Intellectual Property & Provenance

CS489
Shin Yoo

Open Source SW are free to use.

Copyleft means you let go of your copyright.

IP?

- Tangible things such as ideas, inventions, technologies, artworks, music, and literature to which one can claim ownership.

- Copyright, Patent, Trade Secret

Patent & Trade Secret

- Patent: you get an exclusive right to a technology at the cost of making it public

- Trade Secret: you protect a hidden technology. Due to secrecy, it is much harder to characterise formally.

  - Extent of outside knowledge, extent of measures taken by individuals to protect the secret, the value of the information to the owner and the competitor, the cost of development, the difficulty of duplicating the information

Copyright

- Copyright protects *expression:* in general, any original work that has a tangible form and is fixed in a medium is protectable under copyright law (Kizza).

- First established as the right of an author in 1710 in Great Britain.

- Right of the author is protected for a pre-defined duration, after which the work goes into the public domain.

Expiration & Public Domain

- Berne Convention (1886): protects the copyright for 50 years after the death of the author

- US Copyright Law: was life + 50 years, and later extended to life + 70 years

- Korea: was life + 50 years, and later extended to life + 70 years with Korea-USA FTA :)

- Public Domain: work in public domain are not protected by copyright law and can be used by any member of the public without permission

Internet Archive Gets DMCA Exemption To Help Archive Vintage Software

Internet Archive Gets DMCA Exemption To Help Archive Vintage Software

In 2003 the Internet Archive, as part of research into vintage software archiving, discovered possible archiving issues involving the Digital Millenium Copyright Act. This could make it impossible to legally archive early computer software and games, even for accredited institutions wishing to store limited amounts of non-distributable, archival images.

It's vital to make proper archival copies of these artefacts, because the life of magnetic media such as floppy discs has been estimated at 10 to 30 years. Time is running out to properly archive much of this large body of work for safekeeping, to ensure it lives out its term of copyright and is available (in the short-term, under suitable copyright-constrained means) for posterity.

The Copyright Office holds a rulemaking proceeding every 3 years to:

"determine whether there are particular classes of works as to which users are, or are likely to be, adversely affected in their ability to make noninfringing uses due to the prohibition on circumvention of access controls."

As part of this rulemaking process, the Internet Archive submitted an initial comment in early 2003, and followed this up with a reply comment giving further examples of classic software that might be lost if access controls could not be circumvented.

Following deliberation, the Copyright Office ruled in late October 2003 that four exemptions should be added to the anti-circumvention clause of the DMCA, to be valid until the next Copyright Office rulemaking in 2006, including two that are related to the Internet Archive's original comments:

- *Computer programs protected by dongles that prevent access due to malfunction or damage and which are obsolete.*
- *Computer programs and video games distributed in formats that have become obsolete and which require the original media or hardware as a condition of access.*

With the aid of these exemptions, the Internet Archive is continuing its work with institutional and technical partners to research and archive this at-risk software, and would like to thank all those who worked hard to help us achieve our goal.

Eligibility

- US Copyright Office: originality, fixation, expression
    - Originality: Facts cannot be copyrighted, known or unknown, because it cannot be invented. Theories cannot be copyrighted. Ideas are common property and cannot be protected: ideas acquire originality through expression.
    - Fixation: we inherit this from Gutenberg (i.e., the thing protected is printed and therefore *fixed*). There is inherent tension between the concept of fixation and digital contents.

License

- Copyright license is a form of contract (“permissions agreement”) between the holder of the copyright and someone who wants to use the protected work.

- Public copyright license: blanket license that grants permissions to anyone in the general public. The holder can decide to apply public copyright license.

User (Licensee) Right	Berne	CC0
0. free access, and freedom to use the work as you wish ("use" includes to run a program or to execute a music score)	Partial	Yes
1. freedom to access the "source-code" and use it as you wish, for study or change it for personal use.	No	Yes
2. freedom to redistribute copies	No	Yes
2.1 right to quote (freedom to redistribute copies of fragments)	Yes (small amount)	Yes (any amount)
3. freedom to distribute copies of your modified versions to others	No	Yes

CC

- American Non-profit Organisation, devoted to expanding the range of creative works available for others to build upon legally and to share (https://creativecommons.org/faq/)

- Provides template licenses that people can easily adopt. Templates have evolved over time.

