ReDA
Review Data Analyzer

A Web-based Visualization Tool for Analyzing Modern Code Review Dataset

Patanamon Thongtanunam
Xin Yang, Norihiro Yoshida, Raula Gaikovina Kula,
Ana Erika Camargo Cruz, Kenji Fujiwara, Hajimu Iida
Modern Code Review is a tool-based, and occurs regularly in practice nowadays at companies and OSS projects [Bacchelli et. al.]
Modern Code Review
Modern Code Review

Developer → Code change → Gerrit

Code Review
Modern Code Review

Developer

Code change

Gerrit

Code Review

open source project

Change-Id: 8d63e932c8f1563692f52e0a3e37e7b4792f89a5
Owner: Alistair Buxton
Project: platform/system/core
Branch: master
Uploaded: Mar 2, 2013 8:30 AM
Updated: Nov 12, 2013 1:00 PM
Commit Type: Always Merge
Status: Review in Progress

Reviewer | Code-Review | Verified
--- | --- | ---
Oxen Orea | | 
Horacio Spinelli | +1 | 
Deckard Auto verifier | | 

Commit Message

```
adb shell uses terms like disable canonical input processing in order to get raw control codes but does not disable CR/LF translation. The default for Linux terminals is to convert CRLF to LF unless the running program specifically asks for this to be disabled. Since adb does not, there is no way to send a CRLF to any program on an adb shell. Many programs do in fact differentiate and are broken by this behaviour, notably nano. This patch sets the terminal flags to disable all line coding translation.
```

Signed-off-by: Alistair Buxton <aj_buxton@gmail.com>

Add Comment

Patch Set 1

```
Author: Alistair Buxton <aj_buxton@gmail.com> Mar 2, 2013 8:16 AM
Commit: Alistair Buxton <aj_buxton@gmail.com> Mar 2, 2013 8:16 AM
Parent: e7a94708e2fcb4e643e5b7514bae42b8f4472b3
```

Add Comment

```
Method commandline.c
+1, -0 Side-by-Side Unified
+1, -0 All Side-by-Side All Unified
```

Add Comment

```
Deckard Auto verifier Patch Set 1: Verified+1 This is an automated message from the Deckard... Mar 3, 2013
Deckard Auto verifier Patch Set 1: Verified This is an automated message from the Deckard... Sep 9 10:51 PM
Deckard Auto verifier Patch Set 1: Verified+1 This is an automated message from the Deckard... Sep 10 6:16 AM
Horacio Spinelli Patch Set 1: Pretty pathetic... I'm +1'ing this because the code is... 12:40 PM
Horacio Spinelli Patch Set 1: Code-Reviews+1 Forgot to +1, not sure how to edit. God it's so... 12:52 PM
Horacio Spinelli Patch Set 1: 1:00 PM
```

Add Comment
Modern Code Review

Code change ➞ Gerrit Code Review

Developer ➞ Reviewers

Gerrit Code Review

open source project

Commit Message

Disable CR/LF translation for adb interactive shell.

adb shell uses terminology to disable canonical input processing in order to get raw control codes but it does not disable CR/LF translation. The default for Linux terminals is to convert CR to LF unless the running program specifically asks for this to be disabled. Since adb does not, there is no way to send a CR to any program run on adb shell. Many programs do in fact differentiate and so are broken by this behavior, notably nano. This patch sets the terminfo flags to disable all line ending translation.

Change-Id: 1f80902d212b151e2f3f9a6a6962e535e3975c705f0fb41c295
signed-off-by: Alistair Buxton <a.j.buxton@gmail.com>

Reviewer | Code-Review | Verified
---|---|---
Cezar Dorea
Horacio Spinelli | +1 | 
Deckard Autorevisor

Dependences

Reference Version: Base t

Patch Set 1

<table>
<thead>
<tr>
<th>File Path</th>
<th>Comments</th>
<th>Size</th>
<th>Diff</th>
<th>Reviewed</th>
</tr>
</thead>
<tbody>
<tr>
<td>adb/commandline.c</td>
<td>Side-by-Side</td>
<td>Unified</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Commit Message

M adb/commandline.c
+1, -0 Side-by-Side
Unified

Comments

Deckard Autorevisor Patch Set 1: Verified
This is an automated message from the Deckard ...

Mar 3, 2013

Deckard Autorevisor Patch Set 1: Verified
This is an automated message from the Deckard ...

