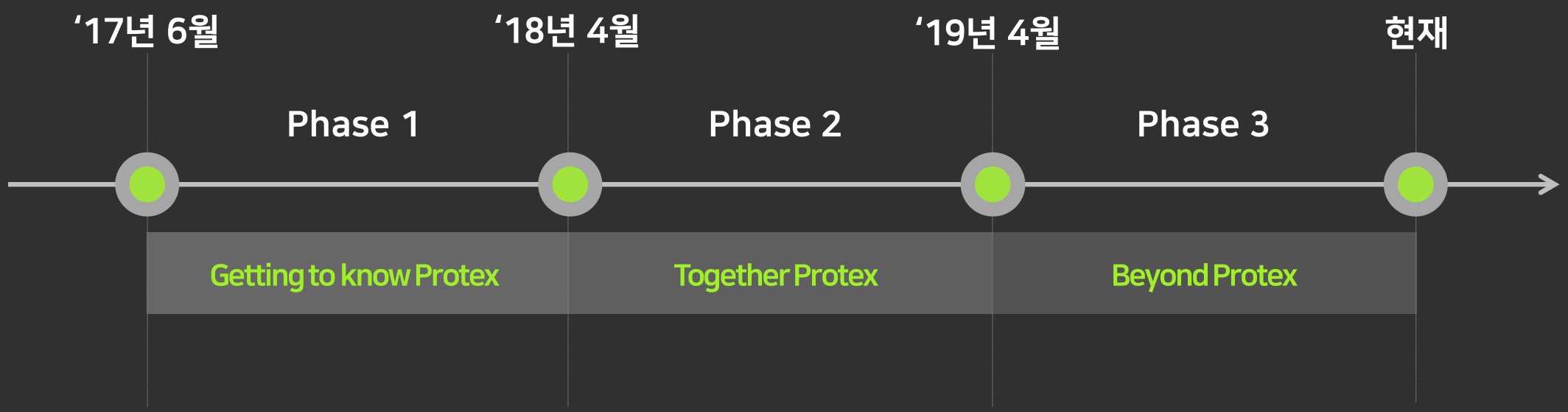
OpenChain Korea WG 공유용

오픈소스 소프트웨어 라이선스 검증記

엔씨소프트 CTO직속 TS팀 한지호

2019. 09. 04

지난 2년 간의 여정을 3단계로 구분해 보았습니다.



공정 사용(Fair Use)

- 기본적으로 저작권으로 보호되는 저작물을 저작권자의 허가를 구하지 않고 제한적으로 이용할 수 있도록 허용하는 미국 저작권법상의 개념, 예) 학문 연구, 평론

- 저작권법 제35조 3 (저작물의 공정한 이용)
  저작물의 통상적인 이용 방법과 충돌하지 아니하고 저작자의 정당한 이익을 부당하게 해치지 아니하는 경우에는 저작물을 이용할 수 있다.

- 저작권법 제28조 (공표된 저작물의 인용)
  공표된 저작물은 보도·비평·교육·연구 등을 위하여는 정당한 범위 안에서 공정한 관행에 합치되게 이를 인용할 수 있다.

Phase 1
Getting to know Protex

소프트웨어에서의 저작권 침해 사례

HANCOM 한글과컴퓨터

Artifex

--

한컴은 지난 2013년부터 아티펙스의 오픈소스 기반 PDF 인터프리터 고스트스크립트(Ghostscript)를 한컴 오피스에 내장함

고스트스크립트는 Dual License로써 한컴은 해당 코드 사용에 대해 전체 소스코드를 공개하거나(GPL 2.0), 상용코드 사용에 대한 저작권료를 지불해야 했으나 아무런 조치도 취하지 않음

2017년 5월, 고스트스크립트의 저작권을 가지고 있는 아티펙스사가 한컴을 GPL 라이선스 위반으로 고소함

▶ 한컴은 고스트스크립트 코드 사용에 대해 205만 달러(약 26억원)를 아티펙스사에 지불함

오픈소스의 세상으로

2017년 6월

어느 날...

관련 교육 수강

2017년 8월

진행중인 과정

[2019 오픈소스 SW 라이선스 전문교육] 일반과정 2차

- 교육기간: 2019-09-27 ~ 2019-09-27
- 신청기간: 2019-07-22 ~ 2019-09-26

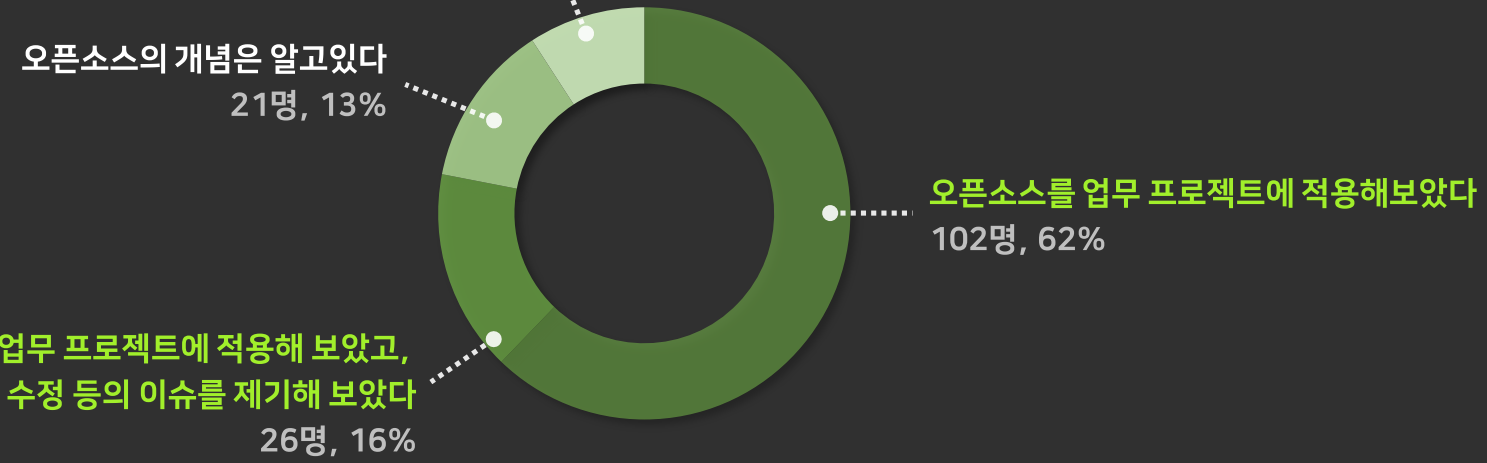
[2019 오픈소스 SW 라이선스 전문교육] 고급과정 2차

- 교육기간: 2019-08-29 ~ 2019-08-30
- 신청기간: 2019-07-29 ~ 2019-08-29

사내 개발자분들의 오픈소스 사용 현황과 니즈를 파악하고자 설문조사를 실시하였습니다.

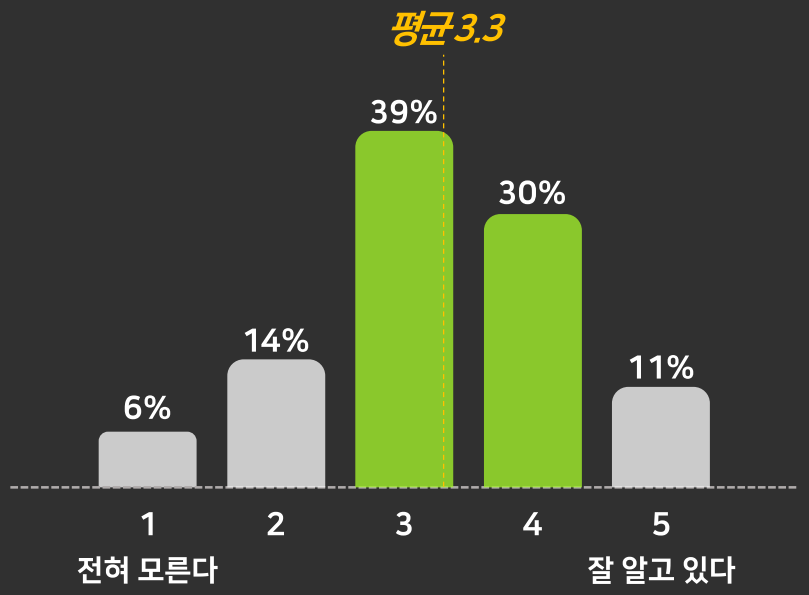
(조사 기간: 2018.3.22 ~2018.3.23 / 응답자: 164명)

Q. 오픈소스의 전반적인 내용에 대해 얼마나 알고 계십니까?

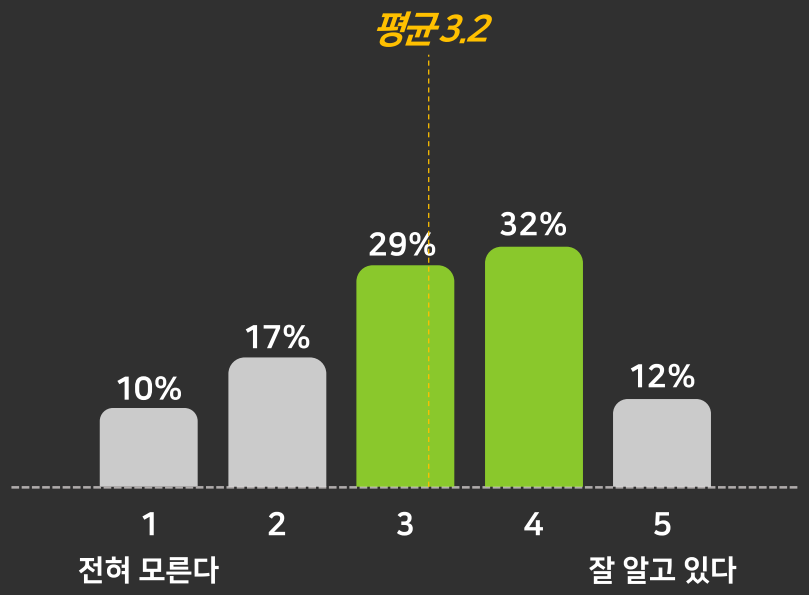


사내 개발자 대상 설문

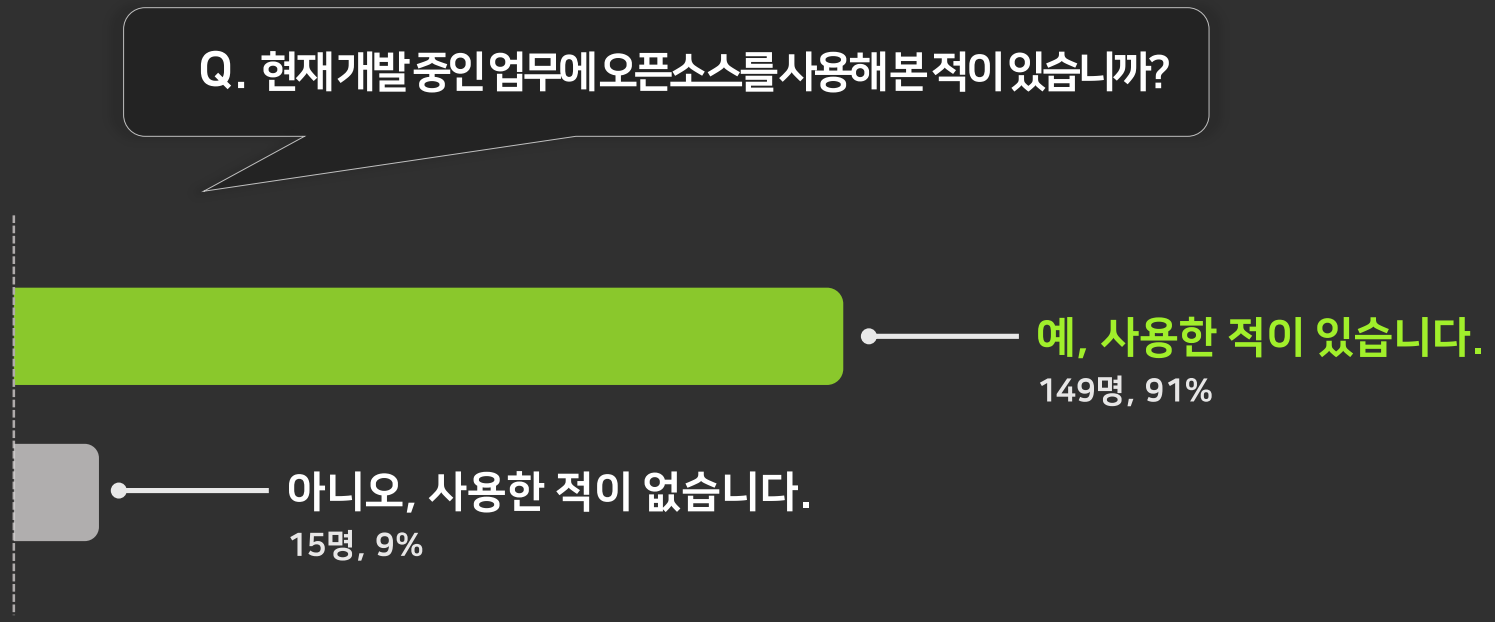
Q. 오픈소스의 주요 라이선스 별 의무사항에 대해 알고 계십니까?