CC Conditions

- Attribution (by): All CC licenses require that others who use your work in any way must give you credit the way you request, but not in a way that suggests you endorse them or their use. If they want to use your work without giving you credit or for endorsement purposes, they must get your permission first.

- ShareAlike (sa): You let others copy, distribute, display, perform, and modify your work, as long as they distribute any modified work on the same terms. If they want to distribute modified works under other terms, they must get your permission first.

CC Conditions

- NonCommercial (nc): You let others copy, distribute, display, perform, and (unless you have chosen NoDerivatives) modify and use your work for any purpose other than commercially unless they get your permission first.

- NoDerivatives (nd): You let others copy, distribute, display and perform only original copies of your work. If they want to modify your work, they must get your permission first.

CC Licenses

- CC0
- Attribution (CC BY)
- Attribution ShareAlike (CC BY-SA)
- Attribution-NoDerives (CC BY-ND)
- Attribution-NoCommercial (CC BY-NC)
- Attribution-NoCommercial-ShareAlike (CC BY-NC-SA)
- Attribution-NoCommercial-NoDerives (CC BY-NC-ND)

CopyLeft

- The simplest way to make a program free software is to put it in the public domain, uncopyrighted. This allows people to share the program and their improvements, if they are so minded. But it also allows uncooperative people to convert the program into proprietary software. They can make changes, many or few, and distribute the result as a proprietary product.

- To copyleft a program, we first state that it is copyrighted; then we add distribution terms, which are a legal instrument that gives everyone the rights to use, modify, and redistribute the program's code, or any program derived from it, but only if the distribution terms are unchanged. Thus, the code and the freedoms become legally inseparable.

- Copyleft is a way of using the copyright on the program. It doesn't mean abandoning the copyright; in fact, doing so would make copyleft impossible. The “left” in “copyleft” is not a reference to the verb “to leave” —only to the direction which is the mirror image of “right”.

https://www.gnu.org/copyleft/

Famous Open Source Licences

License	Linking	Distribution	Modification	Sublicensing
GPL v3	Only GPL v3	Copylefted	Copylefted	Copylefted
GPL Lesser	Under	Copylefted	Copylefted	Copylefted
Apac				ve
MIT	Permissive	Permissive	Permissive	Permissive
Mozilla Public License	Permissive	Copylefted	Copylefted	Copylefted
Beerware	Permissive	Permissive	Permissive	Permissive

```
/*
 * ----------------------------------------------------------------------------
 * "THE BEER-WARE LICENSE" (Revision 42):
 * <phk@FreeBSD.ORG> wrote this file.  As long as you retain this notice you
 * can do whatever you want with this stuff. If we meet some day, and you think
 * this stuff is worth it, you can buy me a beer in return.   Poul-Henning Kamp
 * ----------------------------------------------------------------------------
 */
```

Oracle vs. Google

- Java was developed by Sun Microsystems: it includes a programming language, a VM, and a set of libraries that are documented via APIs.

- Java was released in 1995 under Sun Community Source License
    - Source code is freely available
    - Commercial derivatives should be licensed by Sun

- Later, Sun changed various Java package license to GPL, with exceptions for linking. This led to OpenJDK, an open source implementation of Java SE, initially led by Sun.

Oracle vs. Google

- Android was founded in 2003, and purchased by Google in 2005. Google wanted to license Java SE for its platform, but the negotiation failed.
    - Google claims Sun wanted a shared control over licensed part, which would have made it difficult to open source Android system
    - Oracle claims that Google wanted to fork Java and make it incompatible with the remaining ecosystem.
- Since OpenJDK was not so mature at this point, Google decided to implement Java SE libraries from the scratch, without any access to Sun code. They did reuse part of Apache Harmony, another clean-room reimplementation of Java done by Apache Foundation.

Oracle vs. Google

- Oracle finished acquisition of Sun in January 2010.

- In August 2010, Oracle sued Google for copyright and patent infringement. Oracle asserted Google was aware that they had developed Android without a Java license and copied its APIs, creating the copyright violation.

