

要約：

ライブラリ・クラスの依存関係が引き起こす競合パターンとは？

Do theDependency Conflicts in My Project Matter?

Ying Wang, Ming Wen, Zhenwei Liu

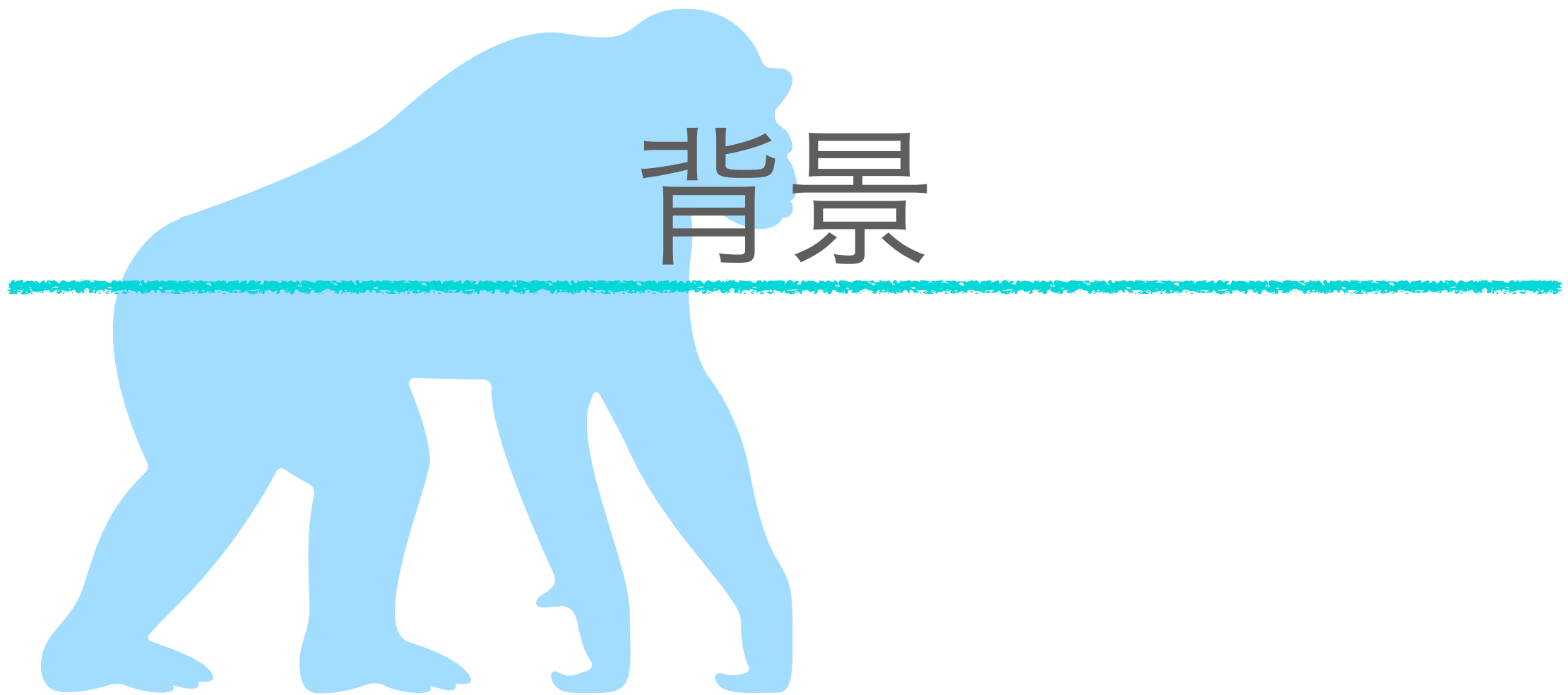
Rongxin Wu, Rui Wang, Bo Yang

Hai Yu, Zhiliang Zhu, Shing-Chi Cheung

(ESEC/FSE 2018) European Software Engineering Conference and
Symposium on the Foundations of Software Engineering: Pages 319 - 330

