

Title	Detection and Visualization of Code Clones for Very-large Source Code Collections
Author(s)	リビエリ, シモネ
Citation	大阪大学, 2008, 博士論文
Version Type	
URL	<a href="https://hdl.handle.net/11094/49288">https://hdl.handle.net/11094/49288</a>
rights	
Note	著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、 <a href="https://www.library.osaka-u.ac.jp/thesis/#closed">〈a href="https://www.library.osaka-u.ac.jp/thesis/#closed"〉</a> 大阪大学の博士論文について <a href="https://www.library.osaka-u.ac.jp/thesis/#closed">〈/a〉</a> をご参照ください。

***Osaka University Knowledge Archive : OUKA***

<https://ir.library.osaka-u.ac.jp/>

Osaka University

氏名	リビエリ シモネ LIVIERI SIMONE
博士の専攻分野の名称	博士 (情報科学)
学位記番号	第 22147 号
学位授与年月日	平成 20 年 3 月 25 日
学位授与の要件	学位規則第 4 条第 1 項該当 情報科学研究科コンピュータサイエンス専攻
学位論文名	Detection and Visualization of Code Clones for Very-large Source Code Collections (大規模ソースコードを対象としたコードクローンの検出と可視化)
論文審査委員	(主査) 教授 井上 克郎  (副査) 教授 萩原 兼一 教授 楠本 真二

#### 論文内容の要旨

昨今のソフトウェアの利用拡大に伴い、ソフトウェアの品質と知的財産の保護が重要となっている。また、我々の生活にソフトウェアの存在が不可欠になるにつれ、ソフトウェアベンダや開発者に対して、これまでに無いほど迅速なソフトウェアの供給が求められるようになってきている。

ソフトウェアシステムの品質に最も強い影響を与える要因の 1 つに、ソースコードの複製がある。ソースコードの複製は、コードクローンと呼ばれる重複したコード片として、ソースコード中に現れる。コードクローンが出現する理由としては、コピーアンドペーストによるソースコードの再利用や定型処理などが挙げられる。

コンピュータサイエンスの分野において、コードクローンの検出法は、1990 年代から活発に研究されてきた。その結果、様々なコードクローン検出システムが開発され、それぞれの検出システムは異なった性質や性能、スケーラビリティを持つに至った。しかし、計算機の性能に限界があるため、極めて大規模なソフトウェアを対象としたコードクローン検出を行えるシステムは存在しなかった。

そこで本研究では、数ギガバイト規模のソースコード集合を対象としたコードクローン検出法の実現に取り組んだ。このようなコードクローン検出法は、知的財産の保護やソフトウェアの変遷の調査に有用であると考え、この目的のために、分散処理環境上でコードクローンの検出を行う方法を実現した。また、大規模なソフトウェア集合を対象としたコードクローン検出と可視化を行う枠組みを提案した。さらに、この枠組みを試作し、以下の 3 つの事例について評価を行い、有効性を確認した。

1. 大規模なオープンソースソフトウェアである FreeBSD Ports Collections に対するコードクローン解析。
2. 研究室内で開発したソフトウェアと FreeBSD Ports Collections 間のコードクローン検出。
3. Linux と BSD 系オペレーティングシステムのカーネルに対するソフトウェアの変遷の調査。

## 論文審査の結果の要旨

ソフトウェアシステムの規模や利用範囲の拡大に伴い、ソフトウェアの品質と知的財産の保護が重要となっている。また、我々の生活にソフトウェアの存在が不可欠になるにつれ、ソフトウェアベンダや開発者に対して、これまでに無いほど迅速なソフトウェアの供給が求められるようになってきている。

ソフトウェアシステムの品質に最も強い影響を与える要因の1つに、コードクローンがある。コードクローンとは、ソースプログラム中に現れる同型のプログラム断片で、コピーアンドペーストや定型処理などで作られ、プログラムの保守作業を困難なものにしている。

ソフトウェア工学において、コードクローンの検出法は、1990年代から活発に研究されてきた。その結果、様々なコードクローン検出システムが開発され、それぞれの検出システムは異なった性質や性能、スケーラビリティを持つに至った。しかし、計算機の性能に限界があるため、極めて大規模なソフトウェアを対象としたコードクローン検出を行えるシステムは存在しなかった。

本論文では、数ギガバイト規模のソースコード集合を対象としたコードクローン検出法の実現に放り組んだ。このような大規模なコードクローン検出法は、知的財産の保護やソフトウェアの変遷の調査に必須な技術である。この目的のために、分散処理環境上でコードクローンの検出および可視化を行う方法を提案し、それを実現した。

さらに、このシステムを利用して、大規模なオープンソースソフトウェアである **FreeBSD Ports Collections** に対するコードクローン分析、研究室内で開発したソフトウェアと **FreeBSD Ports Collections** 間のコードクローン分析、**Linux** と **BSD** 系オペレーティングシステムのカーネル間のコードクローンの変遷調査等への適用を行い、本手法の有効性を確認した。

上記のように本論文の内容は、ソフトウェア工学の研究に、大規模コードクローン分析という新たな分野を切り開く非常に有用なものである。よって、本論文は博士（情報科学）論文として十分な価値があるものと認める。