FUJITSU

shaping tomorrow with you

About License Scanner

2018/9/28

安倍 昌輝<ambai.masaki@jp.fujitsu.com>
株式会社富士通コンピュータテクノロジーズ

1509ka1

Copyright 2018 FUJITSU COMPUTER TECHNOLOGIES

自己紹介

FUJITSU

- 安倍　昌輝　(あんばい　まさき)
  - “あべ”じゃないです。
- 最近のちょっとした自慢
  - サッカーの大会に出場して優勝したこと
- 業務は組込み向けLinuxを開発
  - チームの一員としてYoctoプロジェクトに貢献
- 最近ではOSS コンプライアンスに関わることが多く、ライセンス周りについて勉強中

The names of products are the product names, trademarks or registered trademarks of the respective companies.
Trademark notices ((R),TM) are not necessarily displayed on system names and product names in this material.

Copyright 2018 FUJITSU COMPUTER TECHNOLOGIES

富士通グループでの活動

FUJITSU

- 富士通グループではOSSライセンス遵守活動に長年に渡って取り組む
- 培ってきたノウハウを活かすためにOpenChainに参加している
  - OpenChainについて
    - 2013年にQualcomm主導で作られたLinux Foundation傘下のプロジェクト
    - OSSコンプライアンスをソフトウェア サプライチェーンにまで広め、標準化していく
    - https://www.openchainproject.org/
  - OpenChain Japan Work Group
    - 2018/8/31　第5回会合　2018年8月31日（金）14:30-17:30
      場所：富士通川崎工場
      https://wiki.linuxfoundation.org/openchain/openchain-japanese-working-group

Copyright 2018 FUJITSU COMPUTER TECHNOLOGIES

背景

FUJITSU

- Yoctoプロジェクトのmeta-spdxscannerをより良くする事に取り組んでいます

- 最初はライセンスをスキャンするためにFOSSologyを使用していたが
  - 対応しているSPDXの仕様版数が古い
  - スキャンが遅い
- DoSOCSv2に変更したが、出力精度がイマイチ

- 世の中にあるライセンス　スキャンを調査して
  何がmeta-spdxscannerに適しているか評価する

Copyright 2018 FUJITSU COMPUTER TECHNOLOGIES

評価基準

- CUIやAPIなどを使用してライセンスのスキャンができること
  - Yoctoから使用するため
  - CI/CD環境で定期的にスキャンするため
- SPDXに対応していること
- スキャン対象のライセンスを検出できること
- また、コピーライトを検出できること

Copyright 2018 FUJITSU COMPUTER TECHNOLOGIES

OSSライセンス コンプライアンス ソフトウェアの紹介

FUJITSU

- オープン ソース ソフトウェア ツール
  - FOSSology
  - DoSOCSv2
  - LiD
  - scancode-toolkit
- 商用ソフトウェア ツール
  - Black Duck Protex
  - FlexNet Code Insight
  - FOSS ID
  - WhiteSource
  - Protecode SC
  - Clarity

Copyright 2018 FUJITSU COMPUTER TECHNOLOGIES

オープン ソース ソフトウェア ツールの紹介

FUJITSU

基本情報の比較

Item	FOSSology	DoSOCSv2	LiD	scancode-toolkit
Last Release	3.3.0 (2018/04/17)	0.16.1 (2016/02/17)	1.4.1 (2018/06/21)	2.9.2 (2018/05/19)
License	GPLv2	GPLv2	BSD-3-Clause	Apache-2.0
Support SPDX version	2.1	2.0	-	2.1
SPDX License version	2.6	2.6	3.1	3.1
Scanners	Nomos, Monk, Ninka	Nomos	LiD	scancode
Supported Platform	Linux/Win/Mac	Linux	Linux	Linux/Win/Mac
Graphical user interface	✓	✓	-	-
Command Line	-	✓	✓	✓
Project Activity (http://www.openhub.net)	High	Low	Activity Not Available	Very Low