- The court separated the case into three parts: copyright, patent, and damage.

- On copyright case, Oracle alleged infringement of 37 separate APIs in Android that have derived from Apache Harmony project (by now Google took over Harmony as well).

```
private static void rangeCheck(int arrayLen, int fromIndex, int toIndex) {
     if (fromIndex > toIndex)
          throw new IllegalArgumentException("fromIndex(" + fromIndex +
               ") > toIndex(" + toIndex+")");
     if (fromIndex < 0)
          throw new ArrayIndexOutOfBoundsException(fromIndex);
     if (toIndex > arrayLen)
          throw new ArrayIndexOutOfBoundsException(toIndex);
}
```

Initial verdict from Judge Alsup, District Court of Northern California

“So long as the specific code used to implement a method is different, anyone is free under the Copyright Act to write his or her own code to carry out exactly the same function or specification of any methods used in the Java API. It does not matter that the declaration or method header lines are identical."

https://www.theverge.com/2017/10/19/16503076/oracle-vs-google-judge-william-alsup-interview-waymo-uber

Oracle vs. Google

- Oracle appealed to the federal court, which reversed the initial verdict.

- The court noted that Copyright Act provides protection to **"original works of authorship fixed in any tangible medium of expression"** (p. 17). The legislative history explains that literary works **include "computer programs to the extent that they incorporate authorship in the programmer's expression of original ideas, as distinguished from the ideas themselves"** (p. 18). To qualify for copyright protection a work must be original. 17 U.S.C. § 102(a). The court was therefore **"first to assess whether the expression is original to the programmer"** (p. 24), something that Google had already conceded (p. 21). This led the court to **conclude "that the overall structure of Oracle's API packages is creative, original, and resembles a taxonomy"** (p. 14). It therefore reversed the first instance's decision on the central issue, **holding that the "structure, sequence and organization" of an API is copyrightable**.

- Remanded to district court to consider whether Googles's usage was fair use.

Oracle vs. Google

- District Court again ruled in favour of Google. Oracle appealed.

- The Appeals Court found that Google's use of API code declarations had not met any of the four current criteria for fair use, but was merely untransformed reuse.

- Google has appealed to the supreme court.

Oracle vs. Google

- New in 2021: Supreme Court ruled in favor of Google!
    - “APIs serve as declaring code rather than implementation - it serves organizational function”
    - “Google only used 0.4% of total Java source code”
    - "Google copied those lines not because of their creativity, their beauty, or even (in a sense) because of their purpose. It copied them because programmers had already learned to work with [Java SE], and it would have been difficult … to attract programmers to … Android … without them.”
    - https://en.wikipedia.org/wiki/Google_LLC_v._Oracle_America,_Inc.#Decision

Fair Use Criteria

- Fair use is a copyright principle based on the belief that the public is entitled to freely use portions of copyrighted materials for purposes of commentary and criticism.

- No clear decision process; has to be resolved in court. The judges will consider the following points:

  - Transformative Factor: have you added value?

  - Nature of copyrighted work: is it factual, or fictional?

  - The amount of copying: how much did you take?

  - Effect of use on the market: did you harm the copyright holder?

Structure, Sequence, and Organisation (SSO)

- A term used in the United States to define a basis for comparing one software work to another in order to determine if copying has occurred that infringes on copyright, even when the second work is not a literal copy of the first. (https://en.wikipedia.org/wiki/Structure,_sequence_and_organization)

Code Provenance

- Provenance: *n.* the place of origin or earliest known history of something

- In many cases, IP decisions boils down to where code came from, at the lowest level.

- “Have A copied the code X from B?”

- Copied code are called *code clones*.

Clone Detection

- Clones have been studied for multiple reasons.
  - Provenance in the legal context
  - Plagiarism as a specific context of provenance, as well as in the educational context
  - Productivity as there are various views that clones affect software development lifecycle

Types of Clones

- Type 1: exact copy without modifications (except for whitespaces and comments)
- Type 2: syntactically identical copy, with variable names, types, and/or function identifiers changed
- Type 3: copy with further modifications such as swapped line order, etc
- Type 4: semantically identical computation but written in a different logic

Comparison Methods

- Textual Comparison: simply compare line by line - brittle but also language agnostic.

