

平成22年度科学研究費補助金実績報告書（研究実績報告書）

1. 機関番号

1	4	6	0	3
---	---	---	---	---

 2. 研究機関名 奈良先端科学技術大学院大学
3. 研究種目名 研究活動スタート支援 4. 研究期間 平成22年度～平成23年度
5. 課題番号

2	2	8	0	0	0	4	0
---	---	---	---	---	---	---	---
6. 研究課題名 自然言語処理を応用したコードクローン検出手法
7. 研究代表者

研究者番号	研究代表者名	所属部局名	職名
00582545	ヨシダ 吉田 ノリヒロ 則裕	情報科学研究科	助教

8. 研究分担者(所属研究機関名については、研究代表者の所属研究機関と異なる場合のみ記入すること。)

研究者番号	研究分担者名	所属研究機関名・部局名	職名

9. 研究実績の概要

下欄には、当該年度に実施した研究の成果について、その具体的内容、意義、重要性等を、交付申請書に記載した「研究の目的」、「研究実施計画」に照らし、600字～800字で、できるだけ分かりやすく記述すること。また、国立情報学研究所でデータベース化するため、図、グラフ等は記載しないこと。

識別子名の類似性に基づくコードクローン検出ツールの実装を行った。1つ目に、“ソースコードに含まれる類義語の特定”について述べる。各ソースファイルに含まれる語（基本的には識別子名。ただし、複数の語からなる識別子名は各語に分割）を抽出し、類義語（類似した用いられ方をしている語）の特定を行った。このためには、自然言語の類義語辞書をそのまま用いるという方法が考えられるが、これには“自然言語における類義語は、プログラミング言語における類義語と異なる可能性がある”という問題点と“特定分野のソフトウェアの開発では、分野固有の類義語が存在すると考えられ、そのような類義語は既存の類義語辞書には掲載されていないと考えられる”という問題点が挙げられる。そこで、類義語辞書を用いず、自然文書中から類義語を自動的に抽出する手法を実装した。クラスタリング後、同一のクラスタに含まれた語が類義語として特定される。2つ目に、“コード片間の等価性判定”について述べる。コード片間の等価性を判定するため、構文木中の部分木に対して特徴ベクトルを付加する。特徴ベクトルは、部分木に含まれる語の分布を表し、類義語は同一の語として扱う。次に、特徴ベクトル間の距離に基づいて部分木のクラスタリングを行った。最後に、同一クラスタに属した部分木に対応するコード片を等価なコード片として提示する。本研究は、イギリスの公立大学であるUniversity College of LondonのMark Harman教授との情報交換を通して遂行した。

10. キーワード

(1) コードクローン	(2) 類義語	(3) クラスタリング
(4) 国際情報交換	(5) イギリス	(6)
(7)	(8)	(裏面に続く)

11.研究発表（平成22年度の研究成果）

〔雑誌論文〕 計（ 1 ）件 うち査読付論文 計（ 1 ）件

著 者 名	論 文 標 題				
政井 智雄	テンプレートメソッドの形成に基づく類似メソッドの集約支援				
雑 誌 名	査読の有無	巻	発 行 年	最初と最後の頁	
ソフトウェア工学の基礎	有	XVII	2 0 1 1 0	1 2 5 ~ 1 3 0	

著 者 名	論 文 標 題				
雑 誌 名	査読の有無	巻	発 行 年	最初と最後の頁	
			1 1 1		

著 者 名	論 文 標 題				
雑 誌 名	査読の有無	巻	発 行 年	最初と最後の頁	
			1 1 1		

〔学会発表〕 計（ 5 ）件 うち招待講演 計（ 1 ）件

発 表 者 名	発 表 標 題		
Norihiro Yoshida	Detection of Chained Clone and Its Application		
学 会 等 名	発表年月日	発 表 場 所	
The 9th CREST Open Workshop Code Provenance and Clone Detection	2 0 1 0 年 1 1 月 2 3 日	London, UK	

発 表 者 名	発 表 標 題		
吉田 則裕	ソースコードコーパスを利用したシームレスな再利用支援		
学 会 等 名	発表年月日	発 表 場 所	
電子情報通信学会 ソフトウェアサイエンス研究集会 2010年12月研究集会	2 0 1 0 年 1 2 月 2 5 日	伊香保温泉ホテル天坊（群馬県）	

発 表 者 名	発 表 標 題		
Yu Kashima	A Preliminary Study on Impact of Software Licenses on Copy-and-Paste Reuse		
学 会 等 名	発表年月日	発 表 場 所	
International Workshop on Empirical Software Engineering in Practice 2010	2 0 1 0 年 1 2 月 7 日	奈良先端科学技術大学院大学（奈良県）	

発 表 者 名	発 表 標 題		
齋藤 晃	コードの生存期間を考慮したコードクローンと欠陥修正の関係調査		
学 会 等 名	発表年月日	発 表 場 所	
電子情報通信学会 ソフトウェアサイエンス研究集会 2010年10月研究集会	2 0 1 0 年 1 0 月 1 4 日	岩手県立大学（岩手県）	

発 表 者 名	発 表 標 題		
鹿島 悠	ソフトウェアライセンスがコピーアンドペーストによる再利用に与える影響の調査		
学 会 等 名	発表年月日	発 表 場 所	
電子情報通信学会 ソフトウェアサイエンス研究集会 2010年10月研究集会	2 0 1 0 年 1 0 月 1 4 日	岩手県立大学（岩手県）	

〔図 書〕 計 (0) 件

著 者 名	出 版 社		
書 名		発 行 年	総ページ数

12. 研究成果による産業財産権の出願・取得状況

〔出 願〕 計 (0) 件

産業財産権の名称	発明者	権利者	産業財産権の種類、番号	出願年月日	国内・外国の別

〔取 得〕 計 (0) 件

産業財産権の名称	発明者	権利者	産業財産権の種類、番号	取得年月日	国内・外国の別

13. 備考

※ 研究者又は所属研究機関が作成した研究内容又は研究成果に関する w e b ページがある場合は、U R L を記載すること。