

# 코드 클론에서 잠재적 라이선스 위반에 관한 연구

이상훈  
POSTECH 황승원 교수님 연구실

2011년 1월 7일

# 코드 클론의 예

- java.net의 GlassFish 프로젝트의 XMLChar.java (r1.3)

```
/* ... CDDL, GPL, GPL v2 license fragment ... */
package org.apache.jasper.xmlparser;

import java.util.Arrays;

public class XMLChar {
    // constants
    private static final byte[] CHARS = new byte[1 << 16];
    ....

    // static initialization
    static {
        CHARS[9] = 35;
        CHARS[10] = 19;
    }

    // public static methods
    public static boolean isSupplemental(int c) {
        return (c >= 0x10000 && c <= 0x10FFFF);
    }
    public static int supplemental(char h, char l) {
        return (h - 0xD800) * 0x400 + (l - 0xDC00) + 0x10000;
    }
    ...
}
```

- Apache의 Sling 프로젝트의 XMLChar.java (r599941)

```
/* ... Apache v2.0 license fragment ... */
package org.apache.sling.scripting.jsp.jasper.xmlparser;

import java.util.Arrays;

public class XMLChar {
    // constants
    private static final byte[] CHARS = new byte[65536];
    ....

    // static initialization
    static {
        CHARS[9] = 35;
        CHARS[10] = 19;
        ...
    }

    // public static methods
    public static boolean isSupplemental(int c) {
        return (c >= 0x10000 && c <= 0x10FFFF);
    }
    public static int supplemental(char h, char l) {
        return (h - 0xD800) * 0x400 + (l - 0xDC00) + 0x10000;
    }
    ...
}
```

# 코드 클론의 예

- java.net의 GlassFish 프로젝트의 XMLChar.java (r1.3)
- Apache의 Sling 프로젝트의 XMLChar.java (r599941)

```
/* ... CDDL, GPL, GPL v2 license fragment ... */
```

```
package org.apache.jasper.xmlparser;
```

```
import java.util.Arrays;
```

```
public class XMLChar {
```

```
    // constants
```

```
    private static final byte[] CHARS = new byte[1
```

```
    ....
```

```
    // static initialization
```

```
    static {
```

```
        CHARS[9] = 35;
```

```
        CHARS[10] = 19;
```

```
    }
```

```
    // public static methods
```

```
    public static boolean isSupplemental(int c) {
```

```
        return (c >= 0x10000 && c <= 0x10FFFF);
```

```
    }
```

```
    public static int supplemental(char h, char l)
```

```
        return (h - 0xD800) * 0x400 + (l - 0xDC00)
```

```
    }
```

```
    ...
```

```
}
```

```
/* ... Apache v2.0 license fragment ... */
```

```
package org.apache.sling.scripting.jsp.jasper.xmlparser;
```

```
/*
```

```
 * Licensed to the Apache Software Foundation (ASF) under one or more  
 * contributor license agreements. See the NOTICE file distributed with  
 * this work for additional information regarding copyright ownership.  
 * The ASF licenses this file to You under the Apache License, Version 2.0  
 * (the "License"); you may not use this file except in compliance with  
 * the License. You may obtain a copy of the License at
```

```
 *
```

```
     http://www.apache.org/licenses/LICENSE-2.0
```

```
 *
```

```
 * Unless required by applicable law or agreed to in writing, software  
 * distributed under the License is distributed on an "AS IS" BASIS,  
 * WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.  
 * See the License for the specific language governing permissions and  
 * limitations under the License.
```