- Token Comparison: compare lines as sequences of tokens (concrete values are abstracted)

- Metric Comparison: collect a set of metrics about code, and compare the metric vectors instead of actual code

- Abstract Syntax Tree: partition AST subtrees by hash and use tree matching algorithm

- Program Dependence Graph (PDG): use graph matching algorithm to compare PDGs - approximative matching since it is NP-hard

Checking for License Violation

- We will take the following paper as an example of how lower level techniques are assembled to detect license violation in projects:
  - O. Mlouki, F. Khomh, and G. Antoniol. On the detection of licenses violations in the **Android** ecosystem. In 2016 IEEE 23rd International Conference on Software Analysis, Evolution, and Reengineering (SANER), volume 1, pages 382–392, March 2016.

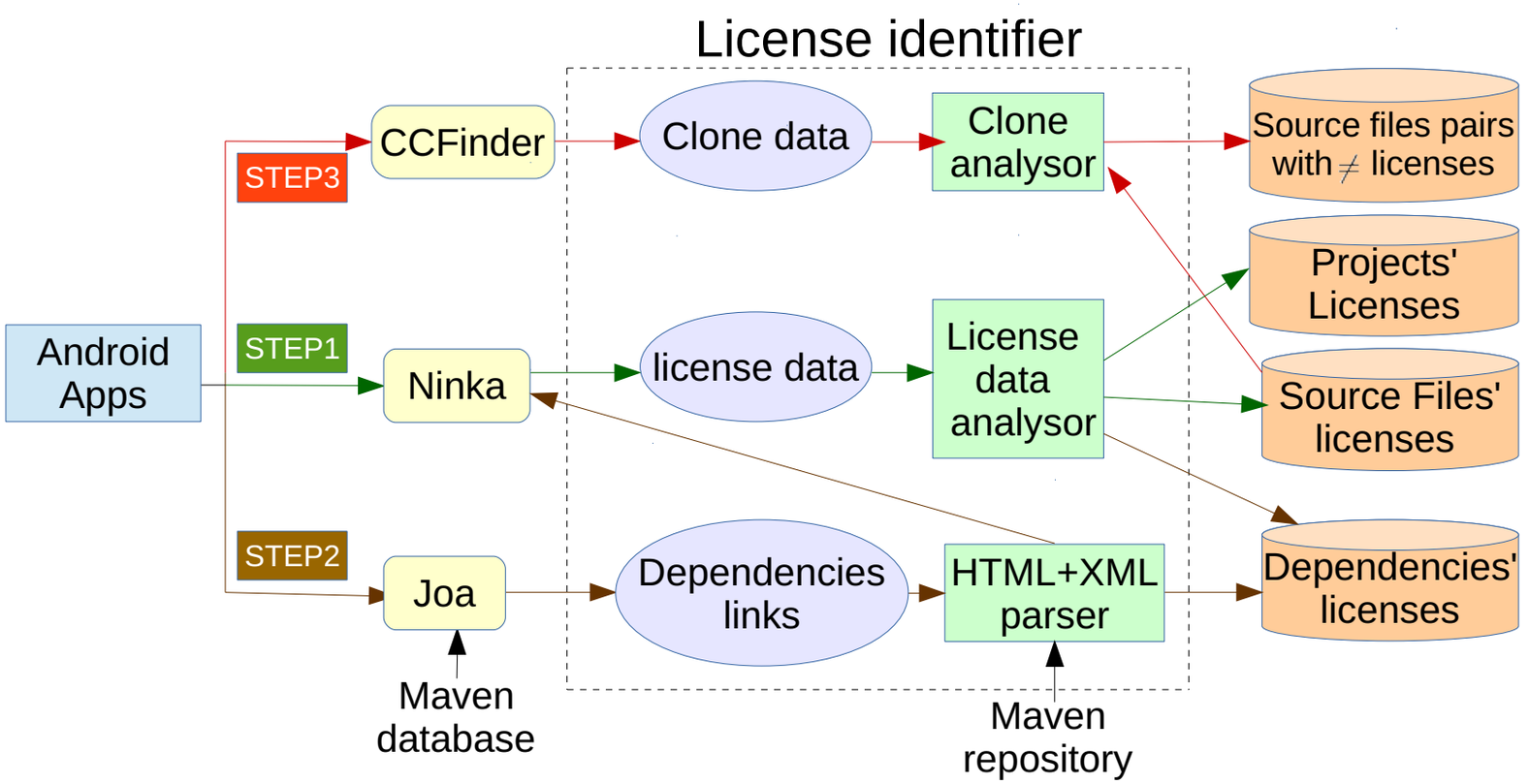


Fig. 1: Identification of license information

- Ninka: license header extraction tool
- Joa: library identification tool
- CCFinder: clone detection tool (to detect files that are cloned under different licenses)

