

ライセンスルール更新にむけた 未知ライセンス文調査

松浦 遊也

全体の流れ

1. OSSの再利用
2. 主流のライセンス特定ツール(正規表現)
3. ライセンス判定例1,2
4. 未知ライセンス記述の問題
5. 未知ライセンス記述削減方法
6. ライセンスルール更新
7. 研究目的: 未知ライセンス文の傾向を把握する
8. アプローチ
9. データセット
10. 結果
11. 考察

OSSの再利用

- 生産性向上のために企業はOSSを再利用することが多くなってきている
- OSSを再利用するためには各ソースファイルにコメントとして記載されているライセンスを遵守する必要がある
- ソースファイルは膨大なので、目視する時間を短縮するために**ライセンス特定ツールが存在している**

主流のライセンス特定ツール

- ライセンス記述を抽出し、**ライセンスルール(正規表現)**と**ライセンス文の比較**を用いてライセンスを特定する
- 特定できない場合、誤判定を防ぐために未知のライセンスと判定する

ソースファイル

```
/*  
*Redistribution ~  
*1 Redistributions of ~  
*2 Redistributions in ~  
*/  
  
#include  
~~~~~
```

ライセンス記述

```
/*  
*Redistribution ~  
*1 Redistributions of ~  
*2 Redistributions in ~  
*/
```

ライセンス文

Redistribution ~

1 Redistributions of ~

2 Redistributions in ~

ライセンス判定:例1

ライセンス記述

ライセンス文1

Redistribution and use~:

ライセンス文2

1. Redistributions of
source code must ~.

ライセンス文3

2. Redistributions in
binary form must ~.

ライセンス文1は
ライセンスルール:BSDpre
にマッチ

ライセンス文2は
ライセンスルール:BSDcondSource
にマッチ

ライセンス文3は
ライセンスルール:BSDcondBinary
にマッチ

読み込んだライセンス記述全体の構成は

BSDpre, BSDcondSource, BSDcondBinary

ライセンス記述: BSDv2とマッチ

出力結果: **BSDv2**

ライセンス判定:例2

ライセンス記述

ライセンス文1

Use and redistribution～:

ライセンス文2

1. Redistributions of
source code must ～.

ライセンス文3

2. Redistributions in
binary form must ～.

ライセンス文1は
対応しているライセンスルールは
存在しない: **未知ライセンス文**

ライセンス文2は
ライセンスルール:BSDcondSource
にマッチ

ライセンス文3は
ライセンスルール:BSDcondBinary
にマッチ

読み込んだライセンス記述全体の構成は
未知ライセンス文,BSDcondSource,BSDcondBinary
マッチしているライセンス
出力結果: **未知ライセンス記述**

未知ライセンス記述の問題

- 未知ライセンス記述の制約内容を理解するには目視をするしかない
- ライセンス特定ツールの利用者に時間的な負担を与えており、少しでも未知ライセンス記述を削減することが望まれている

未知ライセンス記述削減方法

- 未知ライセンス記述を削減する方法としてライセンスルールの更新がある
- 未知ライセンス記述になるかどうかは読み込んだライセンス記述がライセンスルールと対応しているかで判定される
- ライセンスルールの追加や修正を行い、ライセンスに対応できる範囲を広げれば、未知ライセンスを削減することができる

ライセンスルール更新

- ライセンスルールを更新するには現在の未知ライセンス文になる傾向を把握し、どのような更新が必要なのか把握する必要がある
- 未知ライセンスの傾向を把握する研究は存在しておらず、**効率のよい更新方法が分かっていない**

研究目的

- 未知ライセンス文の傾向を把握し、効率のよいライセンスルール更新方法を調査する

アプローチ

- 未知ライセンス文をライセンスルールの更新に基づいて4種類のタイプに分類分けする
- 分類タイプ
 - ライセンスルール修正型(既知ライセンス)
 - ライセンスルール追加型(既知ライセンス)
 - ライセンスルール修正型(未知ライセンス)
 - ライセンスルール追加型(未知ライセンス)

ライセンスルール修正型 (既知ライセンス)

- 実際のライセンス文(Debian)
- This distribution is free software; you **may** redistribute it and/or modify it under the same terms as Perl itself.
- Ninkaのライセンスルール
- `^([^,;]+)` is free software; you **can** redistribute it and/or modify it under the same terms as Perl itself\$

ライセンスルール追加型 (既知ライセンス)

- 実際のライセンス文(Debian)
- Redistribution of this file is permitted under the GNU Public License.
- Ninkaのライセンスルール
- $^([\wedge,;]+)$ can be freely redistributed under the terms of the GPL\$

ライセンスルール修正型 (未知ライセンス)

- 実際のライセンス文(Debian)
- This module is free software; you can redistribute it **or** modify it under the same terms as Perl itself.
- Ninkaのライセンスルール
- `^([^,;]+)` is free software; you can redistribute it **and/or** modify it under the same terms as Perl itself\$

ライセンスルール追加型 (未知ライセンス)

- 実際のライセンス文(Debian)
- This software is subject to the provisions of the Zope Public License, Version 2.1 (ZPL).
- Ninkaのライセンスルール
- ^This Source Code Form is subject to the terms of the Mozilla Public License, v<dot>2<dot>\$0:

データセット

- Debian-7.8.0のデータセットを用いる