Q. 오픈소스의 의무사항을 준수하지 않을 경우 어떤 리스크가 있는지 알고 계십니까?

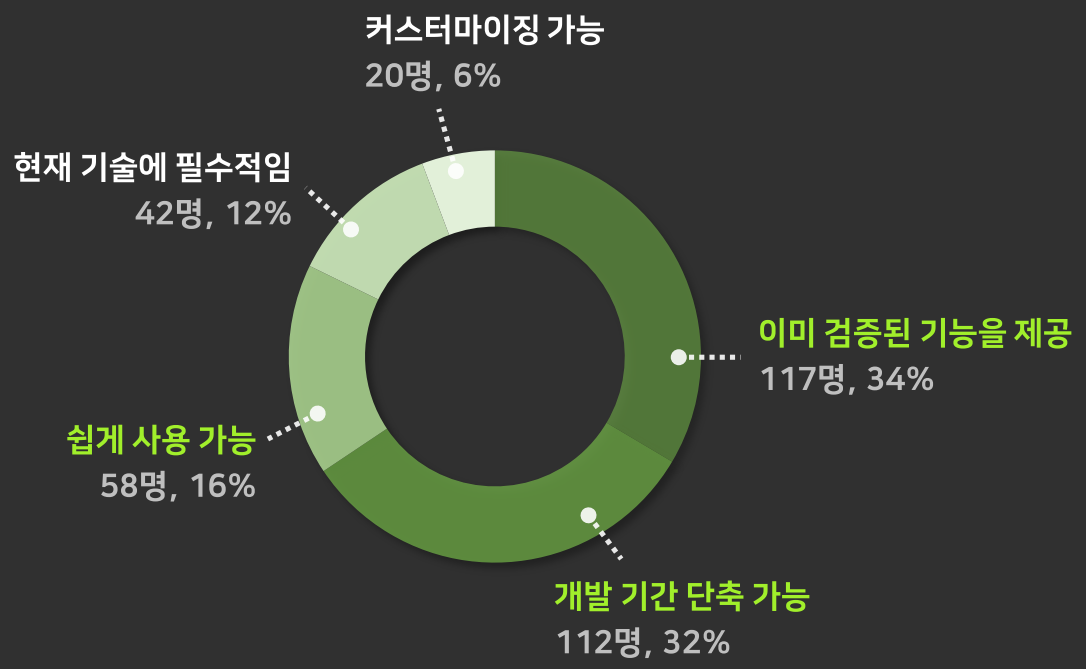


사내 개발자 대상 설문

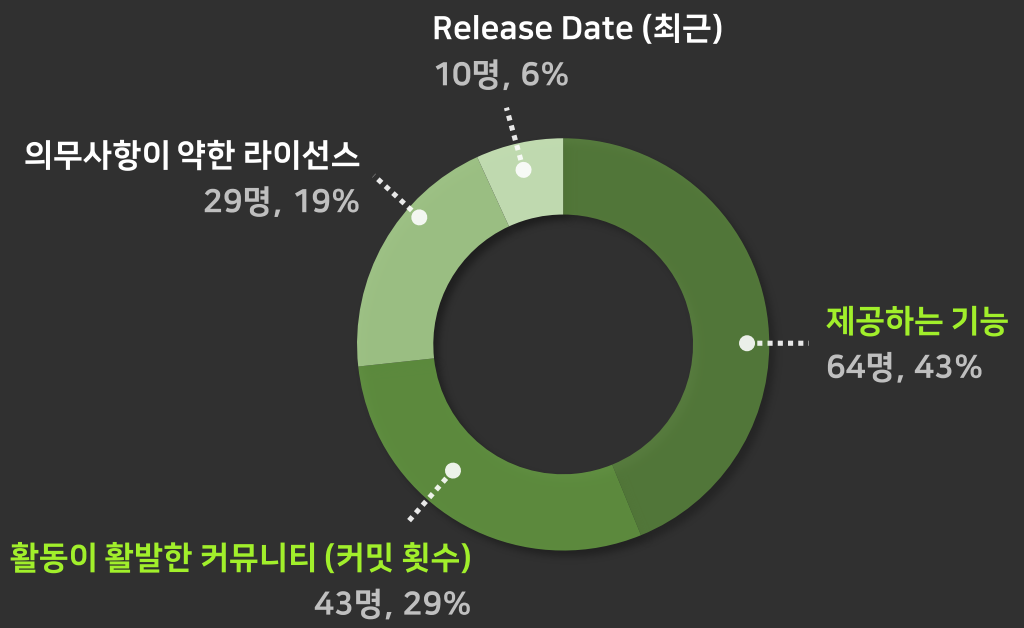

Q. 현재 개발 중인 업무에 오픈소스를 사용해본 적이 있습니까?
예, 사용한 적이 있습니다.
149명, 91%
아니오, 사용한 적이 없습니다.
15명, 9%

사내 개발자 대상 설문

Q. 오픈소스를 사용해 본 경험이 있다면 주로 어떤 이유로 사용하십니까?

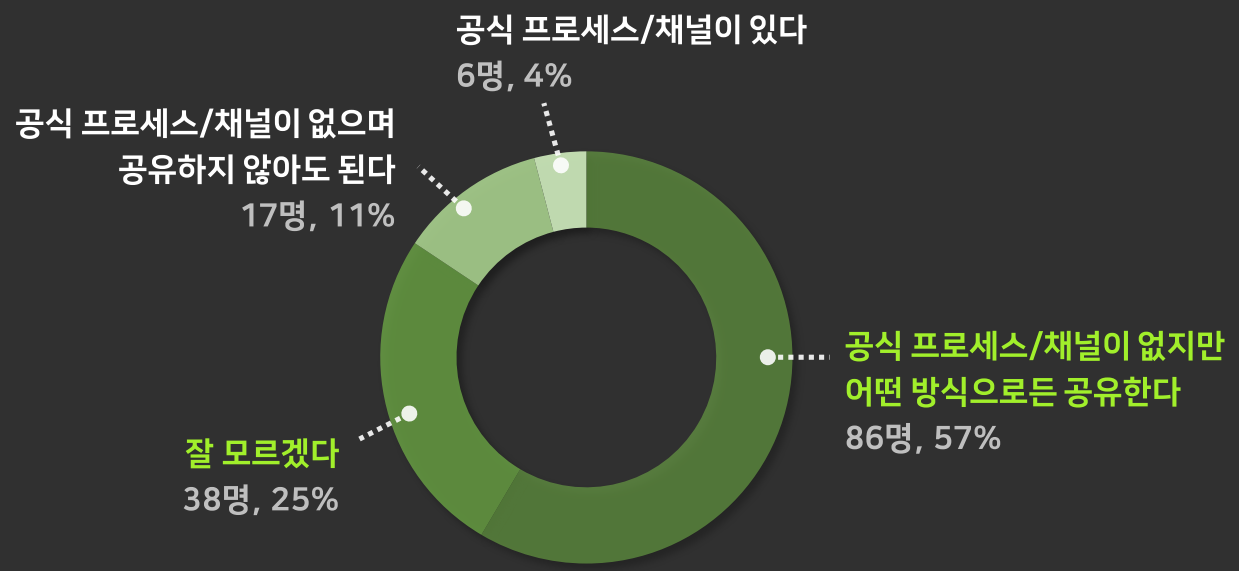


Q. 오픈소스를 선택하는 최우선 기준은 무엇입니까?

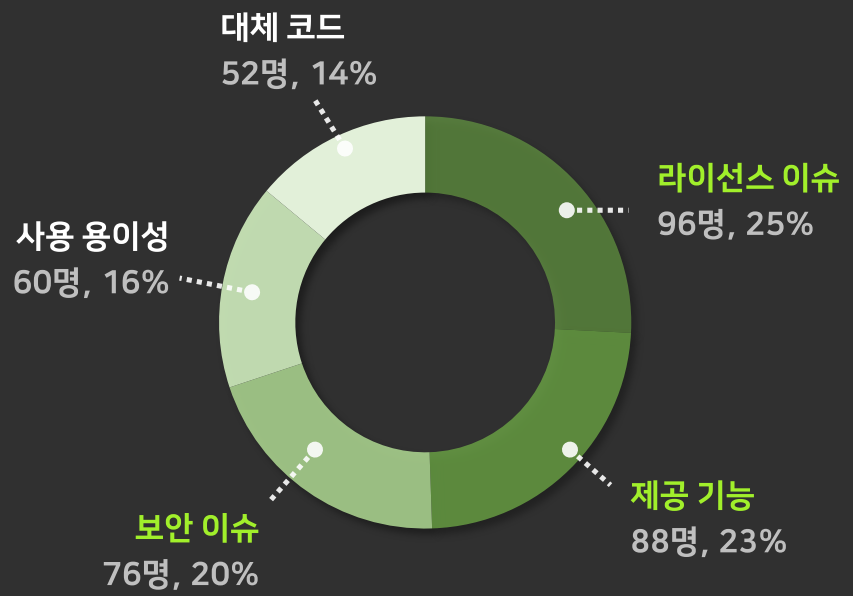


사내 개발자 대상 설문

Q. 부서 내에 오픈소스 사용과 관련한 내용을 공유하는 프로세스/채널이 있습니까?

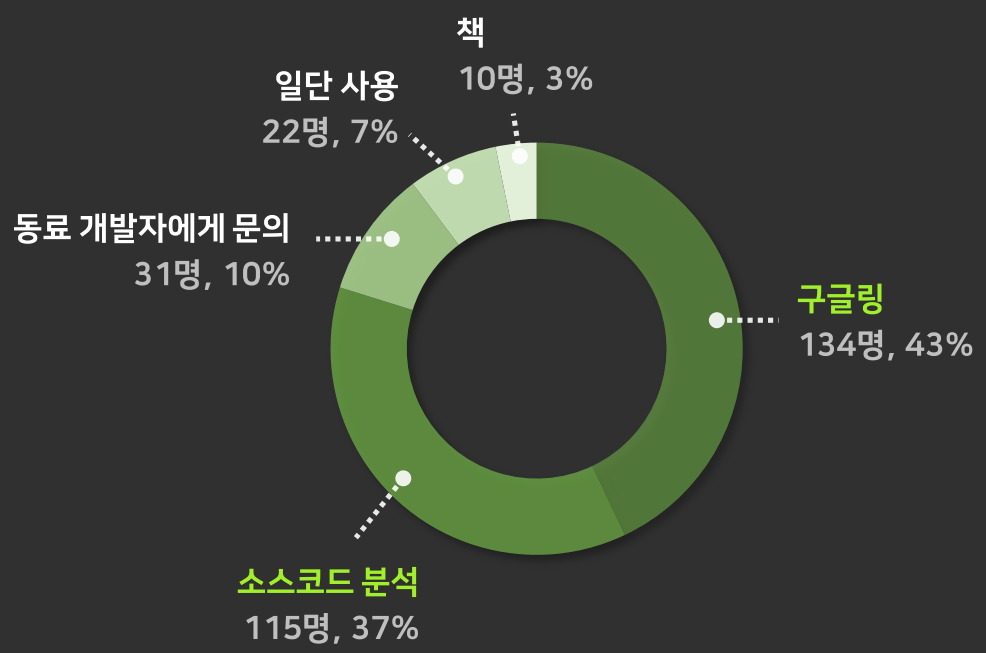


Q. 오픈소스 사용 시 궁금한 점이 있다면 주로 어떤 것들입니까?

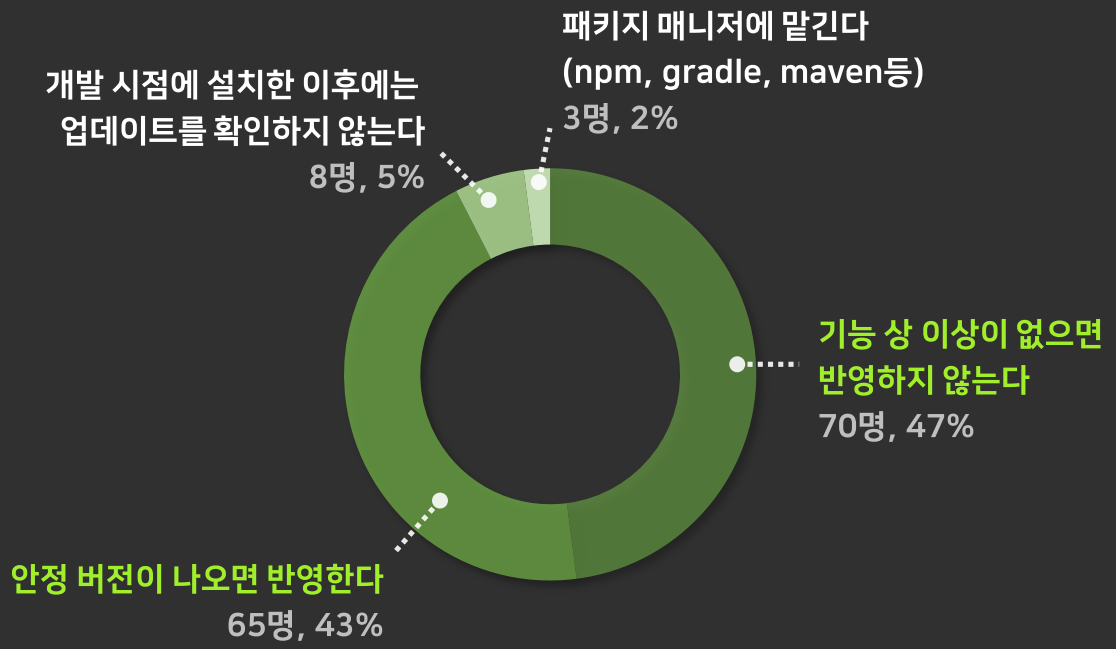


사내 개발자 대상 설문

Q. 오픈소스의 Github 페이지나 프로젝트 페이지에 있는 내용만으로 궁금한 점이 해소되지 않을 경우 어떻게 해결하십니까?



Q. 사용했던 오픈소스의 업데이트를 지속적으로 반영하고 계십니까?



사내 개발자 대상 설문

Q. 오픈소스를 사용할 때 어려운 점은 무엇입니까?

분류	의견
라이선스	라이선스의 범위를 확인하는 것이 가장 번거롭습니다.
	오픈소스 적용 시 법적 의무사항 준수가 신경 쓰입니다.
보안	망, 개발 환경 등의 보안 정책으로 인해 오픈소스에 접근이 쉽지 않습니다.
	보안취약점, 안정성 여부를 파악하기 힘듭니다.
사용 관련	찾고 있던 적절한 기능인지 판단하기 어렵습니다.
	버그 픽스된 새로운 버전이 릴리즈 되면 지속적으로 업데이트 반영을 해야 하고 그에 따른 호환성 이슈를 해결하는 것이 부담입니다.
	오픈소스 내에서 버그 발생 시 해당 부분을 고쳤을 때 사이드 이펙트를 판단하기 어려운 경우가 있습니다.

