

# ライセンス特定のための ライセンス知識獲得における課題

真鍋雄貴<sup>†1</sup>

ソースファイルからそのソースファイルが対応するライセンスを特定する様々な手法が開発されてきた。それらの手法は、主に、ライセンス知識とライセンス知識とソースファイルをマッチングさせる手法から成り立っている。本稿では、ライセンス知識の効率の良い構築をめざし、その際の課題となる点について述べる。

## Challenge to Build License Knowledgebase for license identification

YUKI MANABE <sup>†1</sup>

Several license identification tools which identify a license corresponding to a source file are developed. These approach consists of license knowledge and matching method between source files and license knowledge. This paper describe challenges in license knowledge construction.

### 1. はじめに

ソフトウェア開発コストの削減のため、オープンソースソフトウェアを再利用し、新しいソフトウェアの開発に利用することがある。各ソフトウェア部品を再利用する際には、そのソフトウェア部品のライセンスを特定し、開発中のソフトウェアで想定しているライセンスと不整合が生じていないかを確認する必要がある。

このようなソースファイルのライセンスを特定するツールとして、ライセンス特定ツール

がある。ライセンス特定ツールの概要を図 1 に示す。ライセンス特定ツールは主に、既知のライセンスから作成されるライセンス知識と、ソースファイルとライセンス知識とのマッチングを行う手法から構成されている。ライセンス知識にはライセンスの名前と、それに対応するライセンス記述もしくはそのライセンス記述に対応するパターンの組（ライセンスルール、以下、ルール）が格納されている。近年、ライセンス特定ツールを用いて、オープンソースソフトウェアとライセンスとの関係を対象とした実証的研究がいくつか行われており<sup>1),2),5)</sup>。このため、ライセンス特定ツールはオープンソースソフトウェア工学に貢献できるのではないかと考える。

ライセンス特定をより円滑に行うためには、高い精度でライセンスを特定できる必要がある。一方で、ライセンス知識を構築するコストは低くない。そのため、効率よくライセンス知識を構築できる手法が必要である。本稿では、ライセンス知識を構築する際に問題となる点について述べる。まず、既存のライセンス特定手法を整理し、十分なライセンス知識を構築する意義を述べる。次に、ライセンス知識を自動的に構築するために扱う必要があると考える課題を述べる。

### 2. 既存のライセンス特定手法

本節では、既存のライセンス特定手法をルールを表現するパターンの違いから整理する。ルールを表現するパターンとして正規表現を用いるライセンス特定ツールとして、Ninka<sup>3)</sup>とALSA<sup>6)</sup>、Ohcount<sup>\*1</sup>がある。Ninkaとは正規表現で記述されたメタライセンス文と、メタライセンス文の列で表現されたライセンスルールから成るライセンス知識により、ライセンスを特定するツールである。ASLAは各ライセンス記述に対応した正規表現を用いて、ライセンスを特定する。Ohcountも同様に正規表現を用いるが、ASLAと違い、すべて正規表現はハードコーディングされている。

ルールを表現するパターンとして単なる文字列を用いるライセンス特定ツールとして、FOSSology<sup>4)</sup>やOSLC<sup>\*2</sup>がある。FOSSologyはbSAMアルゴリズム<sup>\*3</sup>を用いて既知のライセンス記述とソースコード中のコメントを比較する。コメント中で類似したライセンス記述を持つライセンスを特定結果として出力する。OSLCはソースコード中のコメントとライセンス記述間で同一となる行を発見し、最長一致列となる部分を探索する。類似度はライ

<sup>†1</sup> 大阪大学  
Osaka University

\*1 <http://www.ohloh.net/p/ohcount>

\*2 <http://sourceforge.net/projects/oslc/>

\*3 [http://fossology.org/symbolic\\_alignment\\_matrix](http://fossology.org/symbolic_alignment_matrix)