発表者 木村 祐太

背景



Javaとライブラリ

- Javaは広く使われている言語のため、ライブラリが多数存在している
- Mavenで管理されているものでも約877万のライブラリ
- 2000以上の人気のあるJavaプロジェクトを調査
 - 直接プロジェクトに関与しているライブラリは平均14ライブラリ
 - 間接的に関与しているものを合わせると平均して48ライブラリ

ライブラリの依存競合問題

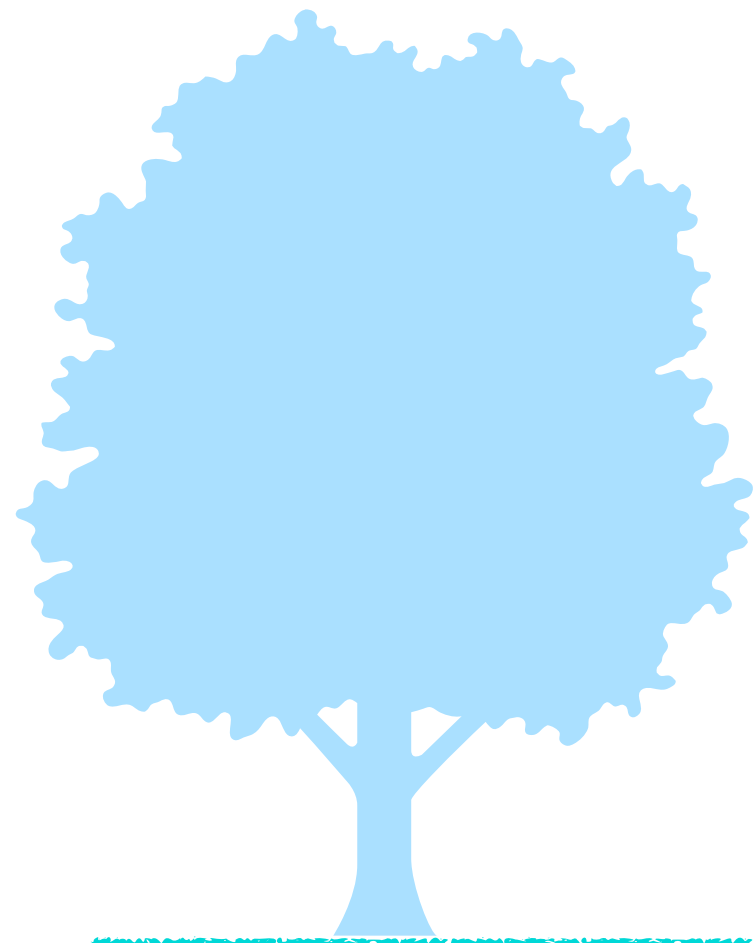
- ライブラリのバージョンが変わるとライブラリの競合が発生する場合がある
→ 同一クラスパス上に複数の同ライブラリが存在
- ライブラリに互換性があれば問題は発生しないが、互換性のないライブラリは多数存在する
- 大抵のビルドツールにはこれらの依存関係を管理するものを提供しているが、適切なバージョンである保証はない
- ビルドツールごとにローディングのメカニズムが違うため、競合発生条件が異なる