사내 개발자 대상 설문

Q. 오픈소스의 사용과 관련하여 회사에 바라는 점은 무엇입니까?

분류	의견
라이선스	법률 및 라이선스 관련 교육이 필요하다고 생각합니다.
	해당 오픈소스 링크를 제출하면 사용 가능한지(라이선스, 보안 등) 판단해주는 지원이 있으면 좋겠습니다.
보안	신개발망은 대부분 사이트에 대한 접근이 막혀있어 일일이 신청하거나 웹하드를 써야 하는데 이를 완화해 주세요.
사용 관련	오픈소스 커뮤니티 기여를 권장했으면 합니다.
	오픈소스 사용과 관련하여 개발자 간 의견을 공유할 수 있는 채널이 있으면 좋겠습니다.

To do list

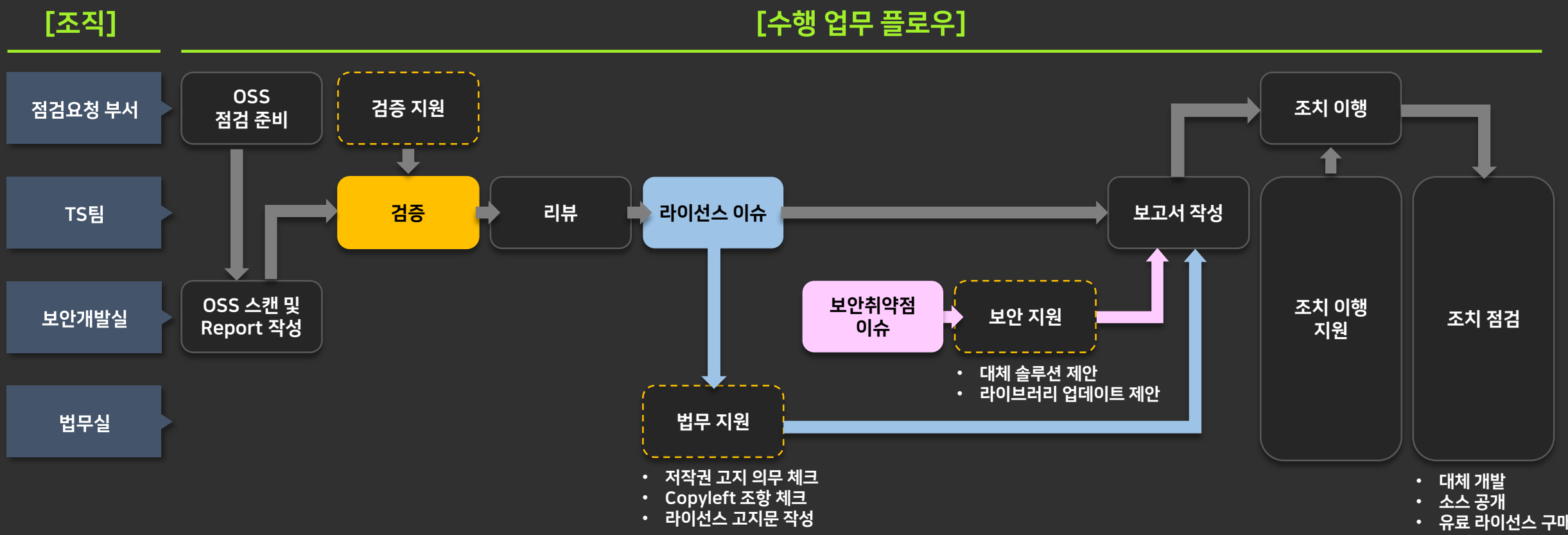
- 라이선스 문의를 할 수 있는 채널 마련
- 개발자들이 개발 주제로 커뮤니케이션 할 수 있는 공간 마련
- 오픈소스 사용과 관련한 교육 시행
- 오픈소스 소프트웨어 기여 정책 검토

Phase 2
Together Protex

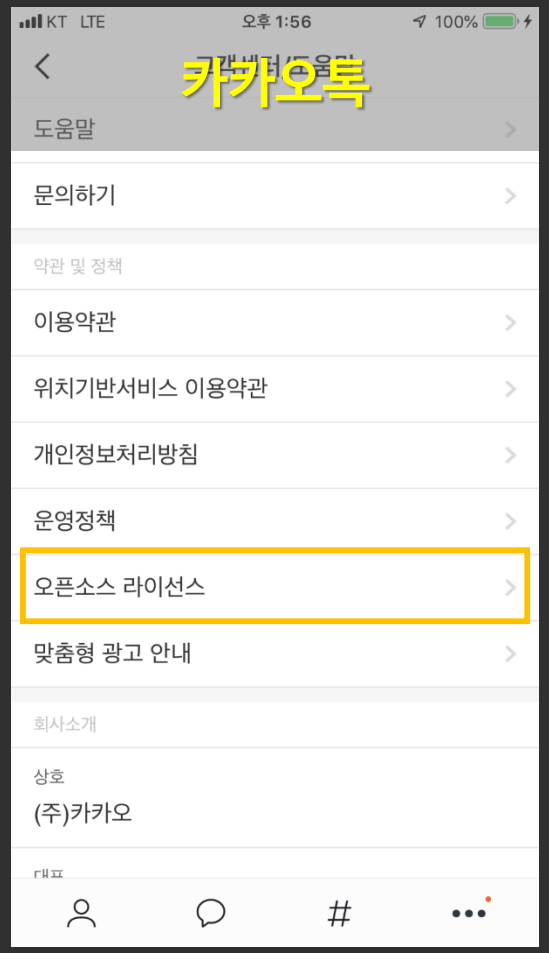
무엇을 검증할 것인가?

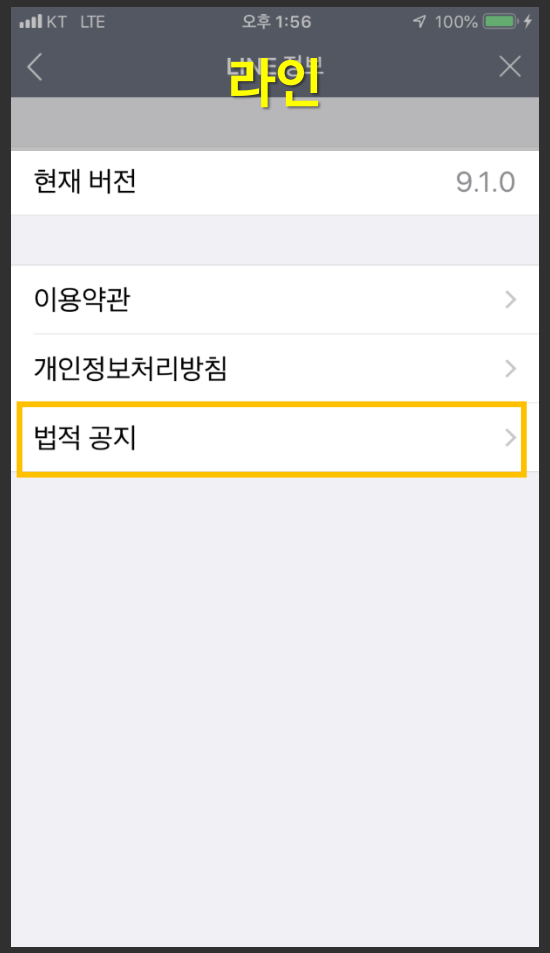
사용자에게 배포되는 프로그램에 어떤 오픈소스 소프트웨어를 사용하고 있는가?	검색하면 찾을 수 있는 수 많은 오픈소스들...
식별된 오픈소스 소프트웨어의 라이선스는 무엇인가?	Apache 2.0, MIT, BSD, GPL, LGPL...
해당 라이선스의 조건과 의무사항은 무엇이며, 해당 의무사항을 이행하고 있는가?	라이선스 고지, 배포 시 라이선스 사본 첨부... (라이선스 종류에 따라 상이함)

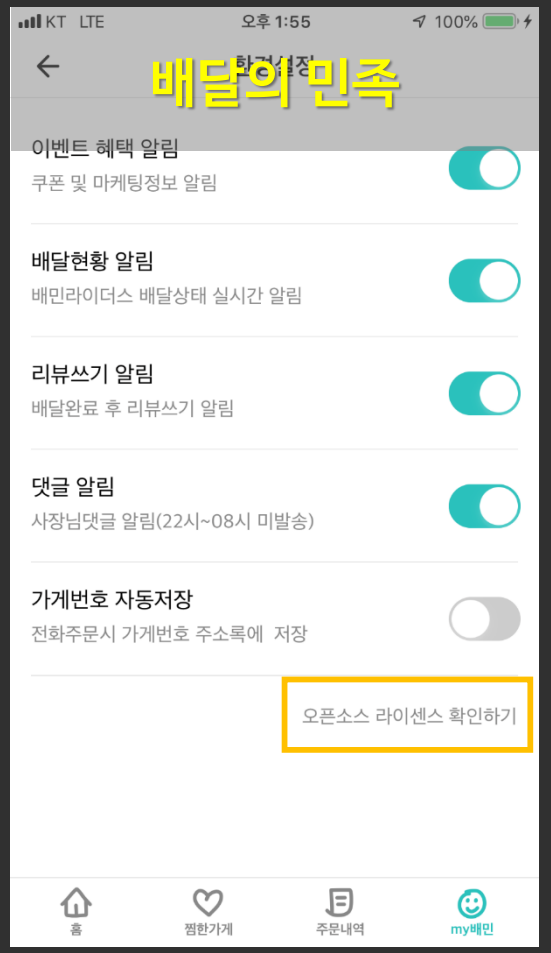
검증 프로세스

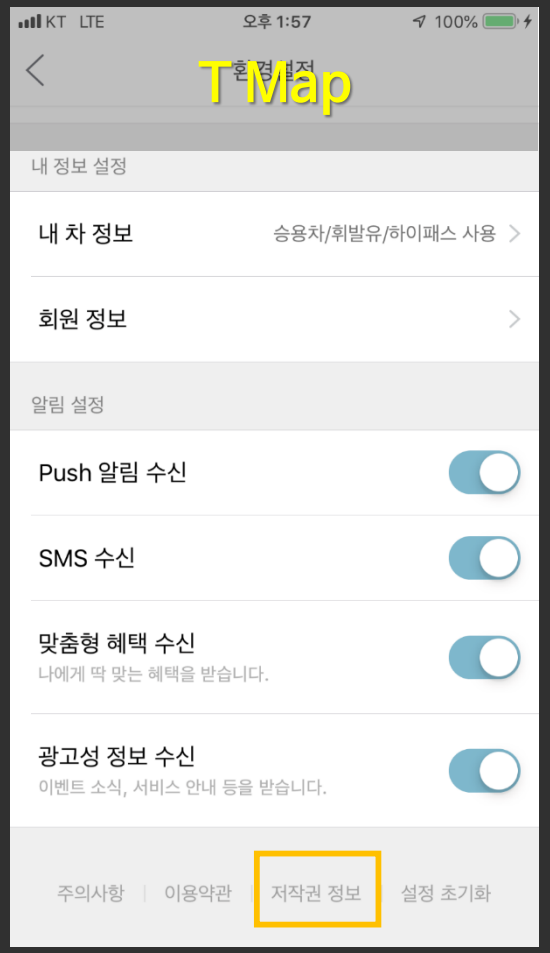

[조직]
[수행 업무 플로우]
점검요청 부서
TS팀
보안개발실
법무실
OSS 점검 준비
검증 지원
검증
리뷰
라이선스 이슈
OSS 스캔 및 Report 작성
보안취약점 이슈
보안 지원
• 대체 솔루션 제안
• 라이브러리 업데이트 제안
법무 지원
• 저작권 고지 의무 체크
• Copyleft 조항 체크
• 라이선스 고지문 작성
보고서 작성
조치 이행
조치 이행 지원
조치 점검
• 대체 개발
• 소스 공개
• 유료 라이선스 구매

고지사례 (일반 앱)


카카오톡
KT LTE 오후 1:56 100%
도움말
문의하기
약관 및 정책
이용약관
위치기반서비스 이용약관
개인정보처리방침
운영정책
오픈소스 라이선스
맞춤형 광고 안내
회사소개
상호
(주)카카오