Sep 5 10:51 PM

Deckard Autorevisor Patch Set 1: Verified
This is an automated message from the Deckard ...

Sep 5 10:51 AM

Horacio Spinelli Patch Set 1: Pretty pathetic... I'm + filing this because the code is ...

12:40 PM

Horacio Spinelli Patch Set 1: Code-Reviews Forgot to +1, not sure how to edit. God its so ...

12:52 PM

Horacio Spinelli Patch Set 1:

1:00 PM

GOT IT! You are the first repository! I download and the first time i ever compile on Linux! THANKS!!!
Modern Code Review

- Determine Quality
- Find Defects
- Comment

Developer

Gerrit Code Review

Reviewers

Code change

Gerrit

open source project

Commit Message

File Path | Comments | Size | Diff | Reviewed
--- | --- | --- | --- | ---
| | | | | 

Dependencies

Reviewer | Code-Review | Verified
--- | --- | ---
| | | 

Developer: Alastair Buxton <a.j.buxton@gmail.com>

Reviewer: Horacio Spinelli

Commit: Alastair Buxton <a.j.buxton@gmail.com>

Parent(s): a15d75d7a101bb46d00d00d5e51218c11 (pikko)

Download:
- repo download
- checkout
- pull
- cherry-pick

Review:
- Cherry Pick To

Commit Message:
- This is an automated message from the Dockard
- This is an automated message from the Dockard
- This is an automated message from the Dockard

Add Comment:
- Dockard Autoreviewer Patch Set 1: Verified+1 This is an automated message from the Dockard Mar 3, 2013
- Dockard Autoreviewer Patch Set 1: Verified+1 This is an automated message from the Dockard Sep 5 10:51
- Dockard Autoreviewer Patch Set 1: Verified+1 This is an automated message from the Dockard
- Horacio Spinelli Patch Set 1: Pretty pathetic... I'm filing this because the code is...
- Horacio Spinelli Patch Set 1: Code-Reviews+1 Forgot to +1, not sure how to edit. God its so...
- Horacio Spinelli Patch Set 1: 

GOT IT! You are the first repository I download! and the first time I ever compile on Linux! THANKS!!!
Modern Code Review

Determine Quality

Find Defects

Comment

Developer

Reviewers

Gerrit Code Review

Patch Sets (Commits)
Modern Code Review

Determine Quality

Find Defects

Comment

Developer

Gerrit

Code Review

Patch Sets (Commits)

Modified Files

Reviewers
Modern Code Review

Determine Quality

Find Defects

Comment

Developer

Gerrit Code Review

Reviewers

Review Scores

Patch Sets (Commits)

Modified Files

Open source project

Code change
Modern Code Review

Developer

Determine Quality

Find Defects

Reviewers

- Code change
- Code Review
- Gerrit

- Review Scores
- Patch Sets (Commits)
- Modified Files
- Comments

- Open source project

Example Gerrit Code Review:

- Change-id: 1b03022c6f6c25c3a4c163b37348fe84a70a
- Owner: Alastair Buxton
- Project: platform/system/core
- Branch: master
- Type: Commit
- Updated: Mar 2, 2013 5:30 AM
- Uploaded: Nov 12, 2013 1:00 PM
- Commit Message:
  
  Disabling CR/LF translation for adb interactive shell.
  
  The adb shell uses terms to disable canonical input processing in order to get raw control codes but it does not disable CR/LF translation. The default for Linux terminals is to convert CR to LF unless the running program specifically asks for this to be disabled. Since adb does not, there is no way to send a CR to any program run on adb shell. Many programs do in fact differentiate and so are broken by this behaviour, notably nano. This patch sets the terminus flags to disable all line coding translation.

- Signed-off-by: Alastair Buxton <aj.buxton@gmail.com>

Example Gerrit Code Review:

- Reviewer: Horacio Spinelli
- Reviewer: Deckard Autoverifier
- Code-Review: +1
- Code-Review: Verified

Example Gerrit Patch Set:

- Author: Alastair Buxton <aj.buxton@gmail.com>
- Commit: 1b03022c6f6c25c3a4c163b37348fe84a70a
- Parent: 1b03022c6f6c25c3a4c163b37348fe84a70a

- File Path: cmd/commandline.c
- Comments: +1, 0 Side-by-Side
- Diff: +1, 0 Side-by-Side

Example Gerrit Comments:

- Deckard Autoverifier: Patch Set 1: Verified +1 This is an automated message from the Deckard... Mar 3, 2013
- Horacio Spinelli: Patch Set 1: Pretty pathetic... Im filing this because the code is... 12:40 PM
- Horacio Spinelli: Patch Set 1: Code-Reviews Forget to +1, not sure how to edit. God its so... 12:52 PM

Add Comment
Modern Code Review
Modern Code Review

New & Rich source
Modern Code Review

New & Rich source Code Quality

open source project

Reviewer | Code-Review | Verified
----------|-------------|---------
Horacio Spinelli | +1 | ✓

Dependencies

Reference Version: Base-1

Patch Set 1

File Path | Comments | Size | Diff | Reviewed
----------|----------|------|------|---------
adb/commands.c | Side-by-Side | Unified

Commit Message

adb shell uses terms to disable canonical input processing in order to get raw control codes but it does not disable CR/LF translation. The default for Linux terminals is to convert CR to LF unless the running program specifically asks for this to be disabled. Since adb does not, there is no way to send a CR to any program run on adb shell. Many programs do in fact differentiate and as are broken by this behaviour, notably nano. This patch sets the termios flags to disable all line ending translation.

Change-Id: 189090228177b33e74e8dc63f2f7a579f39b5e907f

Horacio Spinelli | Patch Set 1: Coded-Reviews. I forgot to +1, not sure how to edit. God its so ... | 12:40 PM
Horacio Spinelli | Patch Set 1: Pretty pathetic... Im +1ing this because the code is ... | 12:52 PM

Add Comment
Modern Code Review

New & Rich source

Code Quality

Popular

open source project

Changes:
- 8195b020f3c0edc0d7a53f7ca8f7e19b5f8ed7e5
- 8b5e82382e37d98759c8e60c8a11f3

- Owner: Alastair Butson
- Project: A platform/system/core
- Branch: master
- Type: Bug
- Status: In Progress
- Updated: Mar 13, 2013 11:11 PM
- Author: Alastair Butson

Commit Message

- Disable GKLF translation for adb interactive shell.
- adk shell uses terms to disable canonical input processing in order to get raw control codes but it does not disable GKLF translation. The default for Linux terminals is to convert GKLF to LF unless the running program specifically asks for this to be disabled. Since adk does not, there is no way to send a CR to any program run on adk shell. Many programs do in fact differentiate and so are broken by this behaviour, notably nano. This patch sets the terminal flags to disable all line ending translation.

Author: Alastair Butson <alastair butson@gmail.com>

Add Reviewer

- Needs Code Review
- Name or Email or Group

Add Comment

GOT IT! You are the first repository I download and the first time I ever compile on Linux! THANKS!!!
Modern Code Review

New & Rich source
Code Quality
Popular

Reliable?
Modern Code Review

“Publicly available data from support tools is a rich source on mining, various potential perils should also be taken into consideration.”

[Kalliamvakou et. al.]
Modern Code Review

New & Rich source
Code Quality
Popular

Reliable?

“Publicly available data from support tools is a rich source on mining, various potential perils should also be taken into consideration.”

[Kalliamvakou et. al.]
ReDA is for facilitating researchers

Review Data Analyzer
ReDA is for facilitating researchers

Review Data Analyzer

Extracting a dataset
ReDA is for facilitating researchers

Extracting a dataset

Showing basic statistical summary
ReDA is for facilitating researchers

Review Data Analyzer

Extracting a dataset

Observing interesting patterns & Identify problems

Showing basic statistical summary
provides three visualizations
ReDA provides three visualizations

Review Statistic

- Number of Reviews vs Time
- Number of Reviewers
- Number of PatchSets
- Number of Modified Files
- Number of Comments
ReDA provides three visualizations

Review Statistic

Activity Statistic
Review Data Analyzer provides three visualizations:

- **Review Statistic**
- **Activity Statistic**
- **Contributor Activities**
Demo

http://reda.naist.jp

Android Open Source Software Project (AOSP) dataset

* Data was captured from October 2008 to January 2012
Example Findings
Android Open Source Software Project (AOSP) dataset
Data Observation using ReDA