ライブラリの依存競合問題

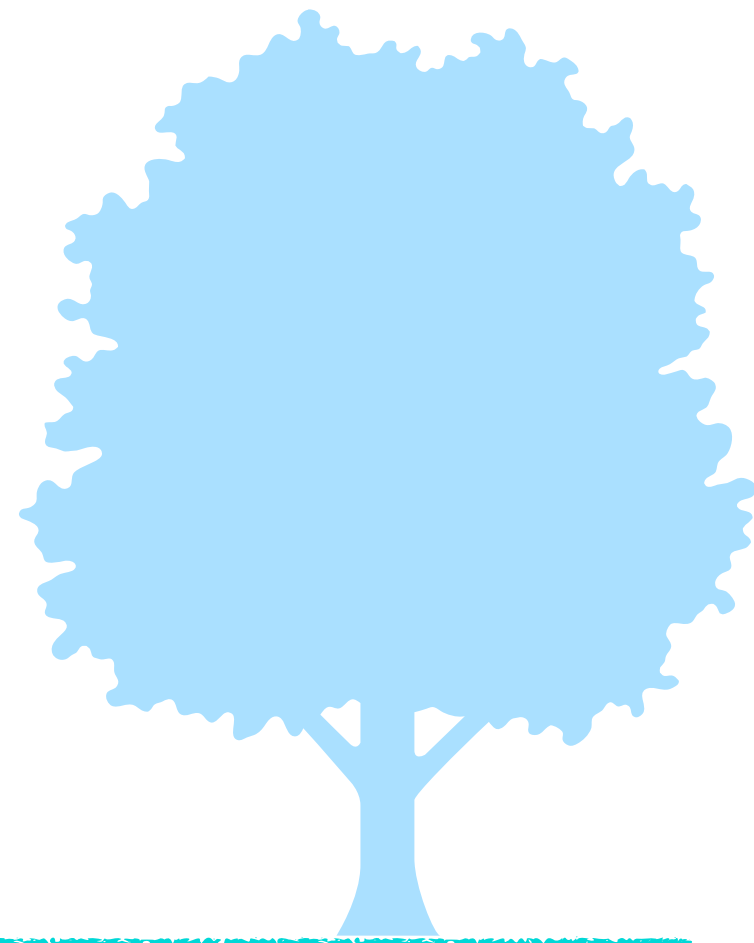
- GitHubの2,289のJavaプロジェクトのうち
 - 同じライブラリでバージョンの違うものを含んでいる 1,457(63.65%)
 - 異なるライブラリに同じクラスが存在する 1,003(43.82%)
 - 上記の2つの場合両方である 954(41.68%)
- 依存関係が引き起こす競合はよく起こる問題
→ 防止するのが難しい

目的

- 競合問題は深刻な問題であるが、はっきりとした依存競合パターンを理解するための体系的な研究がない
- ライブラリの依存関係が引き起こす競合パターンを明確にするための最初の実証研究とすること



実験



データセット

- Apache環境でMavenを用いてビルドされた71のJavaのオープンソースプロジェクト
- バグ管理システムに報告されている135のDC(依存競合)問題を調査
 - "conflicts"または"NoSuchMethodError"のようなキーワードで収集

Table 1: The statistics of subjects covered by collected 135 DC issues

Software	Star			Size (LOC) ¹			Categories ²	Issue tracking systems		Severity ³					Selected issues
71	Min.	Max.	Avg.	Min.	Max.	Avg.	16/28	Jira	BugZilla	M	C	B	N	Mi	135
	10	16398	796	0.9k	2218.4k	375.3k		115	20	74	7	19	13	22	

1. LOC denotes lines of codes; 1 K = 1000; 2. The software category refers to Apache official definition [5]; 3. M: Major; C:Critical; B:Block; N:Normal; Mi:Minor

RQ1 ・ RQ2

(競合パターン・修正パターン)

- 競合パターン (全135件)
 - A. ライブラリのバージョン間で競合 (39/135)
 - B. ライブラリ間のクラスが競合 (90/135)
 - C. ホストプロジェクトとライブラリのクラスが競合 (6/135)
- 修正パターン (全128件)
 - 1. 競合ライブラリのリネーム (25/128)
 - 2. クラスパスの変更 (42/128)
 - 3. 使用ライブラリのバージョンの最適化 (51/128)
 - 4. クラスローダーのカスタマイズ (5/128)

