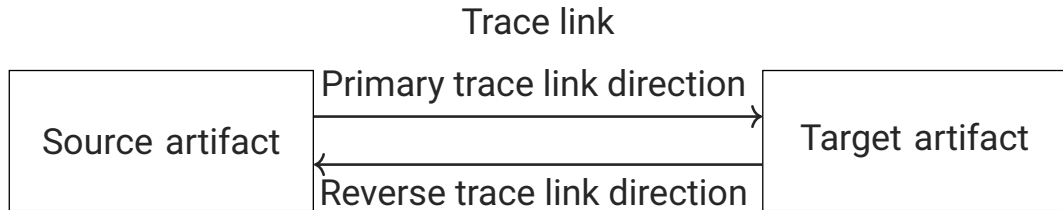# Automatically recovering students' missing trace links between commits and user stories

Sivana Hamer, Christian Quesada-Lopez, Marcelo Jenkins
{sivana.hamer,cristian.quesadalopez,macerlo.jenkins}@ucr.ac.cr

UNIVERSIDAD DE COSTA RICA

Tracing is the activity of establishing links between software artifacts

Trace link

| Source artifact | Primary trace link direction → | Target artifact |
| | ← Reverse trace link direction | |

(Figure and definition from Cleland-Huang and Gotel, 2014)

Tracing students' links between changes and requirements can help...

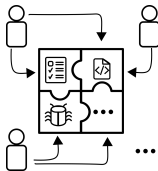**Tracing students' links between changes and requirements can help...**

Track
progress

Tracing students' links between changes and requirements can help...
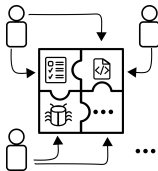

Track
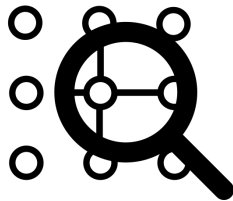progress

Determine
contribution

Tracing students' links between changes and requirements can help...

Track
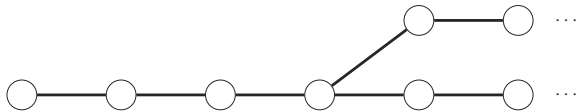progress

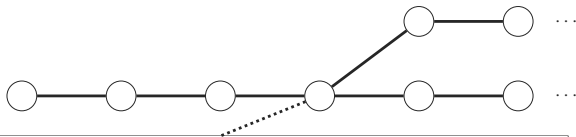Determine
contribution

Discover
patterns

Nowadays, it is common to establish manual change-issue traces with tagging.

Nowadays, it is common to establish manual change-issue traces with tagging.

Nowadays, it is common to establish manual change-issue traces with tagging.

Nowadays, it is common to establish manual change-issue traces with tagging.

Commit

🔴 2020-11-28   committed `ac1b39f`

Changed the EmailContentModel and the MailService to be able to send attachments throught emails.

UserStory:
PIG01IIC20-721

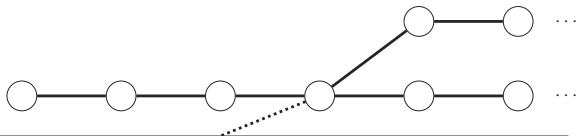ATE - Enviar por correo listado de incidentes del dashboard

Changes:
- Added the attachments list to the email content model DTO
- Changed the SendEmailAsync method to be able to send attachments throught emails.

Developers:
-

Nowadays, it is common to establish manual change-issue traces with tagging.

Commit

committed ac1b39f

2020-11-28

Changed the EmailContentModel and the MailService to be able to send attachments throught emails.

UserStory:
PIG01IIC20-721

ATE - Enviar por correo listado de incidentes del dashboard

Changes:
- Added the attachments list to the email content model DTO
- Changed the SendEmailAsync method to be able to send attachments throught emails.

Developers: PIG01IIC20-721 ATE - Enviar por correo listado de incidentes del dashboard ATE - Administración... 3
-

Nowadays, it is common to establish manual change-issue traces with tagging.

Nowadays, it is common to establish manual change-issue traces with tagging. However, manual links are inaccurate, untrustworthy and incomplete.