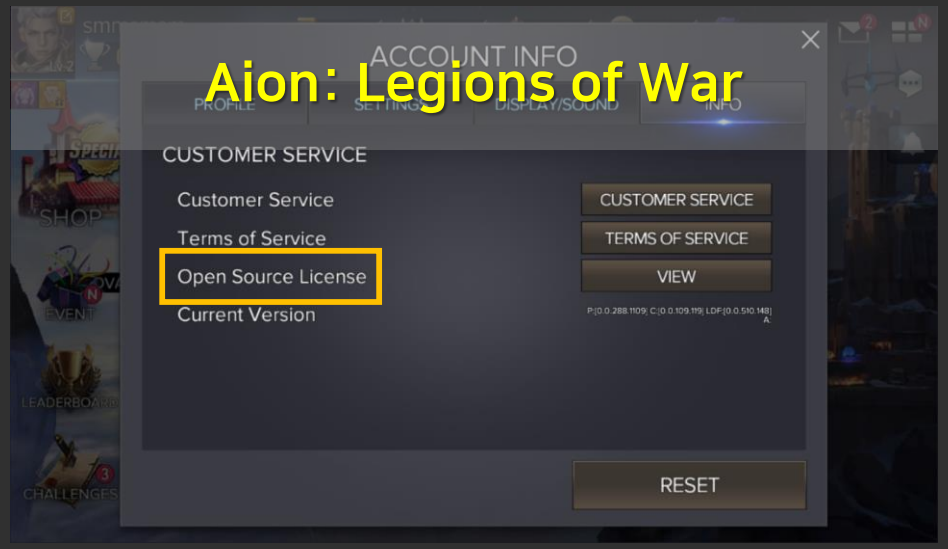

라인
KT LTE 오후 1:56 100%
현재 버전 9.1.0
이용약관
개인정보처리방침
법적 공지


배달의 민족
KT LTE 오후 1:55 100%
이벤트 혜택 알림
쿠폰 및 마케팅정보 알림
배달현황 알림
배민라이더스 배달상태 실시간 알림
리뷰쓰기 알림
배달완료 후 리뷰쓰기 알림
댓글 알림
사장님댓글 알림(22시~08시 미발송)
가게번호 자동저장
전화주문시 가게번호 주소록에 저장
오픈소스 라이센스 확인하기
홈
찜한가게
주문내역
my배민

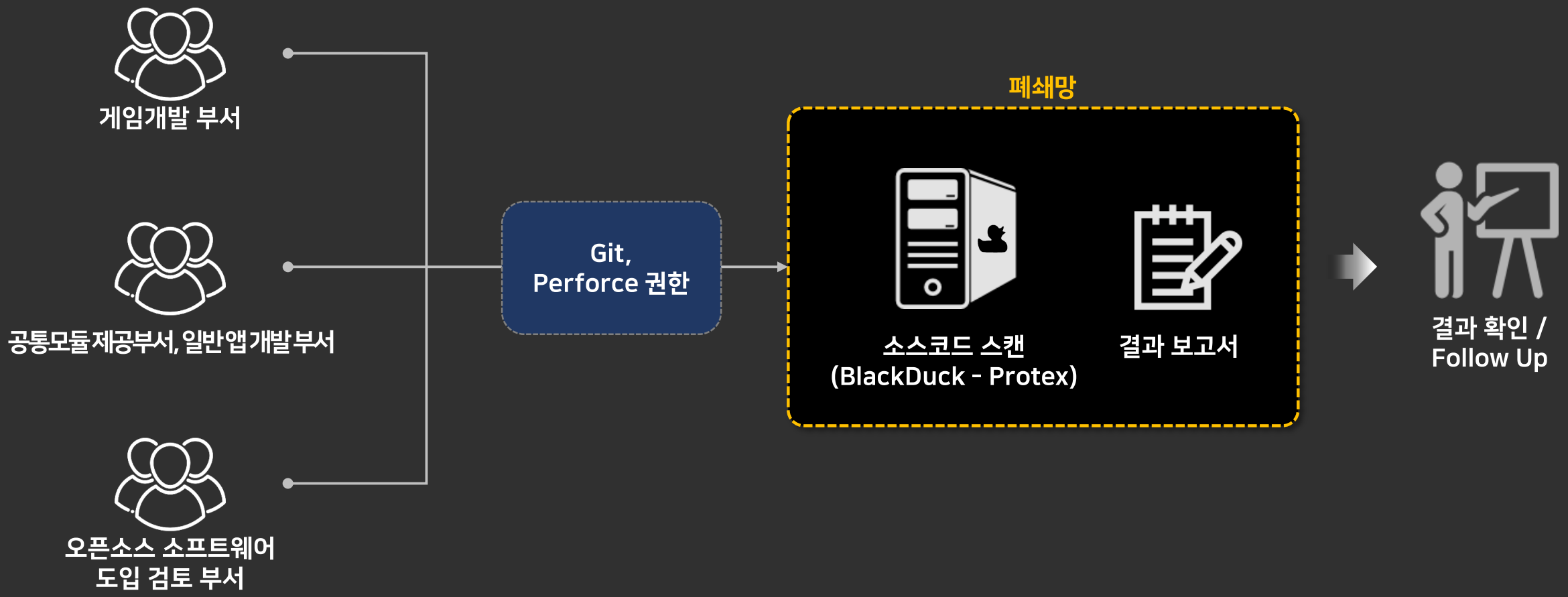

T Map
KT LTE 오후 1:57 100%
내 정보 설정
내 차 정보
승용차/휘발유/하이패스 사용
회원 정보
알림 설정
Push 알림 수신
SMS 수신
맞춤형 혜택 수신
나에게 딱 맞는 혜택을 받습니다.
광고성 정보 수신
이벤트 소식, 서비스 안내 등을 받습니다.
주의사항
이용약관
저작권 정보
설정 초기화

고지사례 (게임)


리니지M
고객센터
커뮤니티정책
앱 접근 권한 안내
유료 상품 확률 안내 (17' 11. 05 기준)
모바일 게임서비스 이벤트 규약
오픈소스 라이선스


Aion: Legions of War
ACCOUNT INFO
CUSTOMER SERVICE
Customer Service
CUSTOMER SERVICE
Terms of Service
TERMS OF SERVICE
Open Source License
VIEW
Current Version
RESET

소스코드 제공받기


게임개발 부서
공통모듈제공부서, 일반앱개발부서
오픈소스 소프트웨어
도입 검토 부서
Git,
Perforce 권한
폐쇄망
소스코드 스캔
(BlackDuck - Protex)
결과 보고서
결과 확인 /
Follow Up

검증의 방향성

1 **개발부서**에서 작성한
오픈소스 소프트웨어 사용 리스트

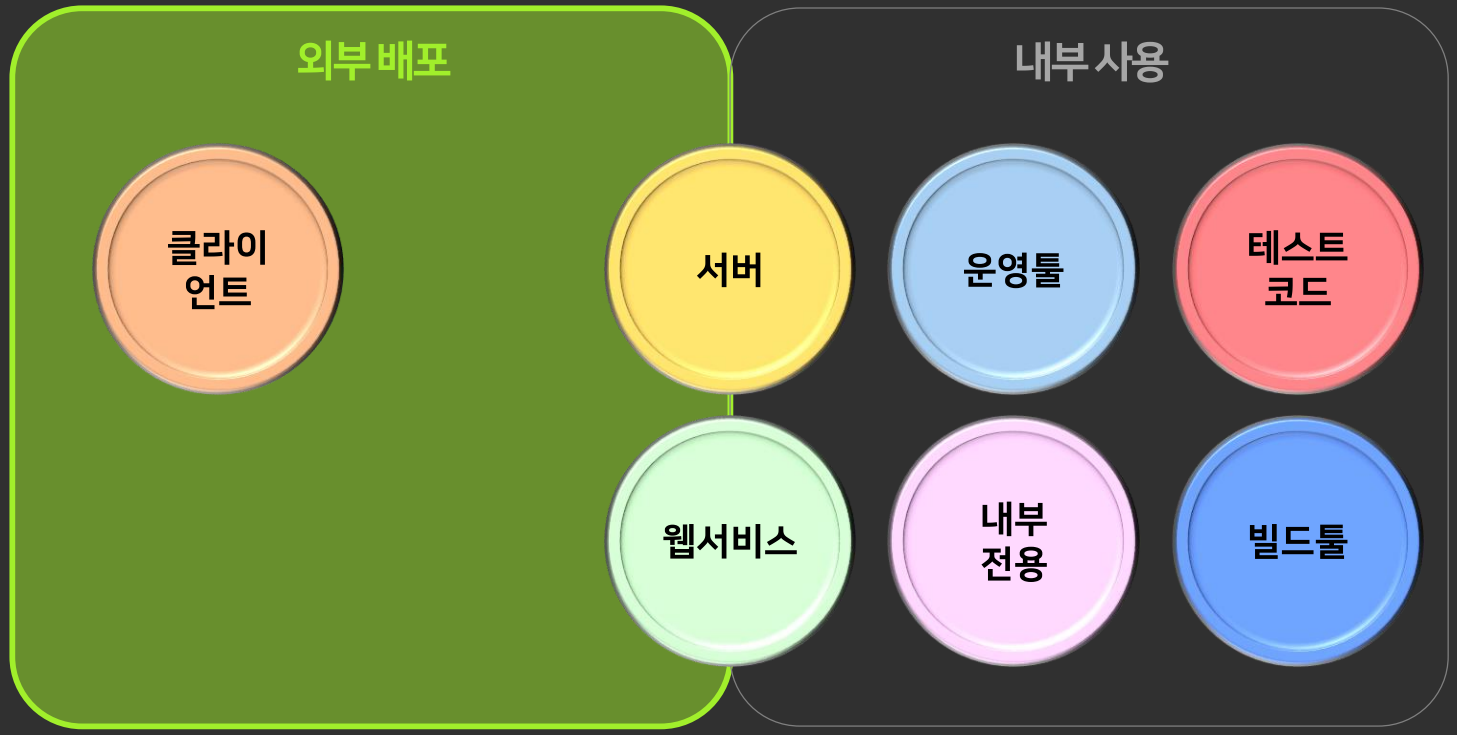
+

2 **검증부서**에서 스캔한
오픈소스 소프트웨어 사용 리스트

1 번과 2 번을 크로스 체크하여

누락되거나 잘못 파악되는 오픈소스 소프트웨어가 없도록 해보자

개발팀과 외부 배포의 대상인지 내부에서만 사용하는 코드인지 구분하였습니다.



- 클라이언트: 배포 대상에 무조건 포함
- 서버: 해외지사, 3rd Party, 퍼블리셔에게 제공되는 경우 배포로 보게 됨
- 웹서비스: AGPL은 네트워크로 연결되어 연동되는 프로그램의 전체 소스코드 공개를 하도록 하고 있음

검증 경과

현재 라이브 서비스 중인 모든 게임과 일반 앱에 대한 오픈소스 사용 고지를 완료하였습니다.