各パターンと修正関連要因

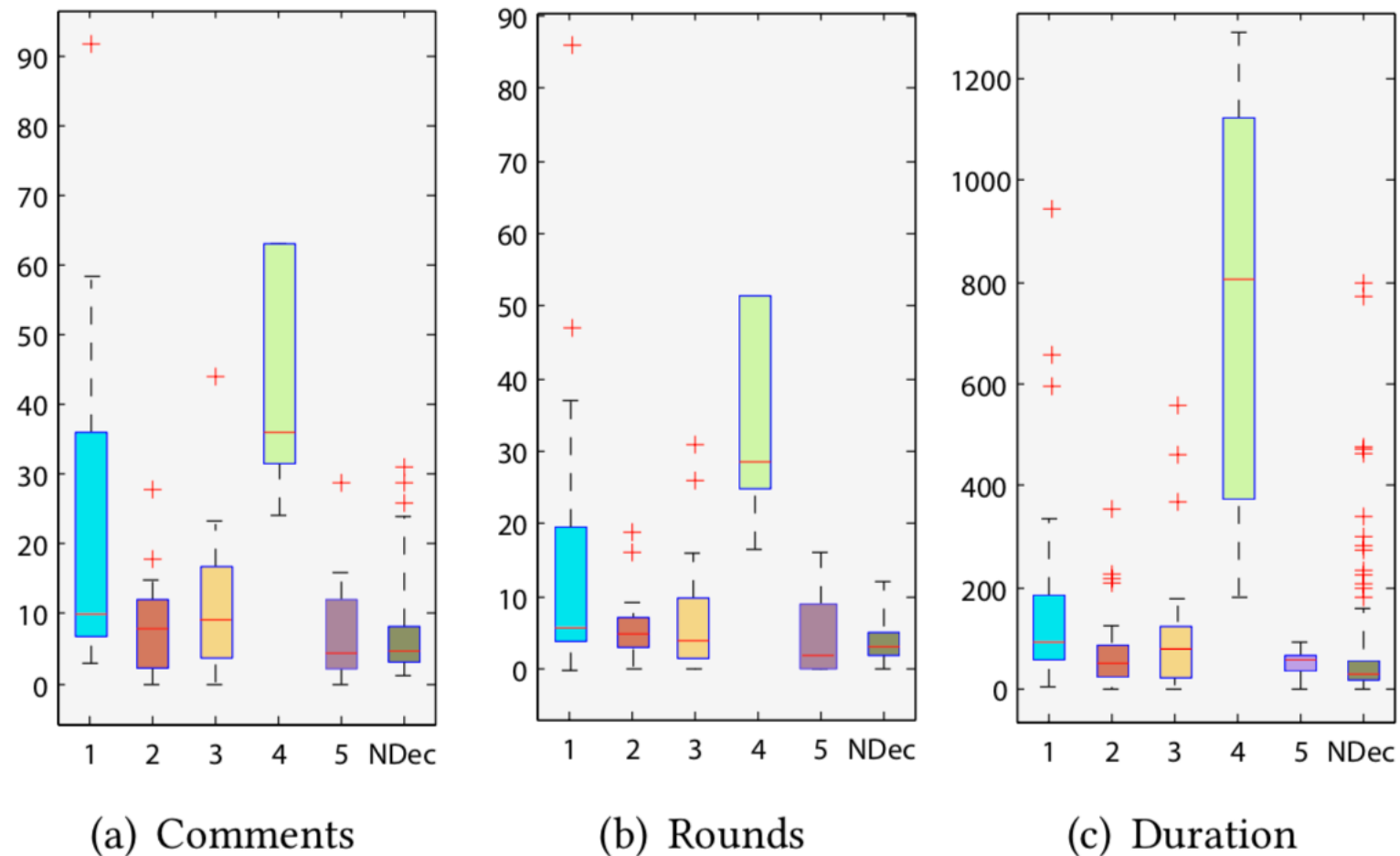


Figure 3: Comparison of the diagnosis and issue-fixing efforts (“1 ~ 4” = “Pattern 1 ~ 4”, “5” refers to other workarounds, “NDec = non-dependency conflict issues”), where (a) concerns the number of comments for each issue, (b) concerns the number of discussion rounds during the fixing process, and (c) concerns the number of days from the time when the issue was reported to the time when it was fixed.