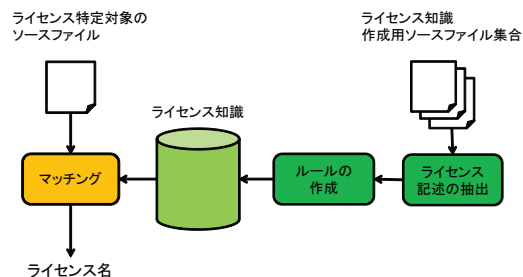


図1 ライセンス特定手法の概要

センス記述に対する最長一致列の割合である。類似度の高いライセンス記述に対応するライセンスを特定結果として出力する。

これらの手法は以下のようにまとめられる。前者は、ライセンス知識の作成を最大限行う必要があるが、マッチングでコストの低い手法を使うことができる。後者は、ライセンス知識の作成はやや容易であるが、ライセンス知識とソースファイル間のマッチングでコストの高いアルゴリズムを使っているといえる。

本稿では、前者の立場を支持する。なぜならば、ライセンス知識を構築することは、あるライセンス記述がどのライセンスに属するか、その領域を決定する行為であるため、ライセンス知識を十分に構築することで、その領域を明確にできると考えられるからである。また、マッチング手法でライセンス知識の不足を補おうとすると、より一層、各ライセンスの領域が不明確になってしまう。

### 3. ライセンス知識作成における課題

始めに、ソースファイル中にあるコメントにおけるライセンス記述の位置が先頭でない場合があることである。多くの場合、ファイルの先頭に存在する。しかし、先頭でない場合もまれにある。また、複数の場所に記載されていることがある。

次に、記述としては近いが、異なるライセンスに対応しているライセンス記述を考慮してルールを作成する必要があることである。ライセンス間で非常に近いライセンス記述が存在する（GNU General Public license version 2 と GNU General Public license version 2 or later, GNU General Public license version 2 と GNU General Public license version 3 など）ことがあるため、ルールが必要以上に広い範囲をカバーしてしまっていた場合、精

度が落ちる可能性がある。

最後に、各ルールに対する名前付けが容易でない場合がある。ライセンス記述にはそのライセンス名が記載されていない場合がある。ライセンス名がない場合は、ウェブページ等からライセンス記述の一致するライセンスを探し、ライセンス名をつけることはできるが、ない場合も多く存在する。

## 4. むすび

本稿では、ライセンス特定におけるライセンス知識を構築する際に課題となる点として、ライセンス記述の位置、類似しているが異なるライセンスに対応するライセンス記述、ルールの名前付けを挙げた。これらの課題に取り組むことにより、オープンソースソフトウェア工学の確立に貢献していきたいと考えている。

**謝辞** 本研究は、日本学術振興会科学研究費補助金基盤研究（A）（課題番号:21240002）の助成を得た。ここに記して謝意を表す。

## 参考文献

- 1) DiPenta, M., German, D.M., Guéhéneuc, Y.-G. and Antoniol, G.: An exploratory study of the evolution of software licensing, *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering - Volume 1, ICSE '10*, pp.145–154 (2010).
- 2) German, D.M., DiPenta, M. and Davies, J.: Understanding and Auditing the Licensing of Open Source Software Distributions, *Proc. ICPC 2010*, Braga, Portugal, pp.84–93 (2010).
- 3) German, D.M., Manabe, Y. and Inoue, K.: A sentence-matching method for automatic license identification of source code files, *Proc. ASE 2010*, pp.437–446 (2010).
- 4) Gobeille, R.: The FOSSology project, *Proc. MSR 2008*, New York, NY, USA, pp. 47–50 (2008).
- 5) Manabe, Y., Hayase, Y. and Inoue, K.: Evolutional analysis of licenses in FOSS, *Proc. IWPSE-EVOL '10*, pp.83–87 (2010).
- 6) Tuunanen, T., Koskinen, J. and Karkkainen, T.: Automated software license analysis, *Automated Software Engineering*, Vol.16, pp.455–490 (2009).