Commit

2020-11-28   committed ac1b39f

Changed the EmailContentModel and the MailService to be able to send attachments throught emails.
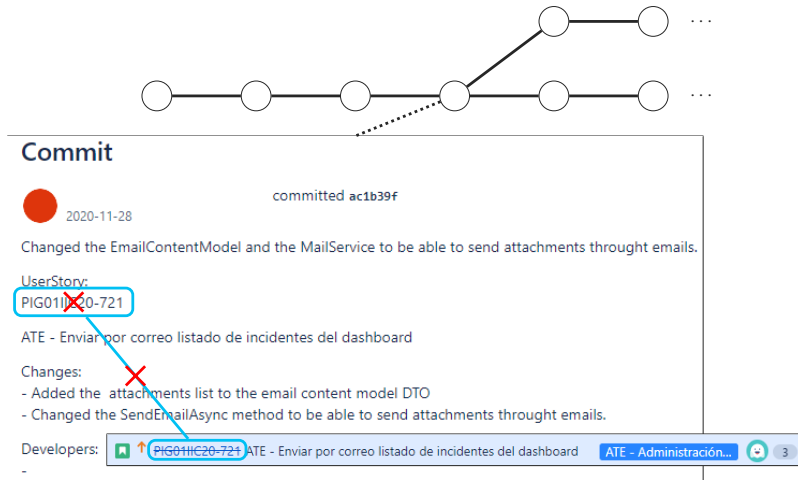
UserStory:
PIG01IIC20-721

ATE - Enviar por correo listado de incidentes del dashboard

Changes:
- Added the attachments list to the email content model DTO
- Changed the SendMailAsync method to be able to send attachments throught emails.

Developers: PIG01IIC20-721 ATE - Enviar por correo listado de incidentes del dashboard   ATE - Administración...   3
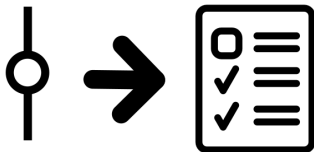-

(Cleland-Huang et al., 2014) (Rath et al., 2018) (Ruan et al., 2019) (Polaczek and Sosnowski, 2021)
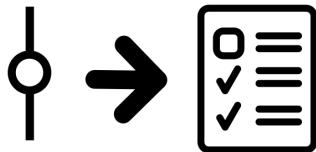
In our project, manual links were also incomplete as…

In our project, manual links were also incomplete as…

43 % of
commits linked

In our project, manual links were also incomplete as...

43 % of
commits linked

56 % of
stories linked

Problem:
What is the effectiveness of automatically recovering trace links between commits and user stories in student projects?

We studied a software engineering student project...

* committed or done

We studied a software engineering student project...

20
students

* committed or done

We studied a software engineering student project...

20
students

4
teams

* committed or done

We studied a software engineering student project...
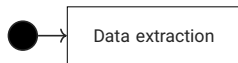
20
students

4
teams

1,035
commits

* committed or done

We studied a software engineering student project...