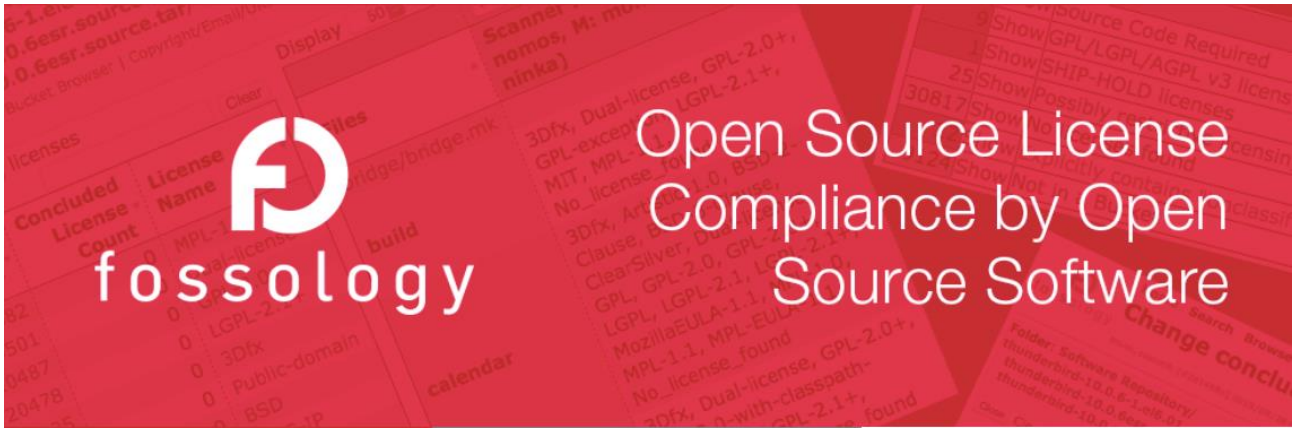
Copyright 2018 FUJITSU COMPUTER TECHNOLOGIES

FOSSology

FUJITSU

- 特徴
  - ライセンスはGPLv2
  - The Linux FoundationのCollaborativeプロジェクト
  - Web経由のFOSSologyサーバでライセンス、著作権のスキャンを行う
  - CUIやAPIからは使用不可
  - ただし、Fossdriver (https://gitlab.com/swinslow/fossdriver)を使用してPythonコードから実行可能

- https://www.fossology.org/



Copyright 2018 FUJITSU COMPUTER TECHNOLOGIES

FOSSologyのスキャン結果

FUJITSU

- FOSSology 3.3.0を使用
  - スキャン対象はbzip2 v1.0.6

SPDXVersion: SPDX-2.1
DataLicense: CC0-1.0

FileName: bzip2-1.0.6/LICENSE
SPDXID: SPDXRef-item1699540
FileChecksum: SHA1: 1c0c6888759a63c3
FileChecksum: MD5: ddeb76cd34e791893c
LicenseConcluded: LicenseRef-bzip2-1.0.6
LicenseInfoInFile: LicenseRef-bzip2-1.0.6
FileCopyrightText: <text> copyright (C) 1996-2010 Julian R Seward. All rights reserved.
copyright notice, this list of conditions and the following disclaimer. </text>
：（省略）
LicenseID: LicenseRef-bzip2-1.0.6
LicenseName: bzip2 and libbzip2 License v1.0.6
ExtractedText: <text> This program, "bzip2", the associated library "libbzip2", and all documentation, are copyright (C) 1996-2010 Julian R Seward. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:
：（省略）