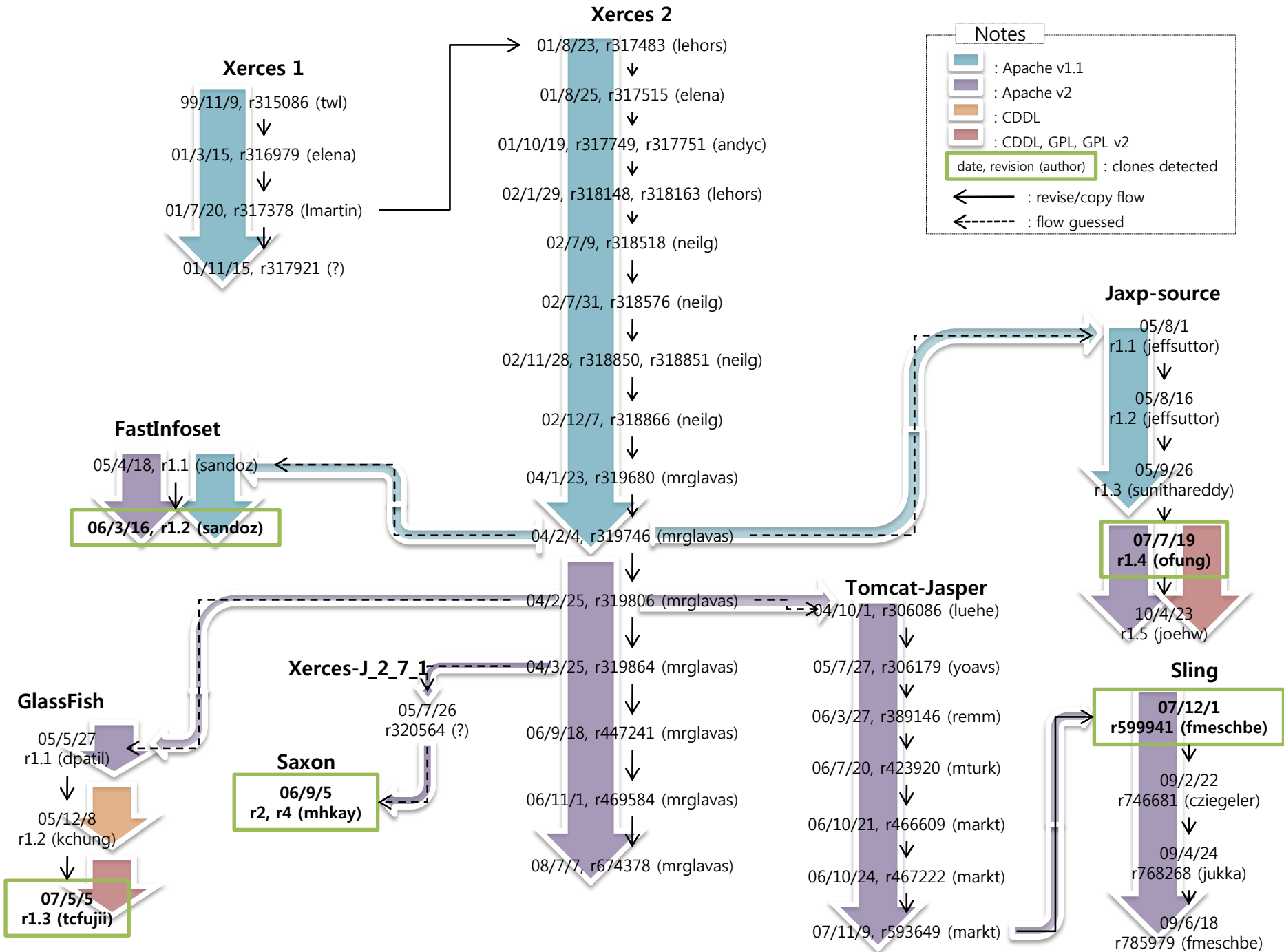
```
 * =====
```

```
 *
```

```
 * This software consists of voluntary contributions made by many  
 * individuals on behalf of the Apache Software Foundation and was  
 * originally based on software copyright (c) 1999, International  
 * Business Machines, Inc., http://www.apache.org. For more  
 * information on the Apache Software Foundation, please see  
 * <http://www.apache.org/>.
```

```
*/
```

```
0000;
```



# 조사 대상 코드 데이터베이스

커뮤니티	프로젝트 이름	하위 프로젝트 수	파일 수	줄 수
Apache Software Foundation	Hadoop	8	5,905	1,327,985
	Xerces	1	823	256,493
	Jakarta	8	2,717	468,118
	...			
	전체	99	97,077	16,929,668
Google Code	Google Web Toolkit	1	3,556	564,598
	jMonkeyEngine	1	1,139	270,202
	FEST	1	3,202	255,836
	...			
	전체	36	23,380	3,670,310
Java.net	GlassFish	1	15,437	2,994,385
	JavaCC	1	195	34,928
	LG3D	2	1,005	195,223
	...			
	전체	147	91,144	16,965,554
SourceForge.net	CUBRID Cluster	1	89	18,120
	Guacamole	1	36	3,707
	...			
	전체	4	1,061	178,807
전체		<b>286</b>	<b>212,657</b>	<b>37,572,912</b>