20
students

4
teams

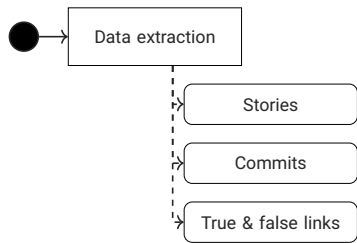1,035
commits

230
user stories*

* committed or done

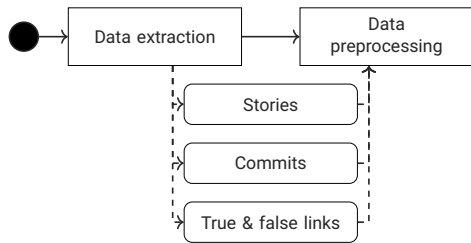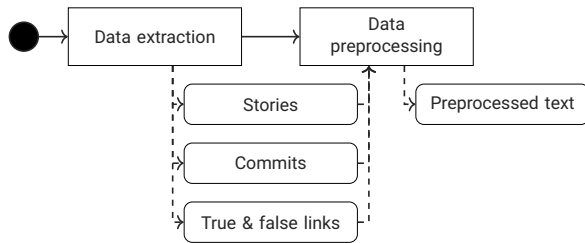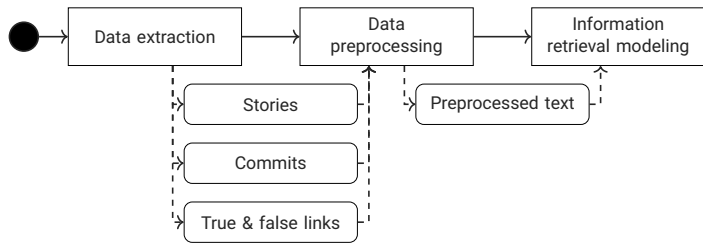**We automatically recover trace links between commits and user stories in students' software engineering projects...**

●

We automatically recover trace links between commits and user stories in students' software engineering projects...
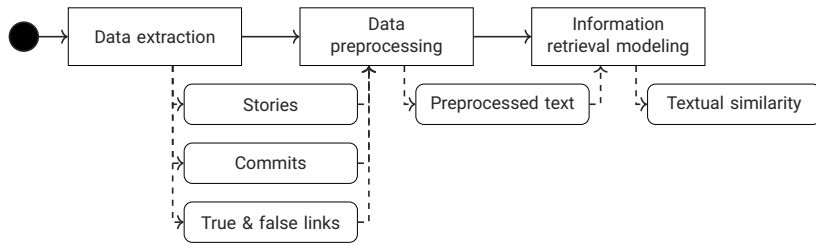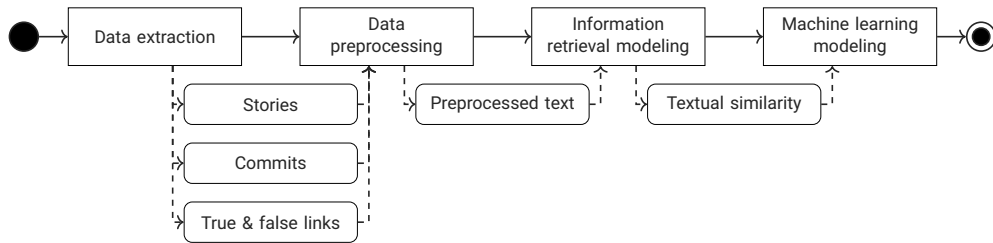
Data extraction

We automatically recover trace links between commits and user stories in students' software engineering projects...

We automatically recover trace links between commits and user stories in students' software engineering projects...
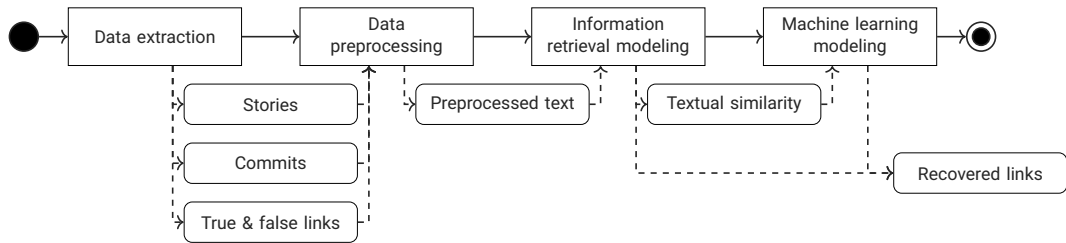
We automatically recover trace links between commits and user stories in students' software engineering projects...

We automatically recover trace links between commits and user stories in students' software engineering projects...

We automatically recover trace links between commits and user stories in students' software engineering projects...