Unusual Peaks in Graphs

Activity Statistic

Reviews without code changes

Review Statistic
Data Observation using ReDA

Unusual Peaks in Graphs

Reviews without code changes

Review Statistic

Activity Statistic
Data Observation by #1

Developers mistakes

- Submit wrong code version
- Accidentally submit changes

Reviews without code changes
Developers mistakes

- Submit wrong code version
- Accidentally submit changes

Reviews without code changes

Data Observation by #1

Review Statistic

Number of Modified Files

Review Statistic
Data Observation by

Developers mistakes

Submit wrong code version

Accidentally submit changes

Reviews related to VCS transactions

E.g., Merging branch transactions

Gerrit Code Review

Gerrit Code Review

Code

< / >
F1: 10% of all reviews were created but not for code review

Developers mistakes
- Submit wrong code version
- Accidentally submit changes

Reviews related to VCS transactions
- E.g., Merging branch transactions

F1: 10% of all reviews were created but not for code review
Data Observation using ReDA #2

Review Data Analyzer

Review Statistic
Data Observation using ReDA #2

Missing data

Review Statistic
Missing data

Git and Gerrit servers were down.
Reported by Google Developers

Incomplete dataset can bias the results.
F2: Review history in AOSP is incomplete.

- Missing data
- Git and Gerrit servers were down.
  Reported by Google Developers
- Incomplete dataset can bias the results.
Data Observation using ReDA #3

Activity Statistic

Contributor Activities
Data Observation using ReDA #3

Activity Statistic

Low activity number

Contributor Activities
Activity Statistic

Weekend

Contributor Activities
F3: Developers usually do code reviews on weekdays.

Activity Statistic

Contributor Activities
F3: Developers usually do code reviews on weekdays.
F3: Developers usually do code reviews on weekdays.
F3: Developers usually do code reviews on weekdays.

Activity Statistic

Main Contributor

Contributor Activities
F3: Developers usually do code reviews on weekdays.

F4: Main contributors are from Google and Android teams.
Modern Code Review is a tool-based, and occurs regularly in practice nowadays at companies and OSS projects [Bacchelli et. al.]
Modern Code Review is a tool-based, and occurs regularly in practice nowadays at companies and OSS projects [Bacchelli et. al.]

New & Rich source
Code Quality
Popular

Reliable?

Can we easily find a characteristic in the dataset?
Modern Code Review is a tool-based, and occurs regularly in practice nowadays at companies and OSS projects [Bacchelli et. al.]

Modern Code Review

Review Data Analyzer

provides three visualizations

Review Statistic

Activity Statistic

Contributors Activities

Can we easily find a characteristic in the dataset?

"Publicly available data from support tools is a rich source on mining, various potential perils should also be taken into consideration." [Kalliamvakou et. al.]

Modern Code Review

New & Rich source

Code Quality

Popular

Reliable?
Modern Code Review is a tool-based, and occurs regularly in practice nowadays at companies and OSS projects [Bacchelli et. al.]

Modern Code Review

Example Findings from Android Open Source Software Project (AOSP) dataset

F1: 10% of all reviews were created but not for code review

F2: Review history in AOSP is incomplete.

F3: Developers usually do code reviews on weekdays.

F4: Main contributors are from Google and Android teams.

Review Data Analyzer

Review Statistic

Activity Statistic

Contributors Activities

Can we easily find a characteristic in the dataset?

Gerrit

New & Rich source Code Quality Popular

Reliable?

Publicly available data from support tools is a rich source on mining, various potential perils should also be taken into consideration.

[Kalliamvakou et. al.]

Example Findings from Android Open Source Software Project (AOSP) dataset

F1: 10% of all reviews were created but not for code review

F2: Review history in AOSP is incomplete.

F3: Developers usually do code reviews on weekdays.

F4: Main contributors are from Google and Android teams.

Example Findings from Android Open Source Software Project (AOSP) dataset

F1: 10% of all reviews were created but not for code review

F2: Review history in AOSP is incomplete.

F3: Developers usually do code reviews on weekdays.

F4: Main contributors are from Google and Android teams.

Can we easily find a characteristic in the dataset?

Gerrit

New & Rich source Code Quality Popular

Reliable?

Publicly available data from support tools is a rich source on mining, various potential perils should also be taken into consideration.

[Kalliamvakou et. al.]

Example Findings from Android Open Source Software Project (AOSP) dataset

F1: 10% of all reviews were created but not for code review

F2: Review history in AOSP is incomplete.

F3: Developers usually do code reviews on weekdays.

F4: Main contributors are from Google and Android teams.