問題パターンと修正パターン

- ・ パターンA, Bはどの修正パターンでも修正可能
- ・ パターンCは主にパターン 3

Table 2: The relations between manifestation and fixing patterns

Manifestation	Pattern 1	Pattern 2	Pattern 3	Pattern 4	Other
Pattern A	5	10	18	3	1
Pattern B	20	32	30	2	1
Pattern C	0	0	3	0	3

評価

問題レベルの定義

- Loaded : Lib3 v2.0
- Shadowed : Lib v1.0

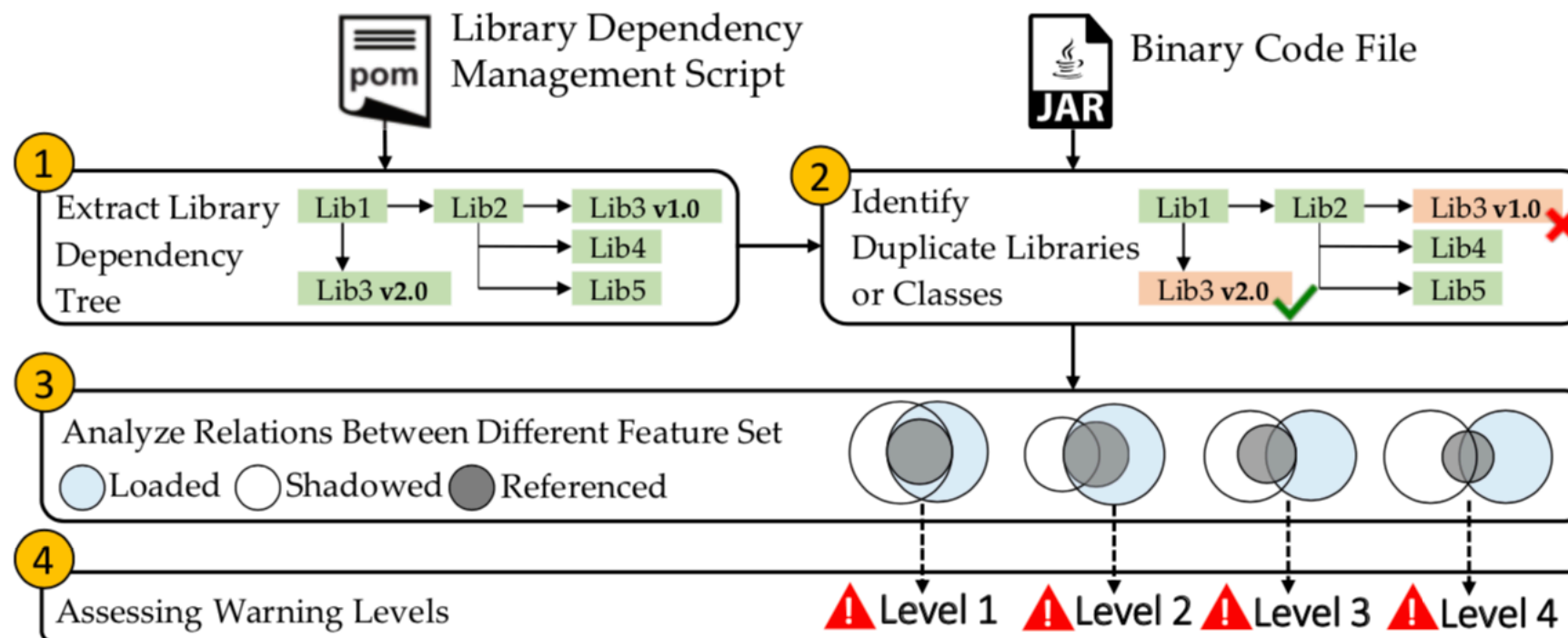


Figure 4: Architecture of DECCA

DECCA

(DependEnCy Conflicts Assessment)