We automatically recover trace links between commits and user stories in students' software engineering projects...
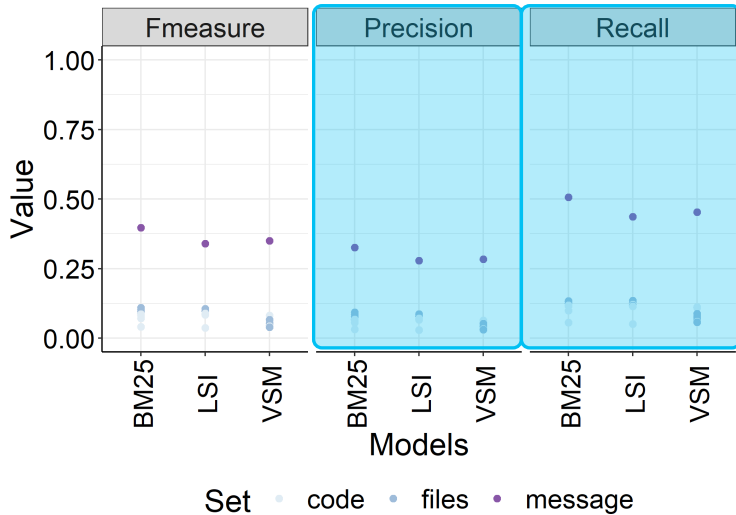
We automatically recover trace links between commits and user stories in students' software engineering projects...
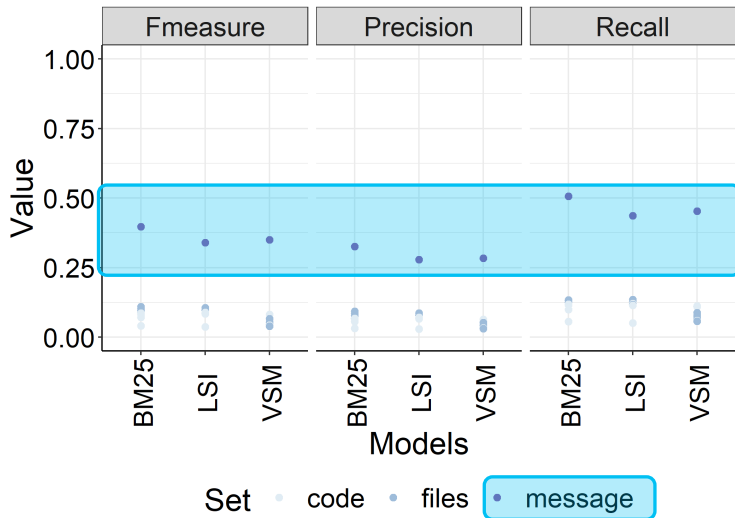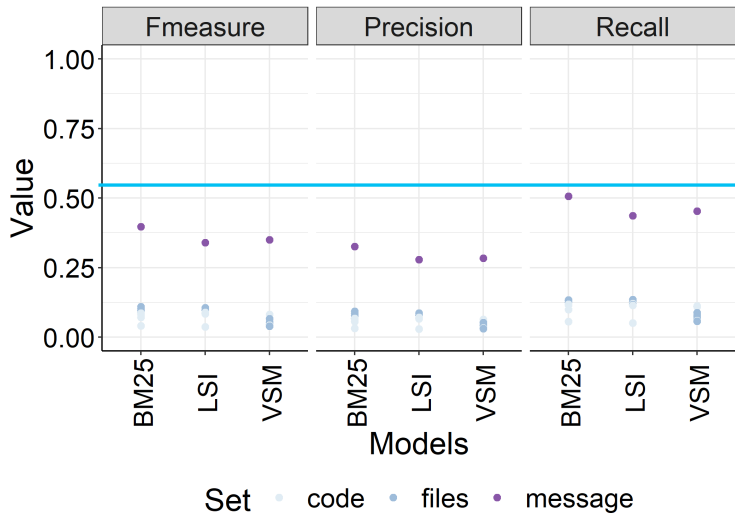
Information retrieval models recovered traces with many incorrect links
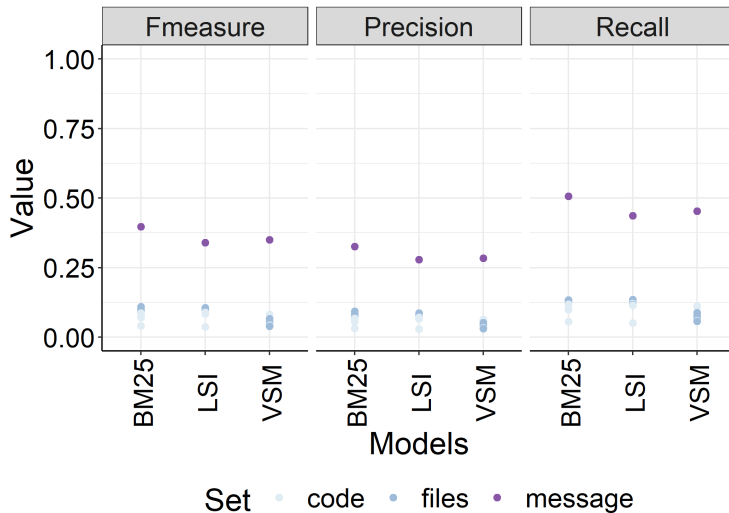
Information retrieval models recovered traces with many incorrect links