ライセンスを正しく検出

コピーライトも検出

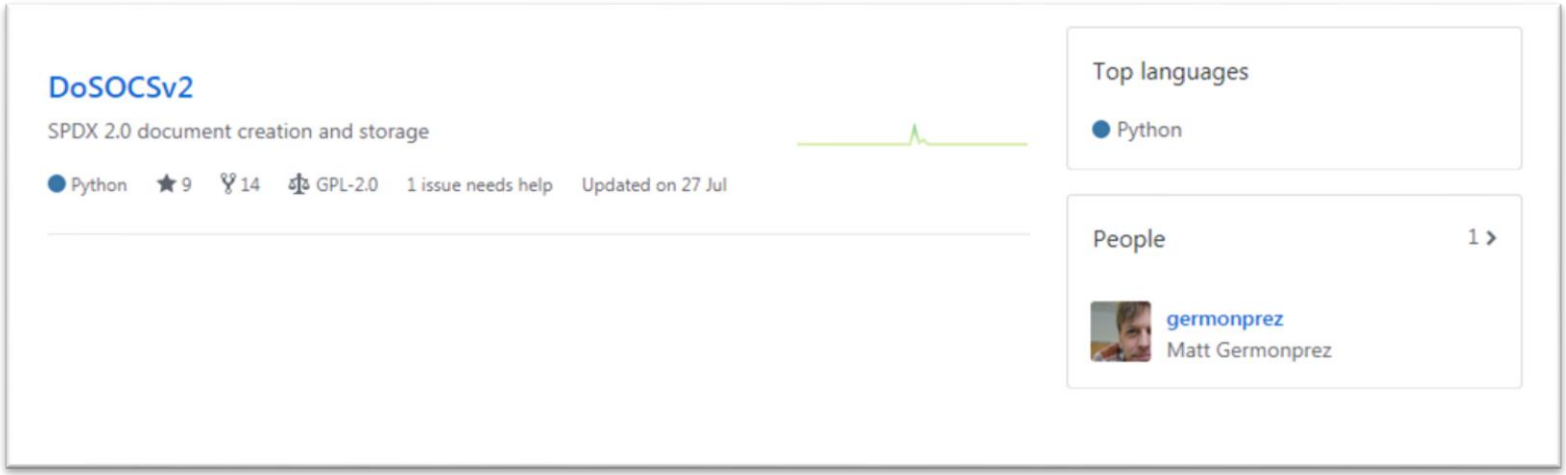
ライセンス文も検出

Copyright 2018 FUJITSU COMPUTER TECHNOLOGIES

DoSOCSv2

- 特徴
  - ライセンスはGPLv2
  - スキャン時間が速い (2回目以降)
  - 環境構築が容易
  - CUIから使用可能
  - dosocs2-ui(https://github.com/pombredanne/dosocs2-ui)を使用する事でGUIからもスキャン可能

- https://github.com/DoSOCSv2/DoSOCSv2



Copyright 2018 FUJITSU COMPUTER TECHNOLOGIES

DoSOCSv2のスキャン結果

FUJITSU

- DoSOCSv2 0.16.1を使用
  - スキャン対象はbzip2 v1.0.6

SPDXVersion: SPDX-2.0
DataLicense: CC0-1.0

FileName: ./LICENSE
SPDXID: SPDXRef-file-LICENSE-4919-7310aaf0
FileType: OTHER
FileChecksum: SHA256:
4919cfb14a73cd64fcef67b10761397[illegible]f373bc7204
LicenseConcluded: NOASSERTION
LicenseInfoInFile: LicenseRef-BSD-style
LicenseComments: <text></text>
FileCopyrightText: NOASSERTION
FileComment: <text></text>
FileNotice: <text></text>
：（省略）

間違いでないが正確でもない

BSD-styleとして検出

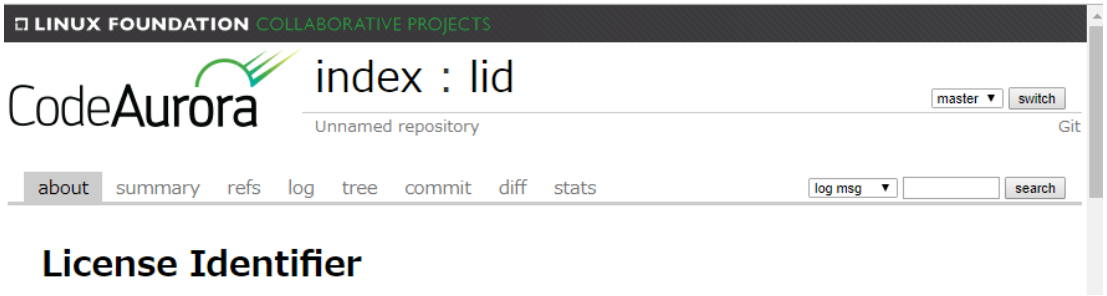
コピーライトを検出しない

Copyright 2018 FUJITSU COMPUTER TECHNOLOGIES

LiD

- 特徴
  - ライセンスはBSD-3-Clause
  - LiDはQualcomm OSTG (Open Source Technology Group)のLicense Identifier(LiD) ツール
  - CUIから使用可能
  - 出力形式はSPDXに対応していない。‘easy_read’と‘csv’の2種類に対応
    - ‘easy_read’はヒューマン リーダブル
    - ‘csv’はマシン リーダブル