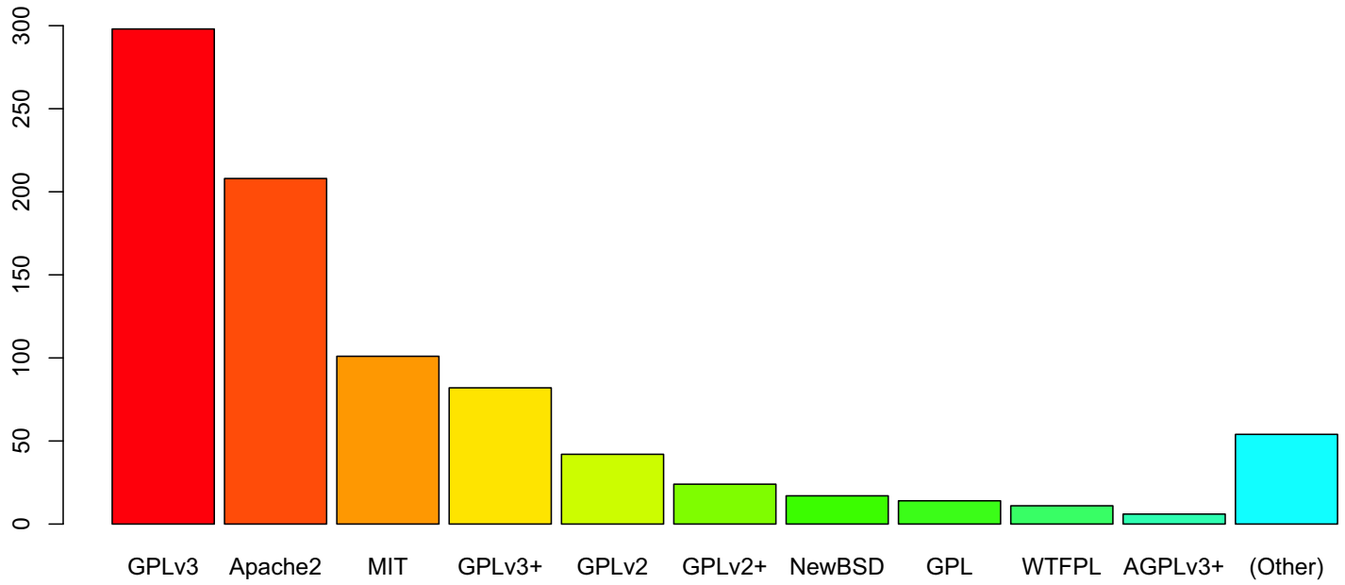


Fig. 4: Projects licenses when considering only the latest release of each app

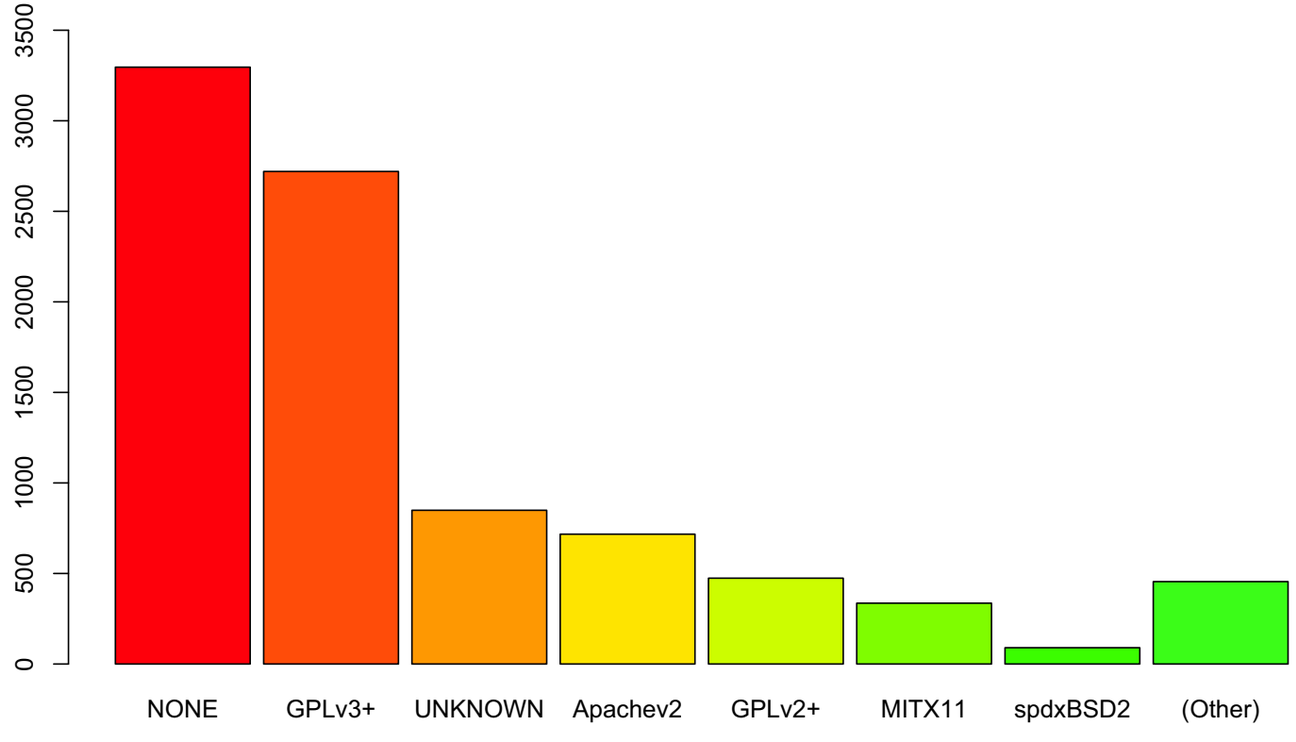


Fig. 5: Projects licenses when considering all the releases of the apps

	spdxBSD3	3 655	
	spdxBSD4	130	
MIT	oldwithoutSelland	5	0,87
	oldwithoutSelland NoDocumentationRequi	147	
	MITVariant	1	
	MITX11	11 965	
	X11BSDvar	4	
	X11noNotice	343	
PublicDomain		2 957	0,21
artifex		136	0,01
BeerWareVer42		6	0
GDDLorGPLv2		475	0,03
CPLv1		20	0
DoWTFYWv2		48	0
EPLv1		13	0
FreeType		1 714	0,12
MPLv1_1		4	0

Distribution of Licenses Detected in Files

Violations

- Out of 857 Android apps from F-Droid,
  - 17 apps showed clear license violation (in total of 229 releases)
  - only 10 out of 17 apps eventually fixed the violation
  - on average, it took an average of 19 releases to fix the violation