1

게임 앱

8종

2

일반 앱

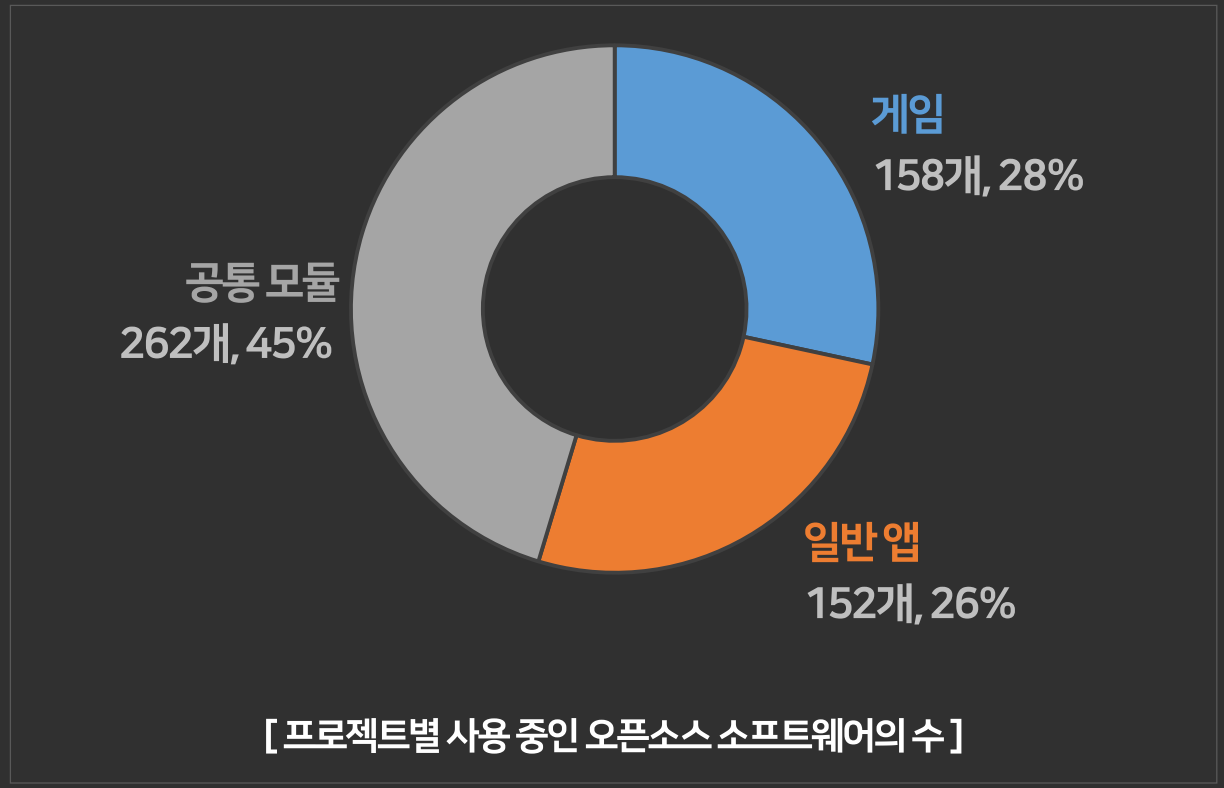
6종

3

공통 SDK

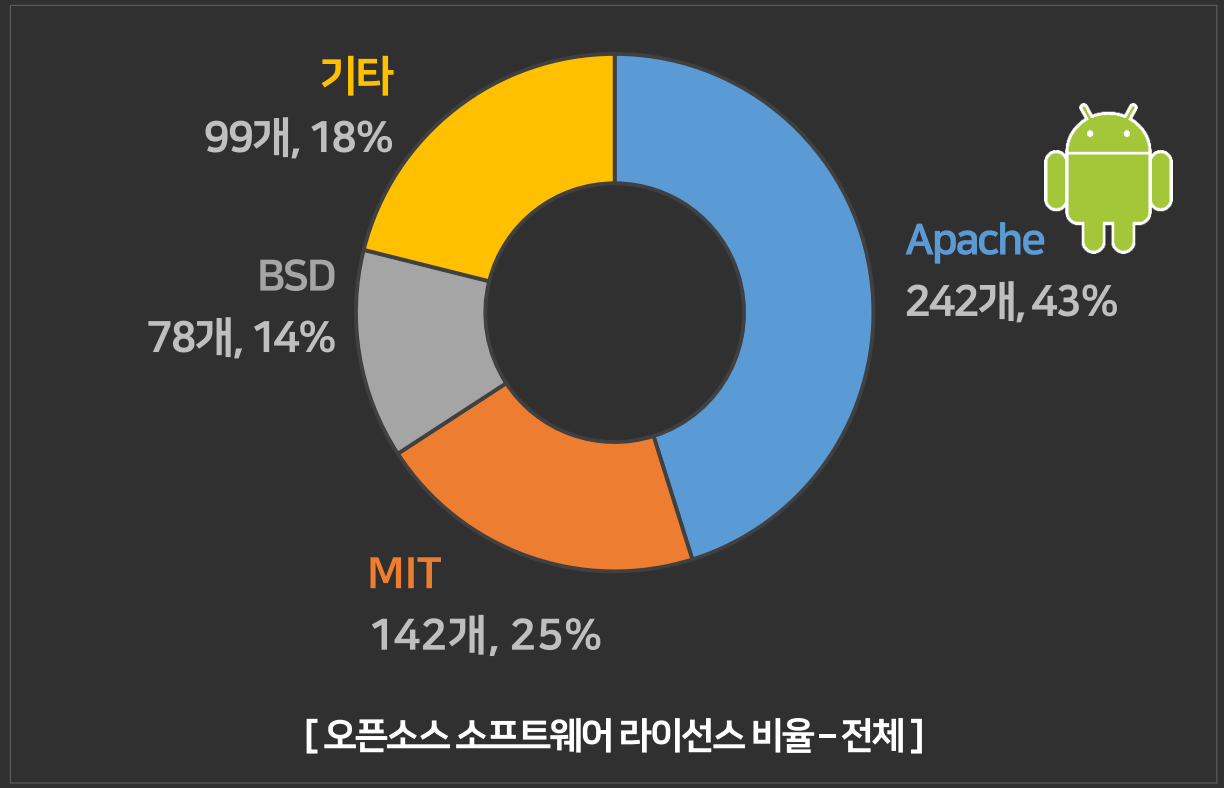
9종

프로젝트 별로 평균 20~30개 정도의 오픈소스 소프트웨어를 사용하고 있습니다.



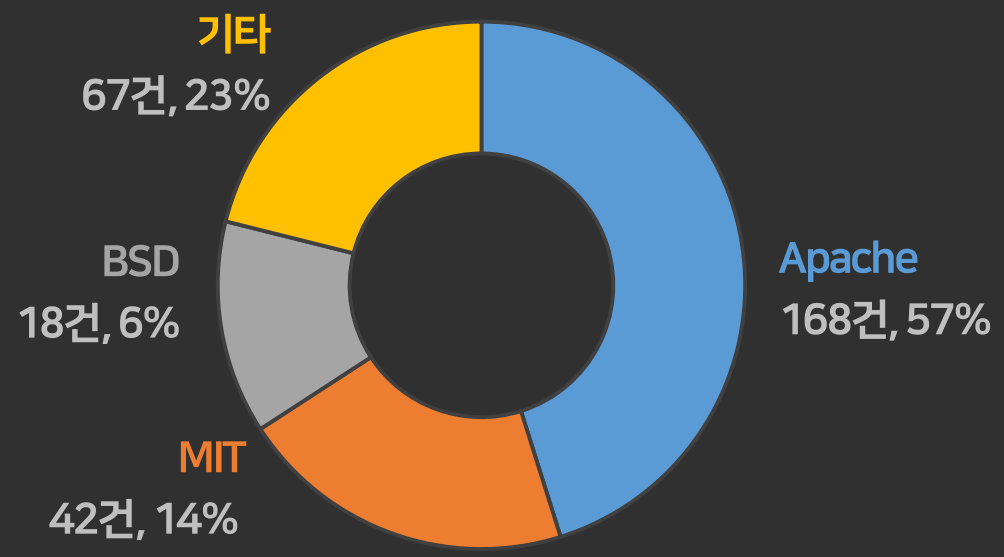
[프로젝트별 사용 중인 오픈소스 소프트웨어의 수]

모바일 프로젝트에서 Android 관련 오픈소스 사용이 많았기 때문에 Apache License 비율이 가장 높습니다.

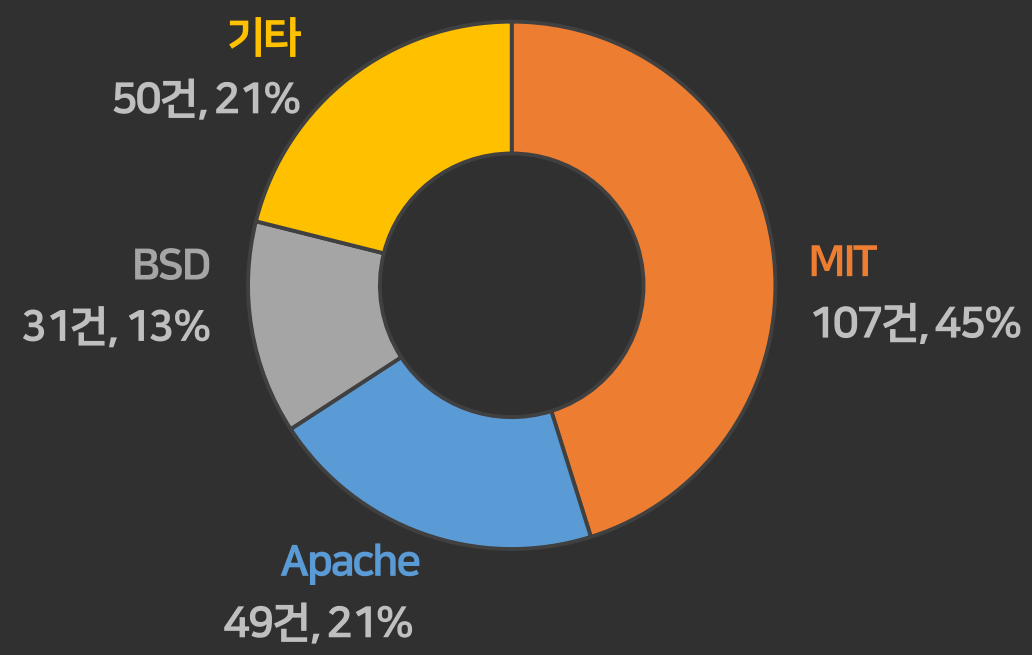


[오픈소스 소프트웨어 라이선스 비율-전체]

iOS의 경우 MIT License의 오픈소스 사용비율이 가장 높습니다.



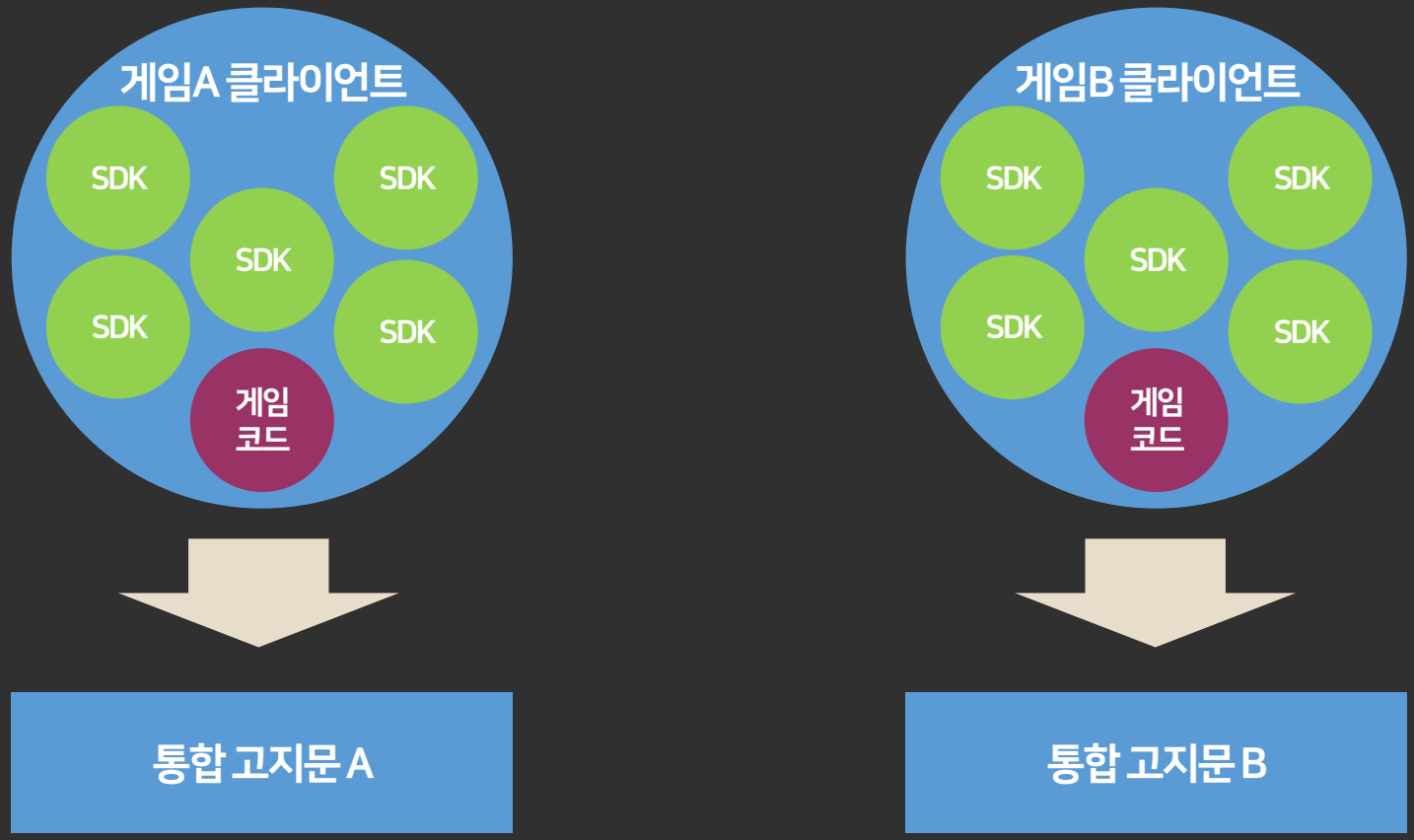
[오픈소스 소프트웨어 라이선스 비율 – Android]



[오픈소스 소프트웨어 라이선스 비율 – iOS]