- 統計分析を用いたDC検出ツール
- Level 3, Level 4の高レベル問題を検出することを目的

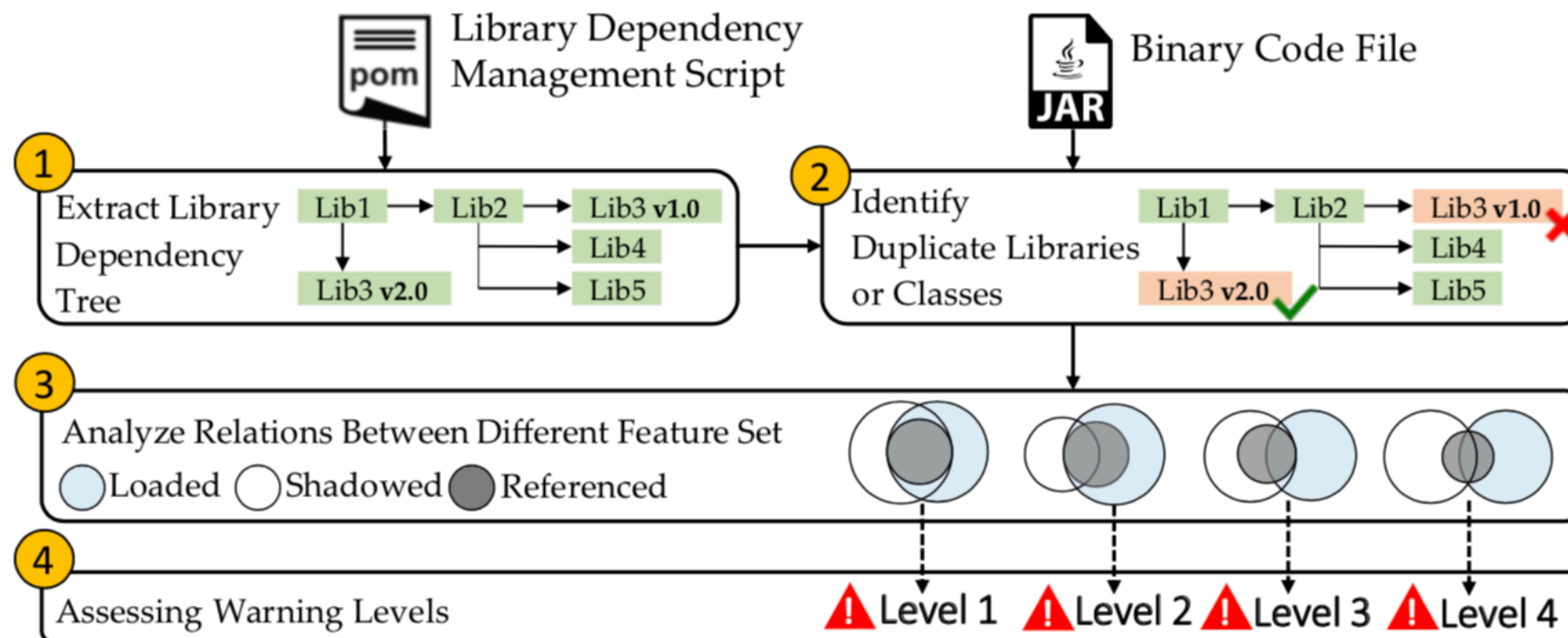


Figure 4: Architecture of DECCA

RQ3

DECCAの有効性

- True Positive : Level 3, Level 4を深刻であると識別
- False Positive : Level 3, Level 4を軽度であると識別
- True Negative : Level 1, Level 2を軽度であると識別
- False Negative : Level 1, Level 2を深刻であると識別