- https://source.codeaurora.org/external/qostg/lid/



Copyright 2018 FUJITSU COMPUTER TECHNOLOGIES

LiDのスキャン結果

FUJITSU

- LiD 1.4.1を使用
  - スキャン対象はbzip2 v1.0.6
- ‘easy_read’ 形式

1ファイルずつならスキャンできるが
まとめてだとスキャンできない

```
=== Found 1 results for '/home/scaner/bzip2-1.0.6/LICENSE':
Summary of the analysis

Name of the input file: /home/scaner/bzip2-1.0.6/LICENSE
Matched license type is bzip2-1.0.6
Score for the match is 1.0
Rank for the match is 5
License text beings at line 2.
License text ends at line 41.
Start byte offset for the license text is 76.
End byte offset for the license text is 1826.
The found license text has the score of 1.0
The following text is found to be license text
-----BEGIN-----

This program, "bzip2", the associated library "libbzip2", and all
documentation, are copyright (C) 1996-2010 Julian R Seward.  All
rights reserved.
 : (省略)
Julian Seward, jseward@bzip.org
bzip2/libbzip2 version 1.0.6 of 6 September 2010

-----END-----
```

正しくライセンスを検出

ライセンス文も検出

コピーライトを検出しない

Copyright 2018 FUJITSU COMPUTER TECHNOLOGIES

LiDのスキャン結果

- LiD 1.4.1を使用
  - スキャン対象はbzip2 v1.0.6
- ‘csv’ 形式

どれだけ一致しているかのスコア

ライセンスを判別

ライセンス文を検出

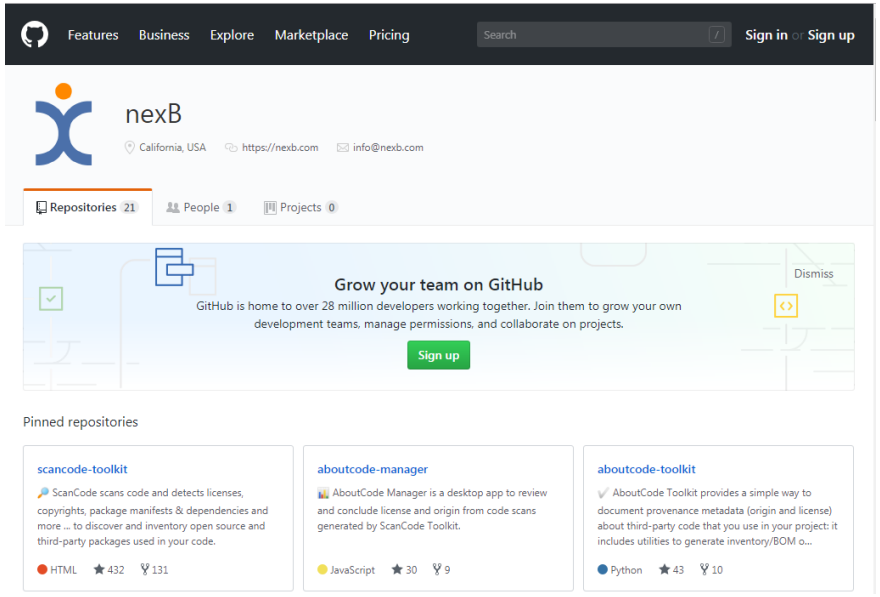
全61ファイル中29ファイルしか解析できなかった

input file path	matched license type	Score usin	Rank based	Start line n	End line nu	Start byte	End byte o	Score usin	Found license text
../scaner/bzip2-1.0.6/LICENSE	bzip2-1.0.6	1	5	2	41	76	1826	1	
../scaner/bzip2-1.0.6/Makefile	GFDL-1.2-only	0.068599	1	5	12	231	467	0.069833	# Copyright (C) 1996-2010 Julian Seward <jseward@bzip.org>
../scaner/bzip2-1.0.6/Makefile-libbz	GFDL-1.2-only	0.071963	1	13	20	525	761	0.069833	# Copyright (C) 1996-2010 Julian Seward <jseward@bzip.org>
../scaner/bzip2-1.0.6/README	bzip2-1.0.5	0.090356	2	118	144	4449	5533	0.085003	
../scaner/bzip2-1.0.6/README	GFDL-1.1-only	0.043053	ScoreOutO	9	15	330	547	0.063413	Copyright (C) 1996-2010 Julian Seward <jseward@bzip.org>
../scaner/bzip2-1.0.6/README.CON	GFDL-1.2	0.055607	ScoreOutO	5	12	222	446	0.062676	Copyright (C) 1996-2010 Julian Seward <jseward@bzip.org>
../scaner/bzip2-1.0.6/README.XML	GFDL-1.2-only	0.06288	1	5	12	228	462	0.069833	Copyright (C) 1996-2010 Julian Seward <jseward@bzip.org>
../scaner/bzip2-1.0.6/blocksort.c	GFDL-1.2-only	0.040499	ScoreOutO	11	18	500	739	0.069833	Copyright (C) 1996-2010 Julian Seward <jseward@bzip.org>