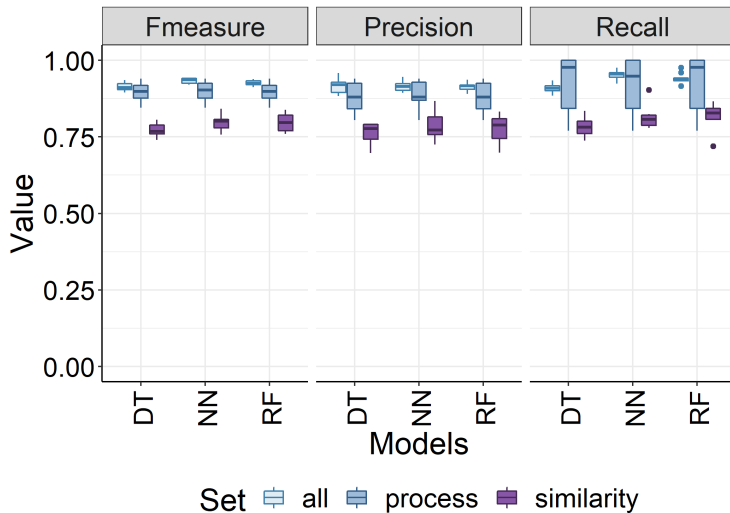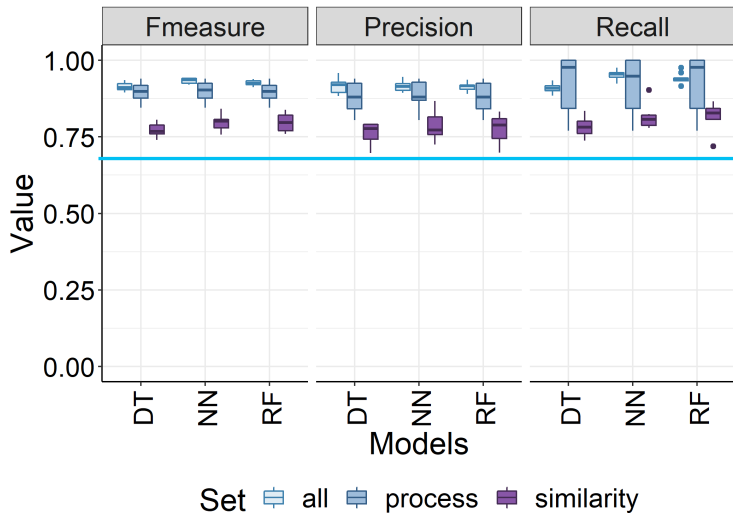# Information retrieval models recovered traces with many incorrect links