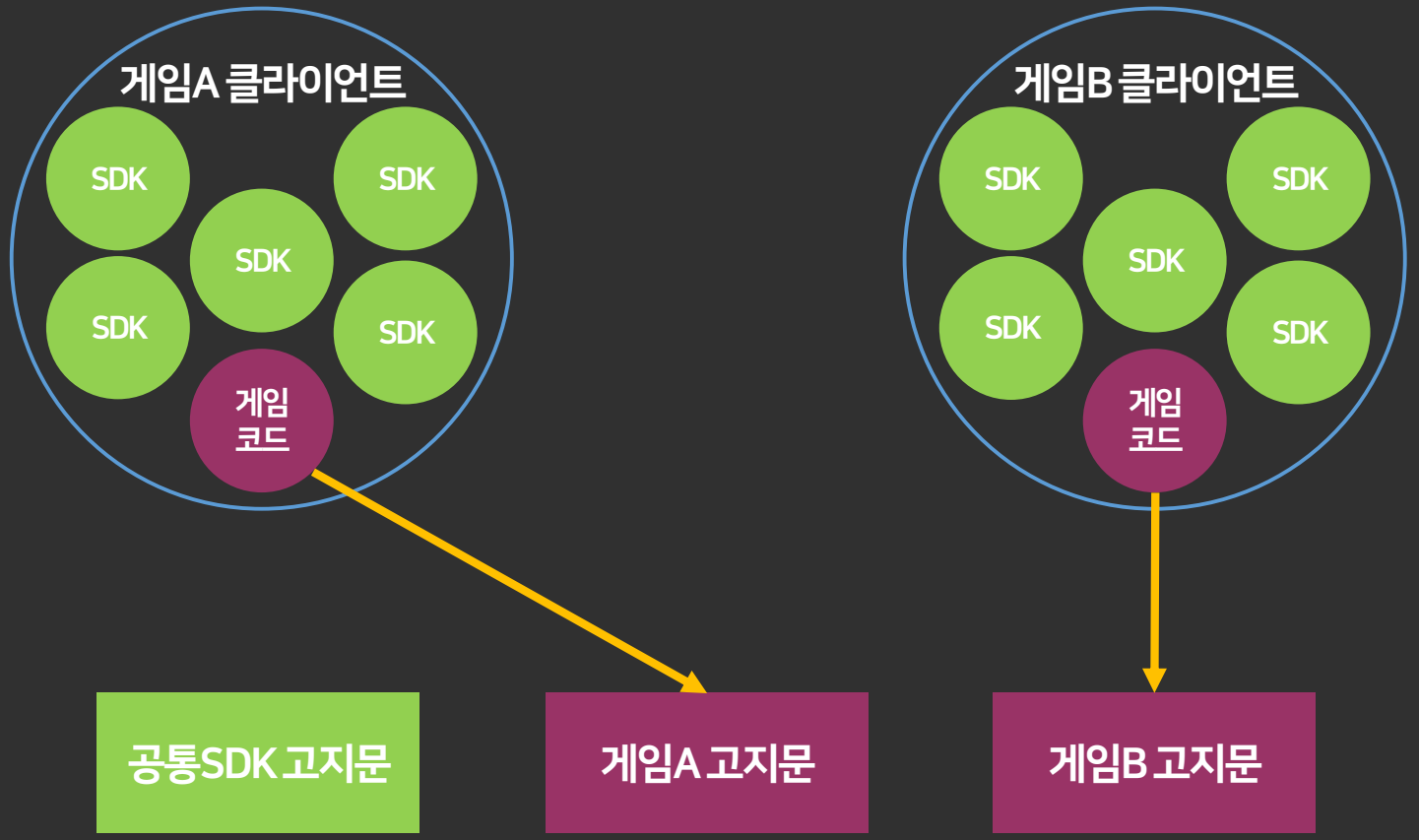
게임과 공통모듈의 분리

지금까지는 게임에 연동되는 SDK까지 개발부서와 커뮤니케이션하면서 검증을 진행했으나,



게임과 공통모듈의 분리

공통SDK를 별도로 고지하여 오픈소스 검증의 Scope을 축소하였습니다.



오픈소스 소프트웨어 라이선스 고지문

게임A 고지문

OSS Notice

Copyright (C) 2017 NCSOFT Corporation. All rights reserved.

This application uses Open Source Software (OSS). You can find the source code of these open source projects, along with applicable license information, below. Also please see this link (https://www.plaync.com/policy/api/view/mobile/shared_module) for NCSOFT shared modules. We are deeply grateful to these developers for their work and contributions.

Any questions about our licensed work can be sent to opensource@ncsoft.com.

aladdin MD5
https://enterprise.dejacode.com/licenses/public/aladdin-md5/#license-text
Copyright (C) 1999, 2002 Aladdin Enterprises
Aladdin MD5 License

공통SDK 고지문

OSS Notice

Copyright (C) 2017 NCSOFT Corporation. All rights reserved.

This is a list of Open Source Software (OSS) that is used for NCSOFT shared modules. You can find the source code of these open source projects, along with applicable license information, below. We are deeply grateful to these developers for their work and contributions.

Any questions about our licensed work can be sent to opensource@ncsoft.com.

AFNetworkActivityLogger
https://github.com/AFNetworking/AFNetworkActivityLogger
Copyright (c) 2013 AFNetworking
MIT License

Phase 3
Beyond Protex

- 소스코드를 오픈소스DB에 있는 코드와 비교하는 Protex와 달리
  오픈소스 내의 **특정 문자열을 검색해서 라이선스를 식별**해 주는 도구
- 2007년에 HP에서 공개했으며, 2015년에 github으로 코드 이관
- Linux Foundation에서 관리
- **Nomos, Monk라는 스캔 에이전트가 있음**
- 단점
  - directory명, file명, readme 파일 등을 통해서 직접 유추해야 함
    (오픈소스의 이름은 사용자가 이미 알고 있다고 가정)
  - License text가 삭제되어 있으면 검출 불가

[Fossology 검색 옵션]

10. Select optional analysis
    - [ ] Copyright/Email/URL/Author Analysis
    - [ ] ECC Analysis, scanning for text fragments potentially relevant for export control
    - [ ] Keyword Analysis
    - [ ] MIME-type Analysis (Determine mimetype of every file. Not needed for licenses or buckets)
    - [x] Monk License Analysis, scanning for licenses performing a text comparison
    - [x] Nomos License Analysis, scanning for licenses using regular expressions
    - [ ] Ojo License Analysis, scanning for licenses using SPDX-License-Identifier
    - [ ] Package Analysis (Parse package headers)

Nomos

- 키워드로 라이선스임을 식별하고, 특정 라이선스를 확인하기 위해 정규식으로 판단
- 정확히 판단 가능한 경우, BSD 3-Clause License, GPL 2.0 등으로 식별
- 라이선스 카테고리만 판단한 경우, BSD, GPL 등으로 식별

Monk

- 알려진 라이선스와 유사하지만 판단할 수 없는 라이선스는 -style 이라고 표현 (예. BSD Style)

[Nomos]

Display 25 licenses

Clear

Scanner Count	Concluded License Count	License Name
193	0	Apache-2.0
126	0	MIT
28	0	MIT-style
26	0	BSD-3-Clause
15	0	JSON
11	0	See-URL
10	0	ISC
8	0	BSD
5	0	MIT-possibility
2	0	WTFPL
2	0	Public-domain
2	0	Dual-license
1	0	UnclassifiedLicense
1	0	See-doc.OTHER
1	0	Public-domain-ref
1	0	MPL-2.0
1	0	MPL-1.1
1	0	MPL
1	0	LGPL-2.1+
1	0	ISC-possibility
1	0	GPL-2.0+
1	0	BSD-possibility
1	0	BSD-2-Clause
1	0	Apache-possibility

Showing 1 to 24 of 24 licenses Page 1 of 1

Hint: Click on the license name to search for where the license is found in the file listing.

Display 50 files (**tree view** or flat)

Clear

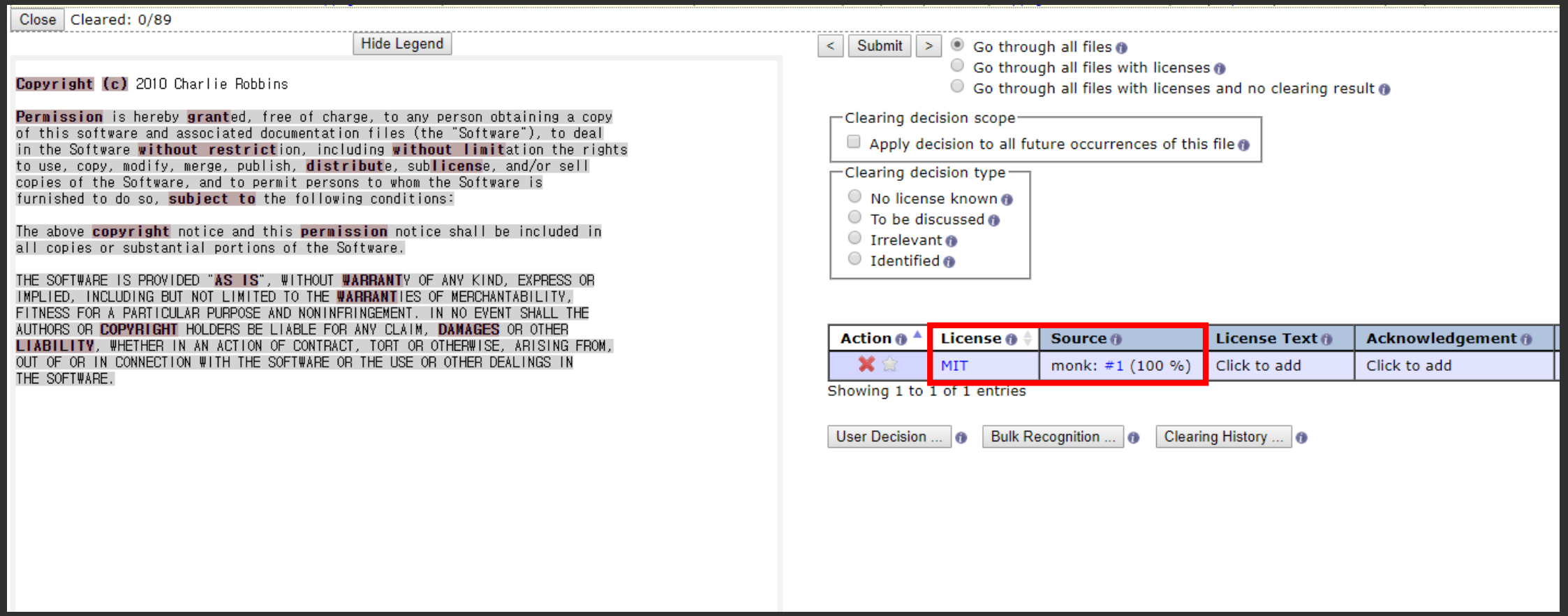
	-- filter for scan results --	-- filter for edited results --	open		MarkAsIrrelevant
Files	**Scanner Results (N: nomos, M: monk, Nk: ninka, I: reportImport, O: ojo)**	**Edited Results**	**Clearing Status**	**Files Cleared**	**Actions**
apidoc	Apache-2.0, BSD, BSD-2-Clause, Dual-license, ISC, ISC-possibility, MIT, MIT-possibility, No_license_found, Public-domain-ref, See-URL, WTFPL		🔴	0/151	[Tag][Edit][Bulk]
data	No_license_found		🟢	0/0	[Tag][Edit][Bulk]
example	No_license_found		🟢	0/0	[Tag][Edit][Bulk]
external	Apache-2.0, MIT-style, No_license_found		🔴	0/34	[Tag][Edit][Bulk]
extras	Apache-2.0, No_license_found		🔴	0/122	[Tag][Edit][Bulk]
gradle/wrapper	No_license_found, UnclassifiedLicense		🔴	0/1	[Tag][Edit][Bulk]
i18n	No_license_found		🟢	0/0	[Tag][Edit][Bulk]
keystore	No_license_found		🟢	0/0	[Tag][Edit][Bulk]
mopsdk	Apache-2.0, BSD, JSON, No_license_found, Public-domain		🔴	0/74	[Tag][Edit][Bulk]
node	Apache-2.0, Apache-possibility, BSD, BSD-3-Clause, BSD-possibility, Dual-license, GPL-2.0+, ISC, LGPL-2.1+, MIT, MIT-possibility, MPL, MPL-1.1, MPL-2.0, No_license_found, Public-domain, See-doc.OTHER		🔴	0/84	[Tag][Edit][Bulk]
script	No_license_found		🟢	0/0	[Tag][Edit][Bulk]
README.md	No_license_found [N]		🟢	0/0	[View][Info] [Download][Tag] [Edit]

Nomos

Monk

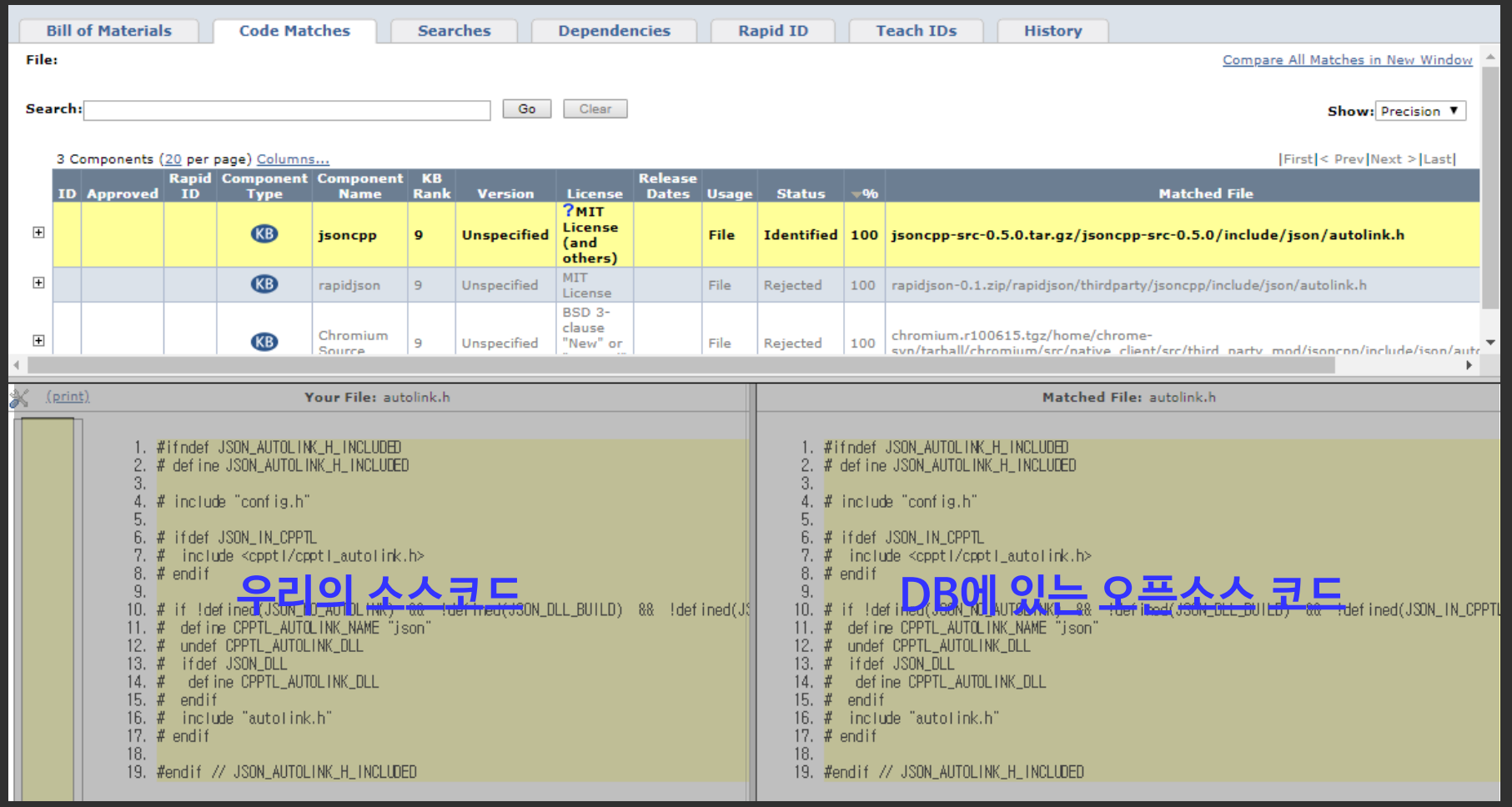
- 라이선스 전문의 텍스트 유사도를 비교하여 라이선스 판단
- Fossology의 라이선스 DB와 비교하여 일치하는 정도를 보여줌
- 라이선스 DB에 400개 정도 있으며, DB에 없는 라이선스는 탐지할 수 없음