# 방법

## 1. 코드 클론 찾기

- Deckard

## 2. 복제된 코드 찾기

- SVM 분류기

## 3. 라이선스 찾기

- FOSSology

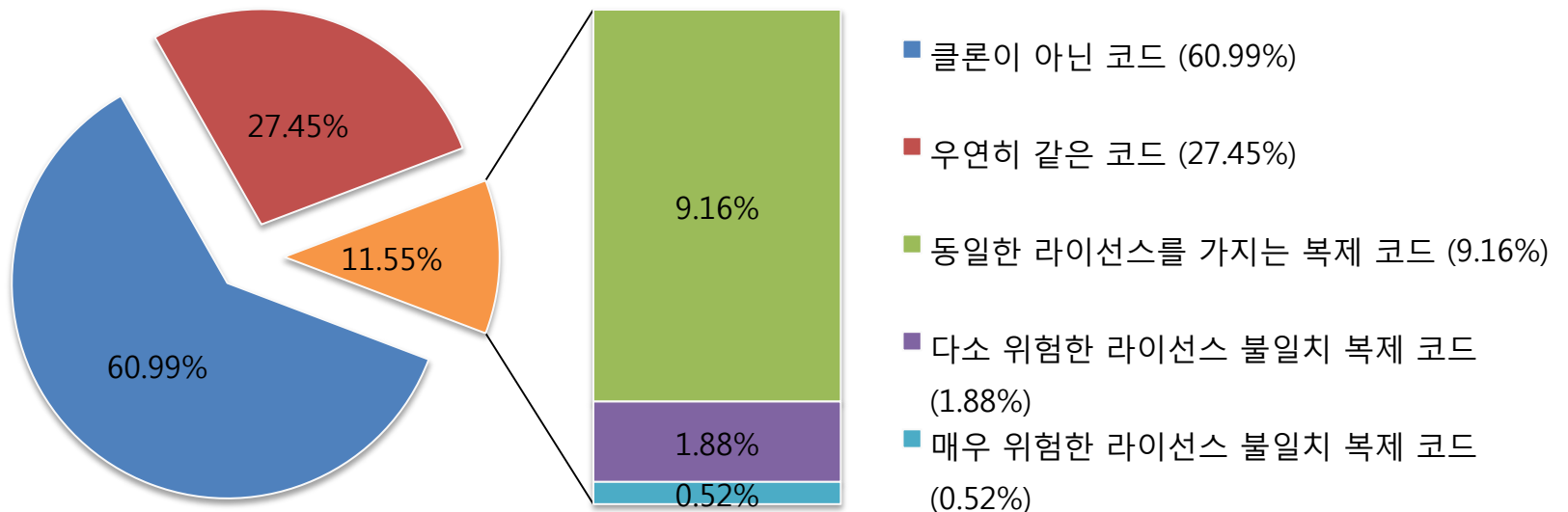
+ 잠재적 라이선스 위반 가능성이 있는 복제 코드들의 족보 추적

# 빈번하게 발생된 라이선스 불일치 복제 코드의 유형

위험 수준	라이선스 불일치 유형	라이선스 1	라이선스 2	파일 수	프로젝트 수	
낮음	라이선스 없음		Apache v2.0	1,799	178	
			CDDL, GPL v2	635	61	
			INRIA-OSL	283	38	
			GPL v2	217	47	
			UNKNOWN	163	39	
			GPL v3	131	15	
			LGPL v2.1+	87	31	
			Apache v2.0, CDDL, GPL v2	71	21	
			MIT	67	17	
			LGPL GNU C Library	60	22	
		라이선스 업그레이드	Apache v1.1	Apache v2.0	230	28
		라이선스 추가	Apache v2.0	Apache v2.0, CDDL, GPL v2	1,224	42
			Apache v1.1	Apache v1.1, CDDL, GPL v2	150	2
	라이선스 추가 및 업그레이드	Apache v1.1, CDDL, GPL v2	Apache v2.0	202	15	
높음	서로 다른 라이선스	Apache v2.0	CDDL, GPL v2	1,030	86	
		Apache v2.0	GPL v2	259	79	
		Apache v2.0	LGPL v2.1+	155	36	
		Apache v2.0	INRIA-OSL	133	62	
		Apache v2.0	UNKNOWN	128	59	
		Apache v2.0	LGPL GNU C Library	89	38	