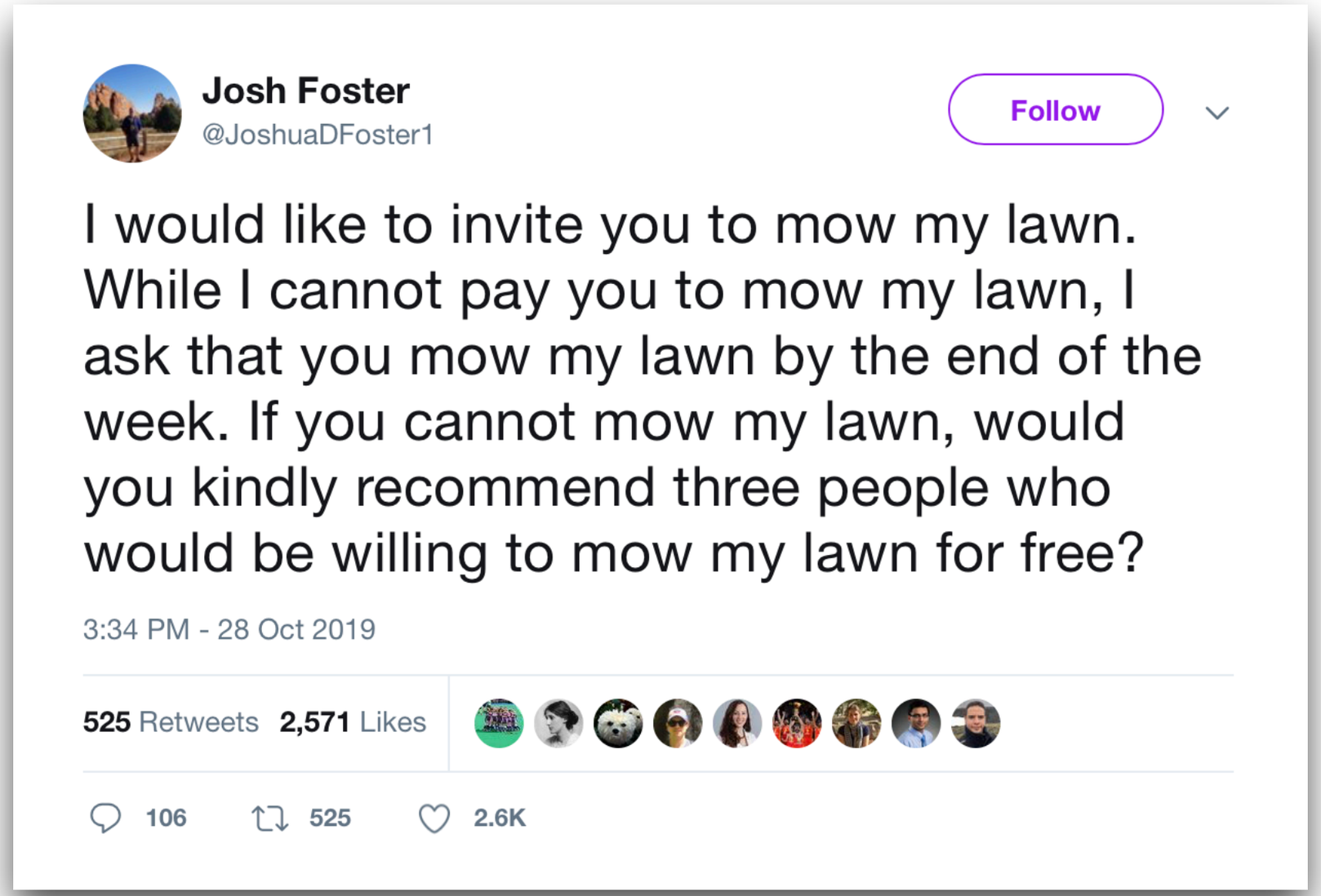
Open Access

Open Access Initiative

- If research is funded by the public, the output should be available for the general public without paywall

- European Union runs EU-wide research programme called Horizon 2020 (over 100 billion Euros). Starting from 2021, all scholarly publications on the results from research funded by public or private grants provided by national, regional and international research councils and funding bodies, must be published in Open Access Journals, on Open Access Platforms, or made immediately available through Open Access Repositories without embargo.

https://avandeursen.com/2019/08/20/europes-open-access-plan-s-and-paper-publishing-in-software-engineering-research/



https://twitter.com/JoshuaDFoster1/status/1188947163398787074

ACM Copyright and Audio/Video Release

Title of the Work: Embedding Genetic Improvement into Programming Languages

Author/Presenter(s): Shin Yoo:Korea Advanced Institute of Science and Technology
Type of material: Full Paper

Publication and/or Conference Name: GECCO '17: Genetic and Evolutionary Computation Conference Companion Proceedings

I. Copyright Transfer, Reserved Rights and Permitted Uses

* Your Copyright Transfer is conditional upon you agreeing to the terms set out below.