[Monk]

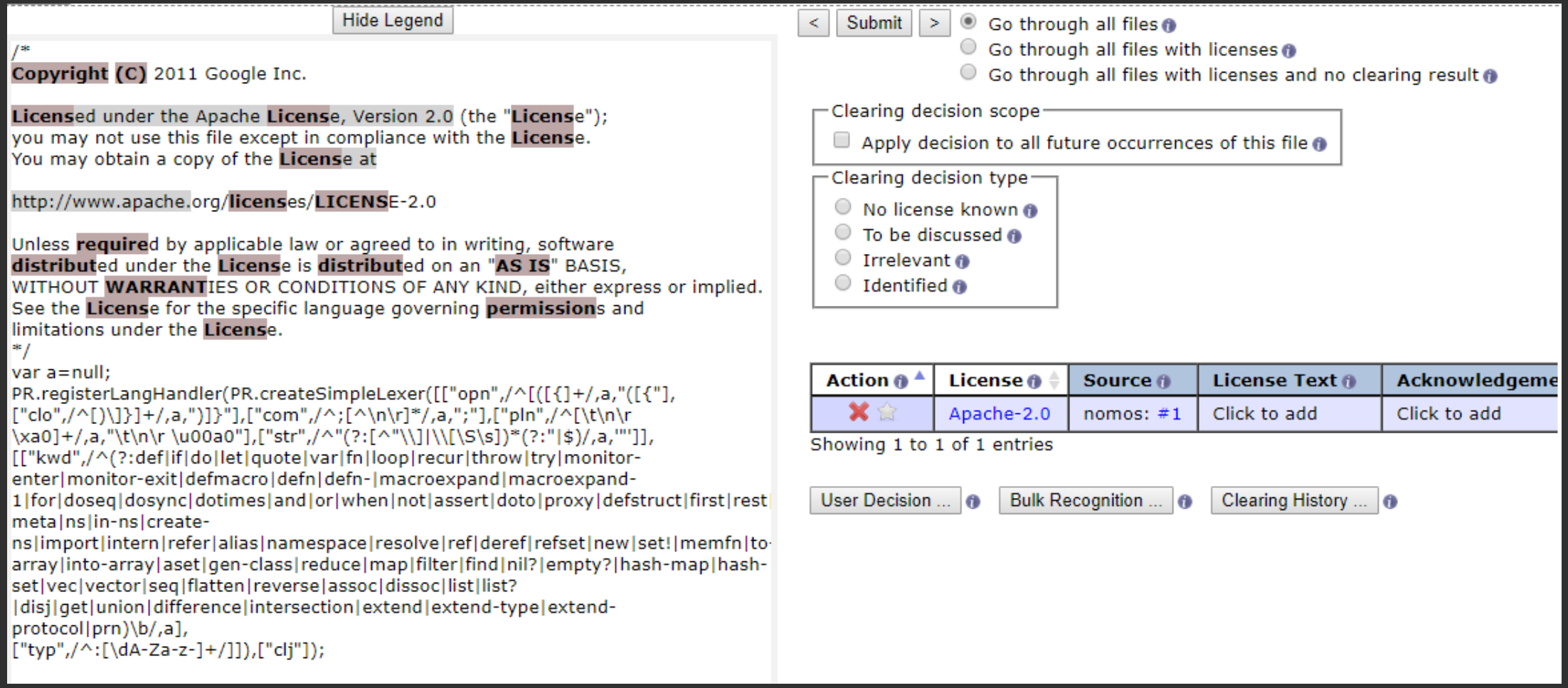

Close
Cleared: 0/89
Hide Legend
Copyright (c) 2010 Charlie Robbins
Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:
The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.
THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.
< Submit >
Go through all files
Go through all files with licenses
Go through all files with licenses and no clearing result
Clearing decision scope
Apply decision to all future occurrences of this file
Clearing decision type
No license known
To be discussed
Irrelevant
Identified
Action
License
Source
License Text
Acknowledgement
MIT
monk: #1 (100 %)
Click to add
Click to add
Showing 1 to 1 of 1 entries
User Decision ...
Bulk Recognition ...
Clearing History ...

생각보다 많은 회사에서 이미 사용하고 있습니다.

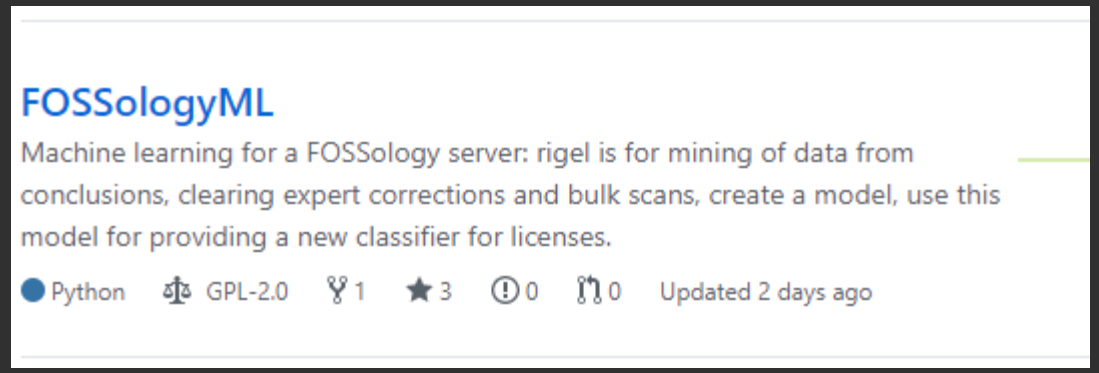
Protex는 소스코드와 DB에 있는 오픈소스 코드를 비교해서 많은 정보를 한번에 제공해주지만,



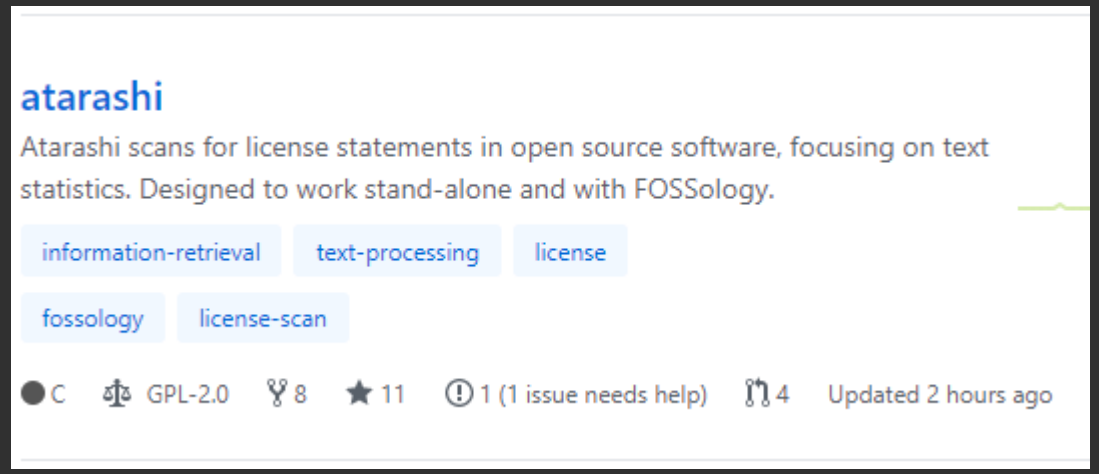
Fossology는 라이선스 텍스트 내의 특정 문자열만 검색해 주기 때문에 후작업이 많이 필요합니다.