# 잠재적 라이선스 위반의 비율

	관련된 파일 수		관련된 프로젝트 수	
전체	212,657	100%	286	100%
코드 클론	82,950	39.0%	280	97.9%
복제된 코드	24,566	11.6%	264	92.3%
다른 라이선스를 가진 복제 코드	5,095	2.4%	219	76.6%
위험한 라이선스 불일치 복제 코드	4,548	2.1%	216	75.5%
매우 위험한 라이선스 불일치 복제 코드	1,104	0.5%	132	46.2%





# 복제 코드의 족보 추적

(InstructionFactory.java)

Jakarta-BCEL 프로젝트의 InstructionFactory.java

리비전	커미터	커밋 날짜	라이선스
r152690	mdahm	2001/10/30	Apache v1.1
r152752	mdahm	2002/3/12	Apache v1.1
r152754	mdahm	2002/3/12	Apache v1.1
r152771	mdahm	2002/4/10	Apache v1.1
r152809	mdahm	2002/6/30	Apache v1.1
r152815	mdahm	2002/7/12	Apache v1.1
r152856	mdahm	2003/5/23	Apache v1.1
r152903	conor	2004/12/15	Apache v2.0
r312570	dbrosius	2005/10/10	Apache v2.0
r386056	tcurdt	2006/3/15	Apache v2.0
r411573	tcurdt	2006/6/4	Apache v2.0
r438659	dbrosius	2006/8/31	Apache v2.0
r509457	dbrosius	2007/2/20	Apache v2.0
r694866	dbrosius	2008/9/13	Apache v2.0
r826333	dbrosius	2009/10/18	Apache v2.0
r947879	sebb	2010/5/25	Apache v2.0
r948092	sebb	2010/5/26	Apache v2.0

Jaxp-Sources 프로젝트의 InstructionFactory.java

리비전	커미터	커밋 날짜	라이선스
r1.1	jeffsuttor	2005/7/31	Apache v1.1
r1.1.2.1	jeffsuttor	2005/7/31	Apache v1.1
r1.2	jeffsuttor	2005/8/16	Apache v1.1
r1.3	ofung	2007/7/19	Apache v1.1, CDDL, GPL v2

Tass 프로젝트의 InstructionFactory.java

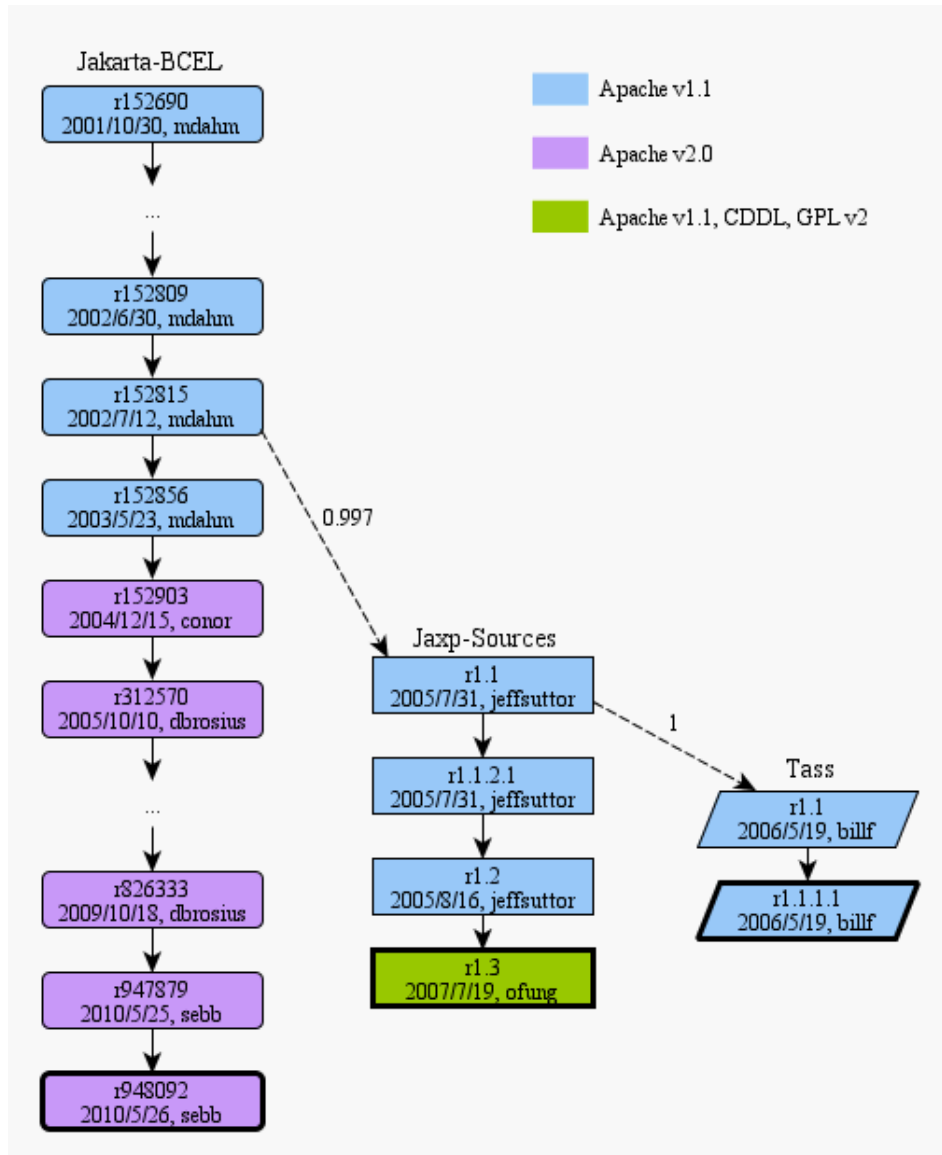
리비전	커미터	커밋 날짜	라이선스
r1.1	billf	2006/5/19	Apache v1.1
r1.1.1.1	billf	2006/5/19	Apache v1.1