Copyright to the Work and to any supplemental files integral to the Work which are submitted with it for review and publication such as an extended proof, a PowerPoint outline, or appendices that may exceed a printed page limit, (including without limitation, the right to publish the Work in whole or in part in any and all forms of media, now or hereafter known) is hereby transferred to the ACM (for Government work, to the extent transferable) effective as of the date of this agreement, on the understanding that the Work has been accepted for publication by ACM.

Reserved Rights and Permitted Uses

(a) All rights and permissions the author has not granted to ACM are reserved to the Owner, including all other proprietary rights such as patent or trademark rights.

(b) Furthermore, notwithstanding the exclusive rights the Owner has granted to ACM, Owner shall have the right to do the following:

(i) Reuse any portion of the Work, without fee, in any future works written or edited by the Author, including books, lectures and presentations in any and all media.

(ii) Create a "Major Revision" which is wholly owned by the author

(iii) Post the Accepted Version of the Work on (1) the Author's home page, (2) the Owner's institutional repository, (3) any repository legally mandated by an agency funding the research on which the Work

UC and Elsevier: Overview

UPDATE AS OF JULY 10, 2019

The University of California has been out of contract with Elsevier since January. UC ended negotiations for a new journal subscription contract on February 28, with the support of the systemwide Academic Senate. The publisher continued to allow access to new articles via ScienceDirect for several months; however, **UC's direct access to 2019 Elsevier articles (and older articles in some journals) is now being discontinued.** (Note: The process for discontinuing access is complex, so access may vary by journal and/or campus until Elsevier's rollout of the changes is complete.)

Over the coming months, the UC Libraries will be carefully evaluating the impact of losing access to new articles on ScienceDirect, and will do our best to ensure that members of the UC community have access to the articles they need. Learn about other ways to access Elsevier articles.

See also the additional August 2, 2019 statement by the negotiation team, Fact check: What you may have heard about the dispute between UC and Elsevier.

https://osc.universityofcalifornia.edu/open-access-at-uc/publisher-negotiations/uc-and-elsevier/

Norway joins the ranks of Germany and Sweden, cancels subscription with Elsevier

https://doi.org/10.34193/EI-A-11541

Fatima Qureshi
Junior Content Writer and Editor, Editage Insights
Mar 19, 2019 · 3.4k views

Norway canceled its subscription with Elsevier after the negotiations between the publisher and Norway's negotiating consortium regarding access to research papers failed to reach an agreement. Consequently, Norwegian researchers will lose access to the new articles published in Elsevier journals, although it's not clear when the access will be halted. Germany, Sweden, Hungary, and the University of California system have

Sci-Hub

From Wikipedia, the free encyclopedia

Sci-Hub is a website that provides free access to millions of research papers and books, without regard to copyright,[2] by bypassing publishers' paywalls in various ways.[2][3][4]

Sci-Hub was founded by Alexandra Elbakyan in 2011 in Kazakhstan in response to the high cost of research papers behind paywalls. The site is widely used in both developed and developing countries.[2][3] As of October 2019, it claimed to contain 76 million academic articles and serve approximately 400,000 requests per day.[5]

Sci-Hub and Elbakyan were sued twice for copyright infringement in the United States in 2015 and 2017, and lost both cases, leading to loss of some of its Internet domain names.[6] The site has cycled through different domain names since then.[2]

Sci-Hub has been lauded by some[7] in the scientific, academic, and publishing communities[8] for providing access to knowledge generated by the scientific community. Others have criticized it for violating copyright,[3][9] threatening the economic viability of publishers,[10] potentially compromising universities' network security and jeopardizing legitimate access to papers by university staff.[11]

Contents [hide]

History [edit]

Sci-Hub

Type of site	File sharing
Available in	English Russian[a]
Created by	Alexandra Elbakyan
Website	sci-hub.tw/about sci-hub.se/about sci-hub.ltd/about
Commercial	No
Registration	Not required
Launched	16 April 2011; 8 years ago[1]
Current status	Active
Content license	Hosts material without regard to copyright

Part of a series on

File sharing

Technologies

File hosting services · Online video platform · Peer to peer · Usenet · Web hosting · WebRTC · XDCC

Concluding Thoughts

- What was the license you applied to your last open source project?
- Do you agree with GPL and free software movement?
- Would you support Sci-Hub as a researcher?