../scaner/bzip2-1.0.6/bzip2.1	LPPL-1.0	0.040279	ScoreOutO	334	386	11257	13453	0.014707	-8 6800K 3300K 2100K 828642
../scaner/bzip2-1.0.6/bzip2.1	LPPL-1.0	0.056963	ScoreOutO	81	122	2006	3276	0.010823	.1 bzip2
../scaner/bzip2-1.0.6/bzip2.1	LPPL-1.0	0.050599	ScoreOutO	138	164	3585	4554	0.007557	.1 BZIP2
../scaner/bzip2-1.0.6/bzip2.1	LPPL-1.0	0.046179	ScoreOutO	240	264	6957	7994	0.004969	information which is primarily of interest for diagnostic purposes.
../scaner/bzip2-1.0.6/bzip2.1.prefor	LPPL-1.0	0.057309	ScoreOutO	45	85	2231	4301	0.014123	will decline to write compressed output to a terminal, as
../scaner/bzip2-1.0.6/bzip2.1.prefor	LPPL-1.0	0.050019	ScoreOutO	175	207	8822	10500	0.00514	verbose mode 要約：show the compression ratio for
../scaner/bzip2-1.0.6/bzip2.1.prefor	LPPL-1.0	0.040169	ScoreOutO	349	359	18087	18628	0.003523	
../scaner/bzip2-1.0.6/bzip2.c	GPL-3.0	0.140504	3	1607	1617	45094	45643	0.257939	Copyright (C) 1996-2010 by Julian Seward.\n
../scaner/bzip2-1.0.6/bzip2.txt	LPPL-1.0	0.056898	ScoreOutO	42	82	1831	3761	0.014123	will decline to write compressed output to a terminal, as
../scaner/bzip2-1.0.6/bzip2.txt	LPPL-1.0	0.049701	ScoreOutO	175	204	8108	9420	0.004898	is primarily of interest for diagnostic purposes.
../scaner/bzip2-1.0.6/bzip2.txt	LPPL-1.0	0.040058	ScoreOutO	346	356	16685	17212	0.003523	

Copyright 2018 FUJITSU COMPUTER TECHNOLOGIES

scancode-toolkit

- 特徴
  - ライセンスはApache-2.0
  - nexB社が公開しており、これ以外にもスキャン結果を管理するためのツールなど様々なツールを公開している
  - CUIから使用可能
  - 出力形式はSPDX以外にもJSON, CSV, HTMLなどが選択できる
- https://github.com/nexB/scancode-toolkit



Copyright 2018 FUJITSU COMPUTER TECHNOLOGIES

scancode-toolkitのスキャン結果

FUJITSU

- scancode-toolkit 2.9.2を使用
  - スキャン対象はbzip2 v1.0.6

必須項目がNOASSERTION

```
# File

FileName: ./bzip2-1.0.6/LICENSE
FileChecksum: SHA1: 1c0c6888759a63c32bca7eb63353af2cd9bd5
LicenseConcluded: NOASSERTION
LicenseInfoInFile: bzip2-1.0.6
FileCopyrightText: <text>copyright (c) 1996-2010 Julian R Seward.
</text>
```

正しくライセンスを検出

コピーライトを検出

SPDXの仕様では
FileName →必須
FileChecksum →必須
LicenseConcluded →必須
LicenseInfoInFile →必須
FileCopyrightText →必須

LicenseConcluded
→このSPDXを作成した人がこのファイルのライセンスはXXXだと決定したライセンス
LicenseInfoInFile
→このファイルの中で見つけたライセンス