Information retrieval models recovered traces with many incorrect links

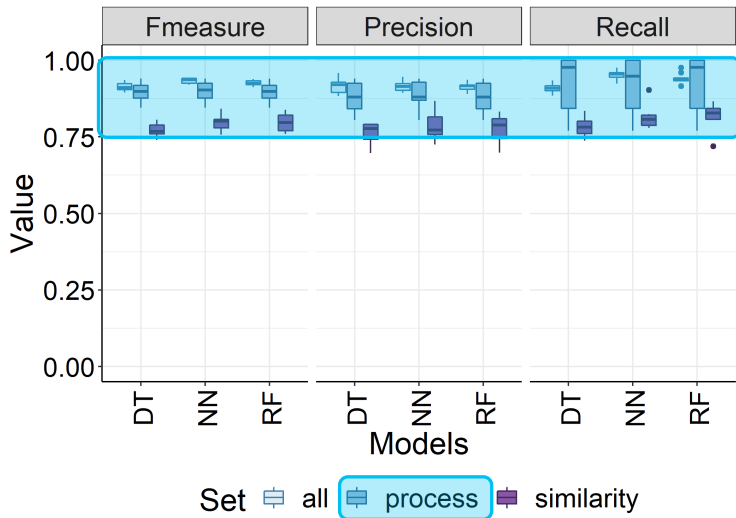# Machine learning recovered commit-story links with few incorrect links

Machine learning recovered commit-story links with few incorrect links

Machine learning recovered commit-story links with few incorrect links
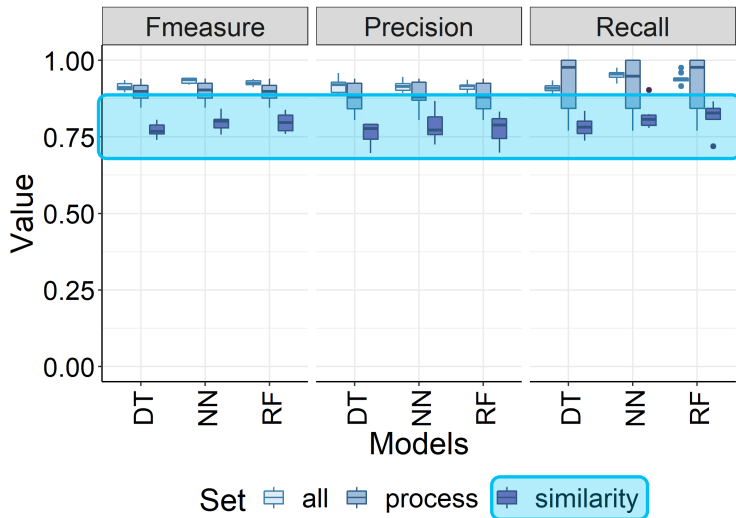
Machine learning recovered commit-story links with few incorrect links

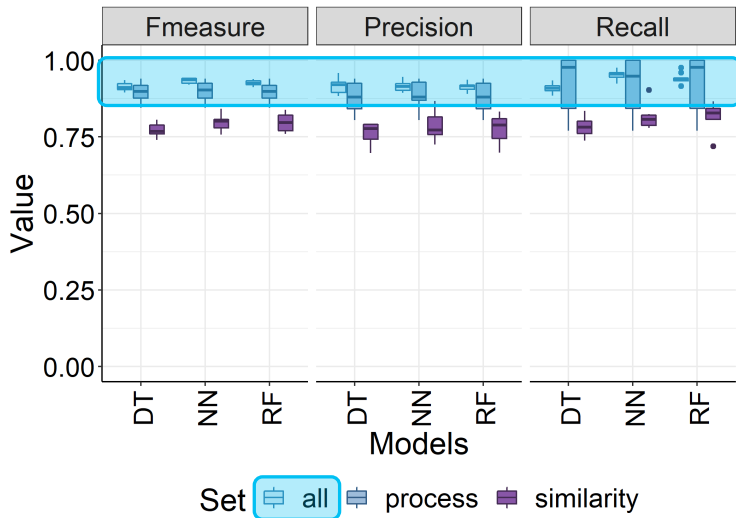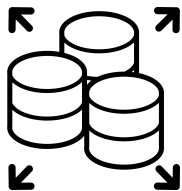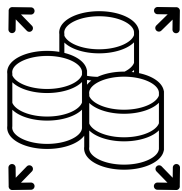# Machine learning recovered commit-story links with few incorrect links

Automatic trace recovery results are encouraging, with future work including…

Automatic trace recovery results are encouraging, with future work including…

Expand datasets
and data

Automatic trace recovery results are encouraging, with future work including...
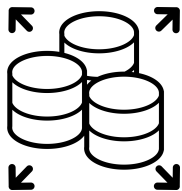

Expand datasets
and data

Create in-vivo
trace links

Automatic trace recovery results are encouraging, with future work including...