$$Precision = TP / (TP + FP)$$

$$Recall = TP / (TP + FN)$$

RQ3

DECCAの有効性

- Precision : 0.923
- Recall : 0.766
- F値 : 0.837
- 検出できなかったもの
 - 動的メソッド処理や反映

RQ4

DECCAの有用性

Table 3: Experimental subjects and checking results

ID	Project	Category	Revision	Size (LOC)	Star	Fork	Availability	Severity level				Bug ID
								L1	L2	L3	L4	
1	Spark [54]	Big data	8077bb0	130.0k	16262	15050	Apache/Github	40	1	0	0	SPARK-23509◇
2	Beam [13]	Big data	a750128	337.0k	1722	1038	Apache/Github	17	2	0	0	BEAM-3690◇
3	Bahir [11]	Extension tool	6ea42a8	0.9k	152	102	Apache/Github	22	0	1	1	BAHIR-159♣
4	Wicketstuff/Core [63]	Container	5cc41f5	228.5k	314	293	Apache/Github	16	1	0	0	Issue #621◇
5	Javaoze clue [38]	Command	23c9da4	2.8k	103	37	Github	18	1	0	0	Issue #61
6	ActiveMQ Artemis [8]	Network server	f6c5408	557.8k	271	326	Apache/Github	24	0	0	0	-
7	Apex Core [6]	Platform	4fb580f	87.0k	277	161	Apache/Github	34	0	0	0	-
8	Ignite [33]	OSGI	4e86660	2218.4k	1505	848	Apache/Github	7	0	0	0	-
9	Wicket [62]	Web framework	b728c69	352.5k	412	293	Apache/Github	2	0	0	0	-
10	Google/Closure-Compiler [25]	JS compiler	900251b	427.6k	4005	777	Github	4	1	0	0	Issue #2815♠
11	Orientdb [47]	Database	56ab1ac	496.3k	3366	718	Github	8	0	0	1	Issue #8111♠
12	Cm [18]	Web application	9e6f45b	19.k	12	6	Github	5	0	0	1	Issue #1◇
13	Brooklyn [15]	Cloud	48dbcc3	276.1k	69	47	Apache/Github	20	0	1	0	BROOKLYN-581
14	CarbonData [17]	Big data	9f2884a	127.9k	612	391	Apache/Github	25	4	0	0	CARBONDATA-2169
15	Prestodb [53]	Big data	89fed3a	0.8k	15	22	Github	16	1	0	0	Issue #29
16	Solr [57]	Network Server	d32048c	31.7k	295	207	Github	10	1	0	0	DATASOLR-447
17	tomcat exporter [60]	Exporter	70ac377	0.9k	19	10	Apache/Github	10	2	0	0	Issue #8
18	Hadoop Common [27]	Database	1e85a99	2042.8.k	5883	3987	Apache/Github	16	0	0	1	HADOOP-15261◇
19	Oozie [45]	Big data	9e662c7	198.6k	364	327	Apache/Github	25	0	1	0	OOZIE-3185♠
20	Accumulo [1]	Database	d98843b	563.8k	343	197	Apache/Github	33	1	0	0	ACCUMULO-4812♣
21	Eclipse jetty [22]	Debugging	b71cd70	375.9k	1868	1134	Github	6	2	0	0	Issue #2232
22	Parquet [50]	Big data	b82d962	0.9k	550	465	Apache/Github	2	1	0	0	PARQUET-1236◇
23	Apex Malhar [6]	Big data	0d98d05	243.7k	110	149	Apache/Github	34	1	0	0	APEXMALHAR-2556
24	Atlas [10]	Framework	6770091	123.4k	33	32	Apache/Github	44	1	1	0	ATLAS-2437

◇:The issues have already been fixed. ♠:The issues were confirmed and being fixed in process.

♣:They have been confirmed as DC issues, and they are suggested to be solved by the upstream third party libraries.

RQ4

DECCAの有用性

- 開発者のフィードバックより，開発者がDECCAについて興味を持っていることがわかった.

→ DECCAが有用であることを表している

まとめ

- Javaには無数のライブラリがあり，よくDC問題が起きる.
- DC問題がどういうものか，色々明らかにした
- DC検出ツールDECCAを作成し，実際に使ってみた
- 有効性・有用性ともに確認できた

所感

- 結果の理由付けがしっかりとされていて、読んでいて納得できるような構成になっていた（と思う）
- 所々のエキスパートなら知っている知識だろう感が滲み出っていて、少々読みづらかった
- 実用化されれば、有用なツールだと思った