그러나,Fossology 단점을 보완하기 위한 많은 하위 프로젝트들이 있어서 앞으로의 가능성은 무궁무진하며,



머신러닝에 의해 구동되는 라이선스 분류기 프로젝트



지금까지 룰 기반 탐지 방법의 오탐률을 낮추기 위한 새로운 검색 알고리즘

지금은 얻을 수 있는 정보가 라이선스와 copyright 뿐이지만,

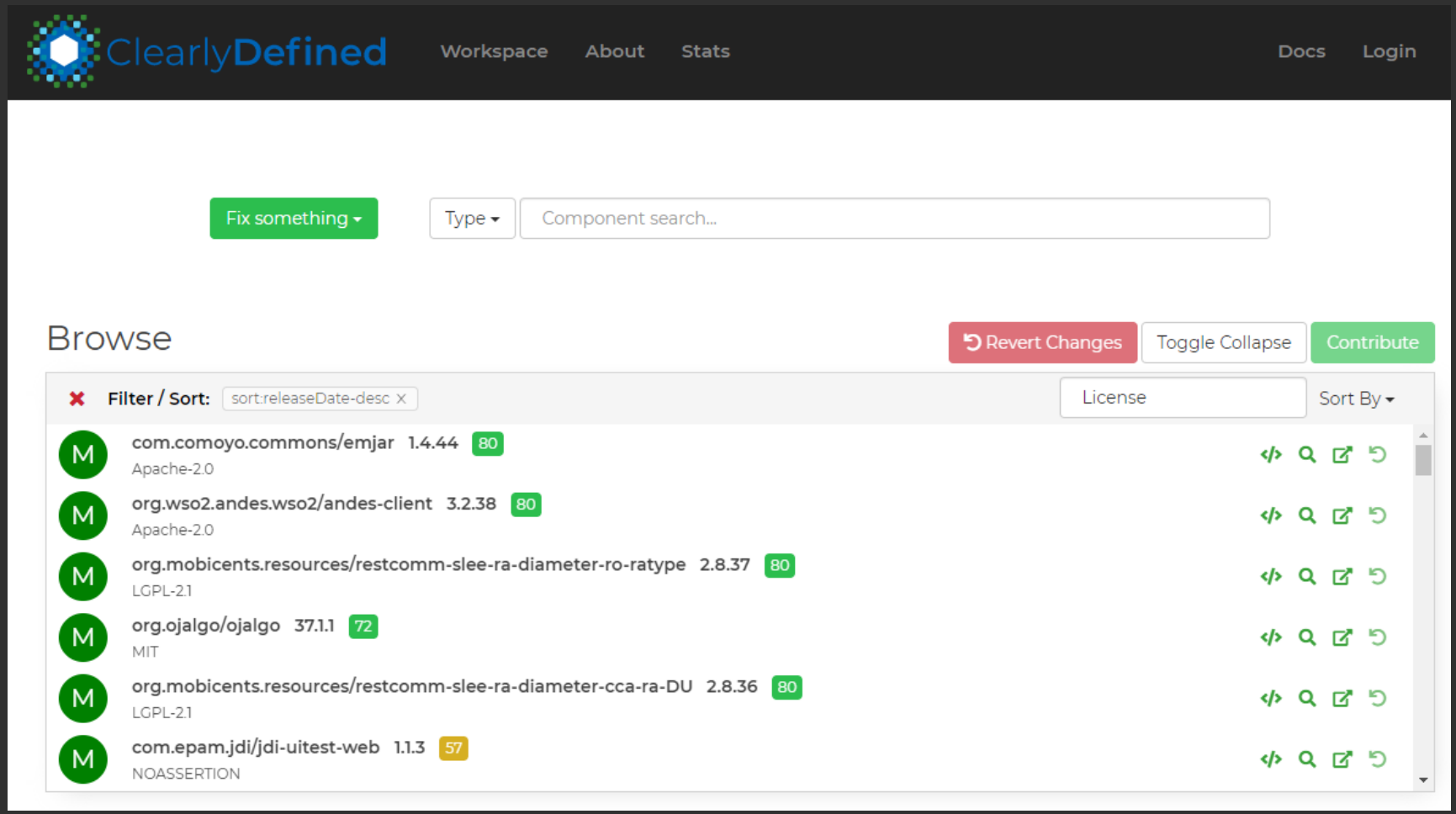
라이선스

Copyright

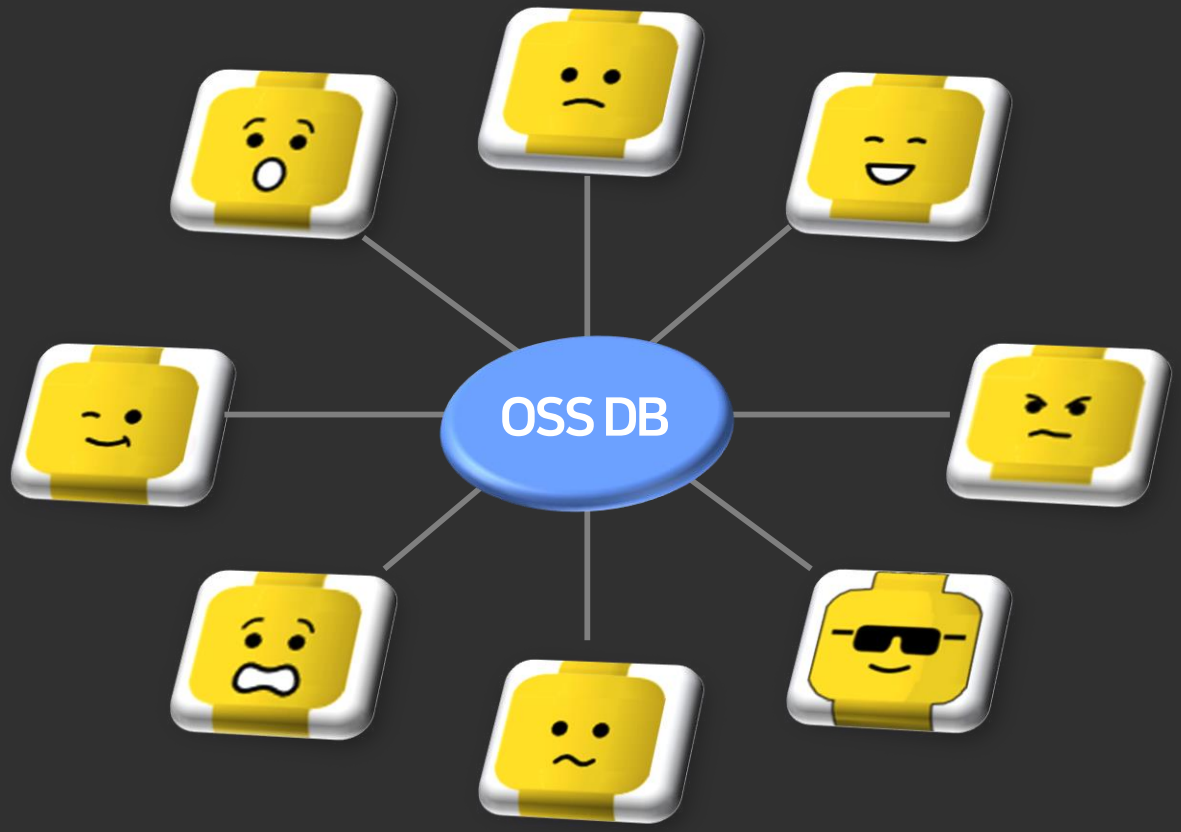
Clearly Defined 프로젝트에서 제공하는 DB와 연계한다면 더 많은 정보를 얻을 수 있을 것 같습니다.

- 라이선스
- Copyright
- 오픈소스의 이름
- 오픈소스의 버전
- 오픈소스의 URL

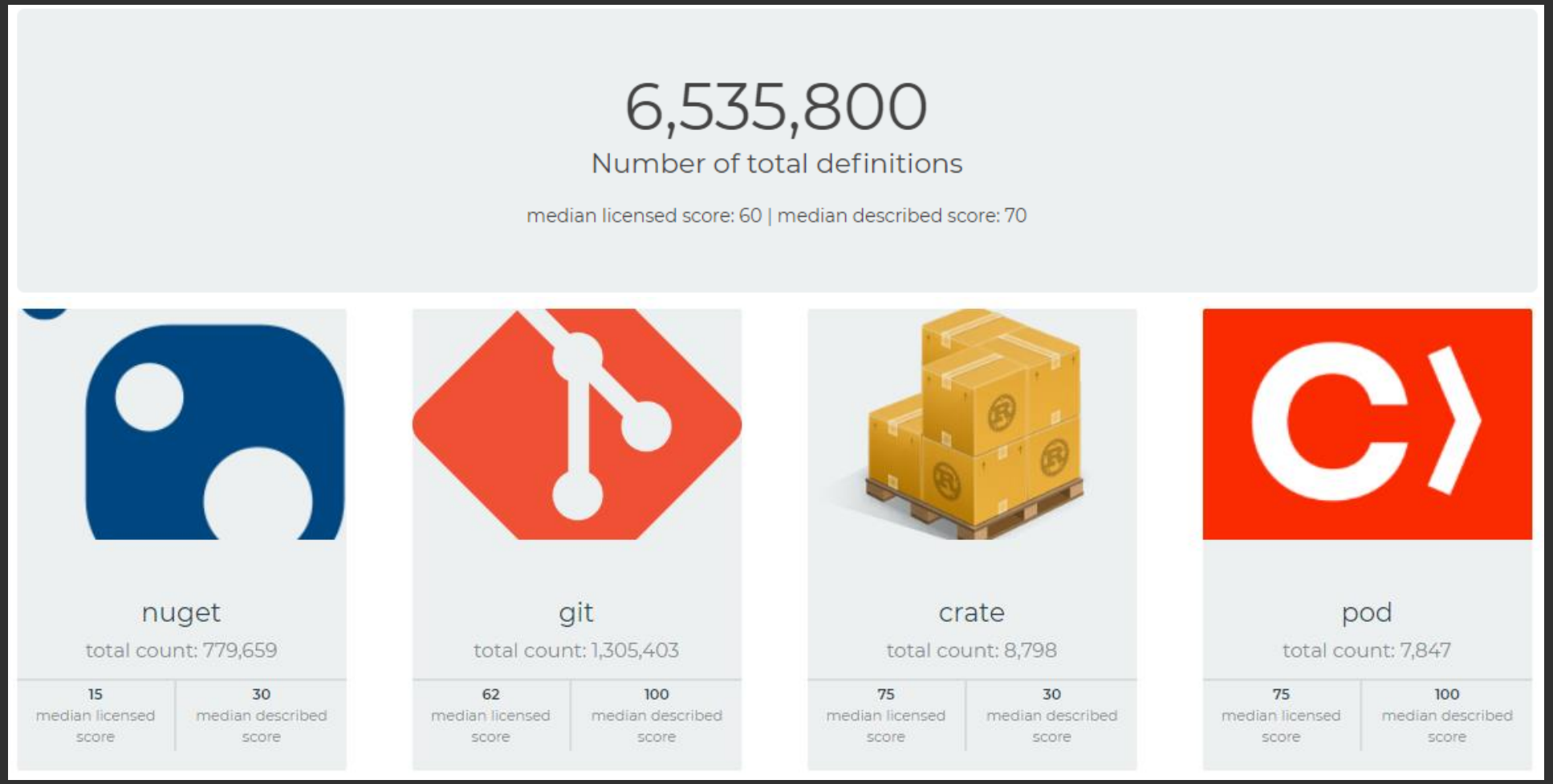
Clearly Defined는 오픈소스 데이터베이스를 만들고 있는 오픈소스 프로젝트인데



십시일반 여러 사람들이 오픈소스의 정보들(오픈소스 이름, 라이선스, 소스 위치 등)를 모아서
오픈소스의 데이터베이스를 구축하고 있습니다.



이미 오픈소스 데이터베이스가 꽤 많이 구축되어 있고,



누구나 데이터베이스를 사용할 수 있도록 API도 제공하고 있습니다.

Data API

The ClearlyDefined service manages both raw, harvested data and curated data, as well as the merge of these. These data can be expressed in relation to source code (e.g., a GitHub repo) or a package (e.g., an NPM, RPM, Maven project, ...). One of the key goals of ClearlyDefined is to correlate the *binary* package with the original source.

> *A quick note on binary. Throughout the ClearlyDefined ecosystem we talk about binary as being the packaged, executable form of a component. An NPM for example is a binary despite the fact it may contain human-readable text that looks a lot like JavaScript source code. In general, the original source for these packages may have been in a very different language (e.g., TypeScript) or the package content may have been minimized, compresses, concatenated, or otherwise swizzled. For the purposes of license detection and ultimately compliance, as well as security scanning etc, consumers need to know the location of the actual developer-authored source code.*

As a result of this separation, the same component may show up in the data in several forms – the NPM and its source are both treated separately. These components may also have different *revision* identifiers (e.g., NPM version and Git commit hash). There are links between the different types and as the ecosystem progresses, this web of components will get richer and enable transitive operations, for example, given a vulnerability in a GitHub repo you will be able to find all the packaged versions and forms that included the vulnerable code.

You can see the swagger API doc at https://api.clearlydefined.io/api-docs/

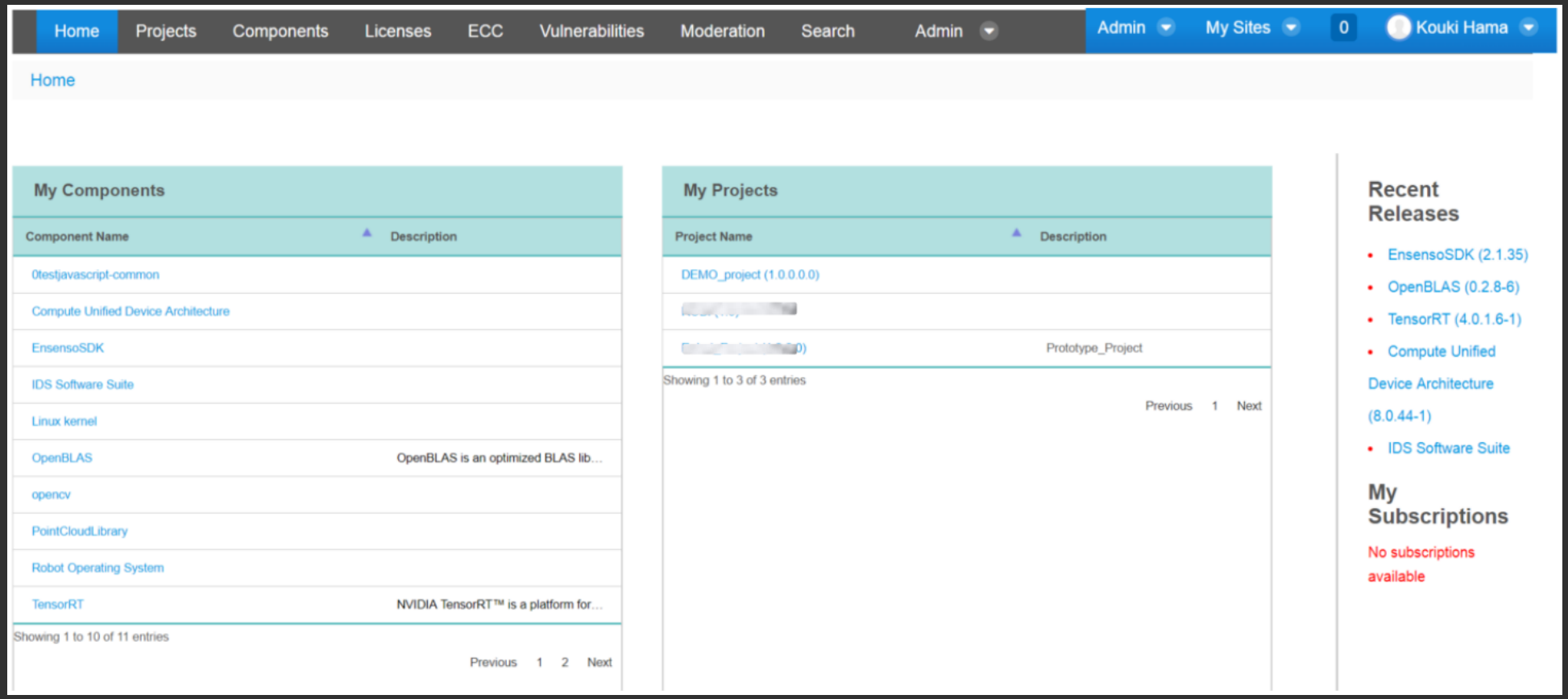
Curation

New curations, or changes to existing curations, are done via PATCHes. Ultimately these surface as PRs in the configured curation repo. They can be manipulated directy there but using this API keeps things regular. Below is an example curation.

```
{
  "described": {
    "sourceLocation": {
      "type": "git",
      "provider": "github",
      "url": "https://github.com/microsoft/redie",
      "revision": "194269b5b7010ad6f8dc4ef608c88128615031ca"
    }
  },
```

- 소프트웨어 카탈로그
- (Kouki Hama 라는 분의 자료에 의하면)
  - 프로젝트 정보 관리
  - 소프트웨어 컴포넌트 관리
  - 취약점 정보 관리
  - Fossology에서 분석한 라이선스 정보 관리 (Connection with FOSSology)
  - csv 파일을 이용해서 일괄 업로드 가능

엑셀을 이용해서 관리할 수 있겠지만 시스템을 통한 관리 방법이 주는 장점이 분명히 있으며
사내 공유를 통해 중복 조사를 방지할 수 있으니 비용도 절감할 수 있습니다.



아직은 파악해야 할 것들이 많지만, 이 세 가지를 잘 파고들어 본다면 꽤 괜찮은 도구가 나올 것 같습니다.

"오픈소스를 오픈소스로 관리해 보자"

Q & A

오픈소스 소프트웨어 라이선스 검증記

The End of Document