Copyright 2018 FUJITSU COMPUTER TECHNOLOGIES

商用ソフトウェア ツールの紹介

FUJITSU

ツールの一覧

Item	Vender	Target
Protex	Black Duck by Synopsys	Source code
FlexNet Code Insight	Flexera Software	Source code Binary code
FOSS ID	FOSS ID	Source code
WhiteSource	WhiteSource Inc.	Source code
Protecode SC	Synopsys	Binary code
Clarity	Insignary Inc.	Binary code

特徴

- ソース/バイナリ コードを解析し、OSSを特定することでライセンスを識別
  - OSSやコード スニペットのハッシュ値をDBと比較して特定する
- 各ツールのデータベースが非常に重要 (DBの情報が少ないと上手く特定できない)
- 機能は多岐に渡り、OSSの管理、ライセンスの管理、脆弱性情報の管理など

Copyright 2018 FUJITSU COMPUTER TECHNOLOGIES

商用ソフトウェア ツールの紹介

FUJITSU

- Protex：Black Duck by Synopsys
  - https://www.blackducksoftware.com/products/protex
- FlexNet Code Insight：Flexera Software
  - https://www.flexera.jp/enterprise/products/software-vulnerability-management/flexnet-code-insight/
- FOSS ID
  - http://fossid.com/
- WhiteSource
  - https://jp.whitesourcesoftware.com/

Copyright 2018 FUJITSU COMPUTER TECHNOLOGIES

商用ソフトウェア ツールの紹介

FUJITSU

- Protecode SC：Synopsys
  - https://www.synopsys.com/apps/japan/today-tomorrow/articles/tt110-protecode-sc.html
- Clarity：Insignary Inc.
  - https://www.insignary.com/

Copyright 2018 FUJITSU COMPUTER TECHNOLOGIES

まとめ

FUJITSU

- License Scannerとして
  - 総合的にFOSSologyが良いと考える
    - SPDXの出力でライセンスやコピーライトを示すLicenseConcluded、FileCopyrightTextの検出が良い
    - Linux Foundation傘下のプロジェクトである
    - コミュニティが健全であり、とても活発である
    - FOSSology単体ではCUIが使えないがFossdriver がある
  - 商用ソフトウェア ツールはOSS、ライセンスを特定することは出来るがコピーライトを検出できない
    - 使用しているOSSを管理したり、ライセンス違反がないか確認したり、脆弱性の対応状況を管理するなどシステム全体のOSSを管理する事ができる
    - 目的や利用シーンに応じて、使用するツールを使い分ける事でより効果的である
    - 例)製品の出荷前にバイナリを解析する事でOSSやライセンスの最終確認をする
      Protecode SCやClarityなど
    - 例)コンプライアンスのリスクだけではなく、脆弱性のリスクにも対応したい
      FlexNet Code InsightやWhiteSourceなど

Copyright 2018 FUJITSU COMPUTER TECHNOLOGIES

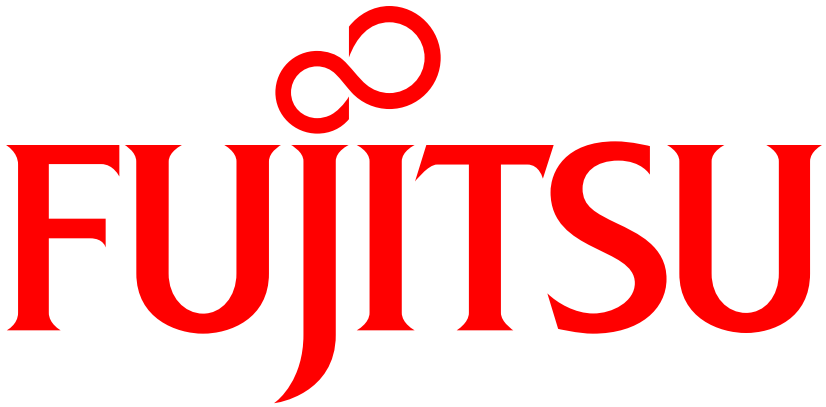
次回

FUJITSU

- 今回は広く浅く、ライセンスのスキャン ツールを紹介した

- 次回は狭く深くで
- FOSSologyで出力したSPDXはSPDX 2.1の仕様にどれだけ準拠しているか
- などを紹介できればと思っています。

Copyright 2018 FUJITSU COMPUTER TECHNOLOGIES



shaping tomorrow with you