Expand datasets
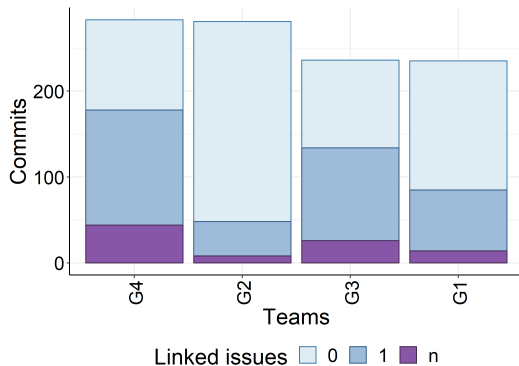and data
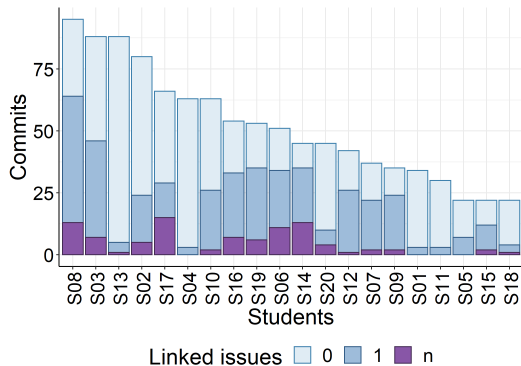

Create in-vivo
trace links

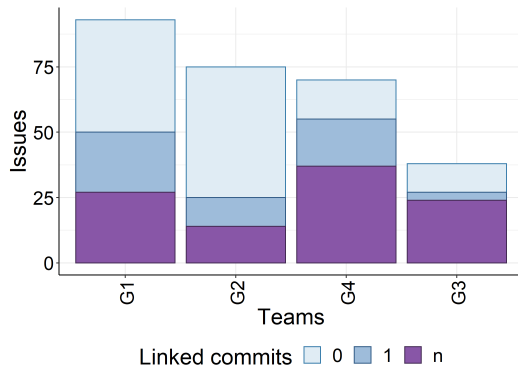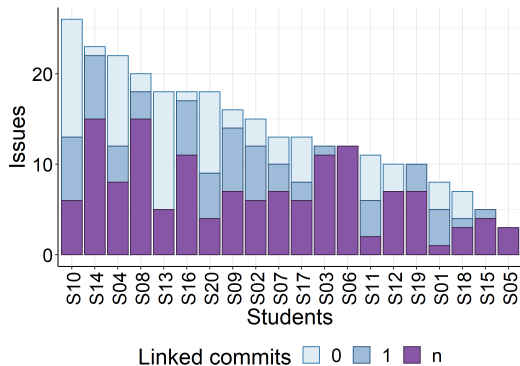
Gather
percepeptions

Thank you! Questions?

Automatically recovering students' missing trace links between commits and user stories

Sivana Hamer, Christian Quesada-Lopez, Marcelo Jenkins
{sivana.hamer,cristian.quesadalopez,macerlo.jenkins}@ucr.ac.cr

Linked commits by student and team

Linked issues by student and team

Machine learning model's features

Process. Incorporates process related information. We capture: (p_1) committer; (p_2) committer team; (p_3) story assignees; (p_4) teams assigned to story; (p_5) committer is an assignee (committer $==$ an assignee); (p_6) committer team is an assigned team (committer $==$ an assigned team); (p_7) story created date $<=$ commit date; and, (p_8) commit date $<=$ story resolution date. Features p_5, p_6, p_7 and p_8 are binary.

Similarity. Contains the textual similarity scores. We used: (s_1) commit message and story message similarity; (s_2) commit files and story message similarity; (s_3) commit code and story message similarity. Since we have multiple similarity scores, we select the highest performing information retrieval models for each similarity feature.

All. All the features ($p_1, \cdots, p_8, s_1, s_2, s_3$) from both the process and similarity set are selected.