# 복제 코드의 족보 추적

(InstructionFactory.java)

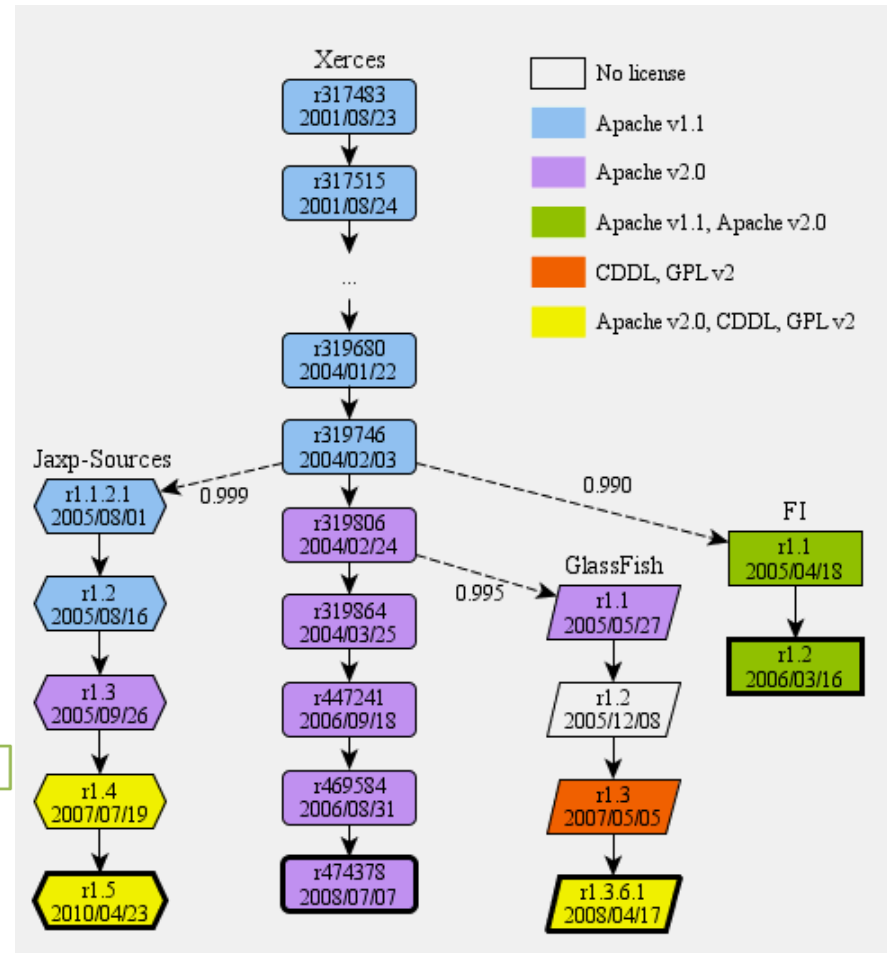
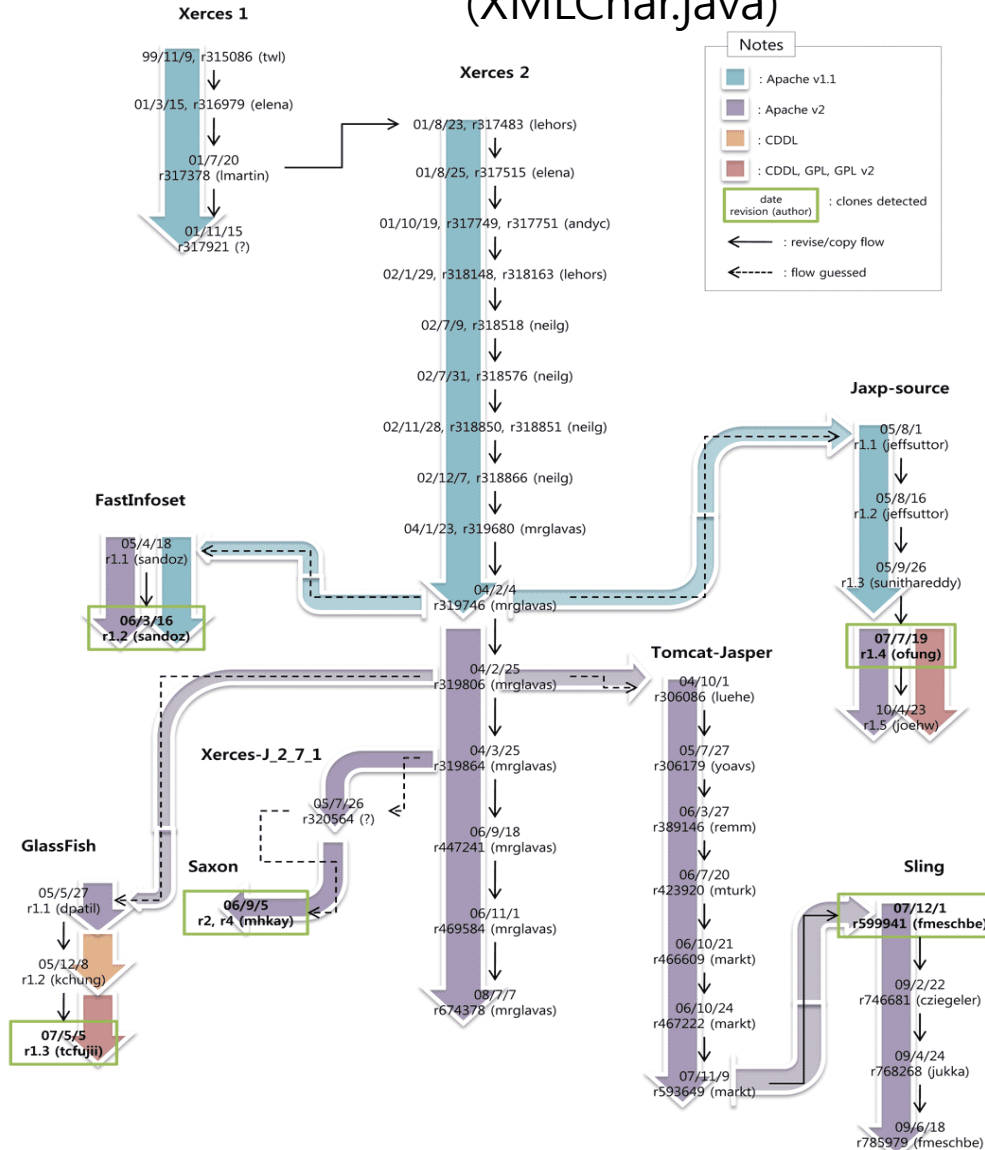
각 코드의 첫 번째 리비전과 다른 복제 코드의 유사도

프로젝트 이름	리비전	Jaxp-Sources r1.1	Tass r1.1
Jakarta-BCEL	r152690	0.9841	0.9394
	r152752	0.9927	0.9713
	r152754	0.9935	0.9744
	r152771	0.9947	0.9792
	r152809	0.9976	0.9904
	<b>r152815</b>	<b>0.9992</b>	0.9968
	r152856	0.9983	0.9936
	r152903	0.9697	0.8891
	r312570	0.9692	-
<b>r1.1</b>	-	<b>1</b>	
Jaxp-Sources	r1.1.2.1	-	0.9968
	r1.2	-	0.9968



# 복제 코드의 족보 추적

## (XMLChar.java)



감사합니다.