

Supplement

1 SWRL rules

SWRL rule has the form:

antecedent \rightarrow consequent

where both antecedent and consequent are conjunctions of atoms written as $a_1 \wedge \dots \wedge a_n$. Variables in SWRL rules are denoted by prefixing them with a question mark (e.g., ?x). Below is an example showing the syntax:

Person(?p) \wedge hasAge(?p, ?age) \wedge
swrlb:greaterThan(?age, 17) \rightarrow Adult(?p)

This rule states that if there is a person p whose age is greater than seventeen, then this person is an adult (swrlb:greaterThan() is a built-in function in SWRL used to increase its expressiveness).

We next list all the SWRL rules for all unethical behavior detected by Etor.

(S1) *No attribution to the author in code.*

The SWRL rule for **S1** is listed below:

GHIssue(?i) \wedge isIssueMessageBody(?i, ?b) \wedge
swrlb:contains(?b, "https://stackoverflow.com/") \wedge isWebLink(?b, ?w) \wedge
isIssueOwner(?i, ?u1) \wedge openStream(?w, ?s) \wedge
swrlb:booleanNot(contains(?s, ?u1)) \wedge isFileCount(?r, ?fc) \wedge
swrlb:greaterThan(?fc, 0) \wedge isFileContent(?r, ?ffct) \wedge
isContentCount(?ffct, ?ffctc) \wedge swrlb:greaterThan(?ffctc, 0) \wedge
isContent(?ffct, ?ffcc) \wedge isSimilarCode(?ffcc, ?s) \wedge
swrlb:booleanNot(contains(?ffcc, ?w)) $\wedge \rightarrow$ S1(?i)

(S2) *Soft forking.* The SWRL rule for **S2** is listed below:

GHRepository(?r1) \wedge GHRepository(?r2) \wedge swrlb:isSimilarCode(?r1, ?r2)
 \wedge isForkCount(?r1, ?fc) \wedge swrlb:greaterThan(?fc, 0) \wedge isFork(?r2, null)
 \rightarrow S2(?r2)

(S5) *No license provided in public repository.* The SWRL rule for **S5** is listed below:

GHRepository(?r) \wedge isFileCount(?r, ?fc) \wedge swrlb:greaterThan(?fc, 0) \wedge
isLicenseFile(?r, null) \wedge isReadmeFile(?r, ?rmf) \wedge isContentCount(?rmf,
?rmcc) \wedge swrlb:greaterThan(?rmcc, 0) \wedge isContent(?rmf, ?c) \wedge
swrlb:booleanNot(contains(?c, [License])) \rightarrow S5(?r)

(S6) *Uninformed License change*. The SWRL rule for **S6** is listed below:

```

GHRepository(?r) ∧ isFileCount(?r, ?fc) ∧ swrlb:greaterThan(?fc, 1) ∧
  isLicenseFile(?r, ?lf) ∧ isContentCount(?lf, ?fctc) ∧
  swrlb:greaterThan(?fctc, 0) ∧ isContent(?lf, ?c) ∧ isPathCount(?lf,
    ?pc) ∧ swrlb:greaterThan(?pc, 0) ∧ isPath(?lf, ?p) ∧
  isCommitCountByPath(?r, ?ctp) ∧ swrlb:greaterThan(?ctp, 1) ∧
  isCommitByPath(?r, ?cl) ∧ isCommitCodeChange(?cl, ?cc) ∧
    swrlb:contains(?cc, [License]) ∧
    swrlb:booleanNot(contains(CHANGELOG, ?cc)) ∧
isPullRequestCountByCommit(?r, ?prc) ∧ swrlb:lessThan(?prc, 1) → S6(?r)

```

(S8) *Self-promotion*. The SWRL rule for **S8** is listed below:

```

GHIssue(?i) ∧ isIssueMessageBody(?i, ?b) ∧ swrlb:contains(?b,
  "https://github.com/") ∧ isWebLink(?b, ?r2) ∧ swrlb:notEqual(?r1, ?r2)
  ∧ isIssueOwner(?i, ?u) ∧ swrlb:booleanNot(isContributor(?r1, ?u)) ∧
  swrlb:booleanNot(contains(?b, [DEMO])) ∧ isContributor(?r2, ?u) → S8(?i)

```

(S9) *Unmaintained Android Project with Paid Service*. The SWRL rule for **S9** is listed below:

```

GHRepository(?r) ∧ isLatestRelease(?r, ?lr) ∧ isPublishedDate(?lr, ?d)
  ∧ durationGreaterThan(0.5, ?d, "Years") ∧ isFork(?r, null) ∧
  isFileCount(?r, ?fc) ∧ swrlb:greaterThan(?fc, 0) ∧ isReadmeFile(?r,
    ?rm) ∧ isContentCount(?rm, ?rmcc) ∧ swrlb:greaterThan(?rmcc, 0) ∧
  isContent(?rm, ?c) ∧ swrlb:contains(?c, "https://play.google.com") ∧
  isWebLink(?c, ?l) ∧ openStream(?l, ?s) ∧ swrlb:contains(?s, "in-app
    purchase") → S9(?r)

```

2 Keywords

Table 1: Ethical guidelines and keywords

Ethical guideline	Keywords
To respect copyright	copy, plagiarism, credit, fork
To help individuals make informed consent decisions easier via licensing	license
To avoid license violatio	proprietary, closed source
To respect expectations between people through goodwill	promote, advertise, promotion
To be responsible for the project maintenance	maintain, support, pay, purchase
To avoid fraudulent activities	malware, vulnerable, hacking
To be responsible for naming	name
To be responsible for explaining public actions	unfair
To avoid offensive language	offensive
To allow individuals to choose which tasks to perform	opt-in, option
To protect the right of an individual of personal information	privacy