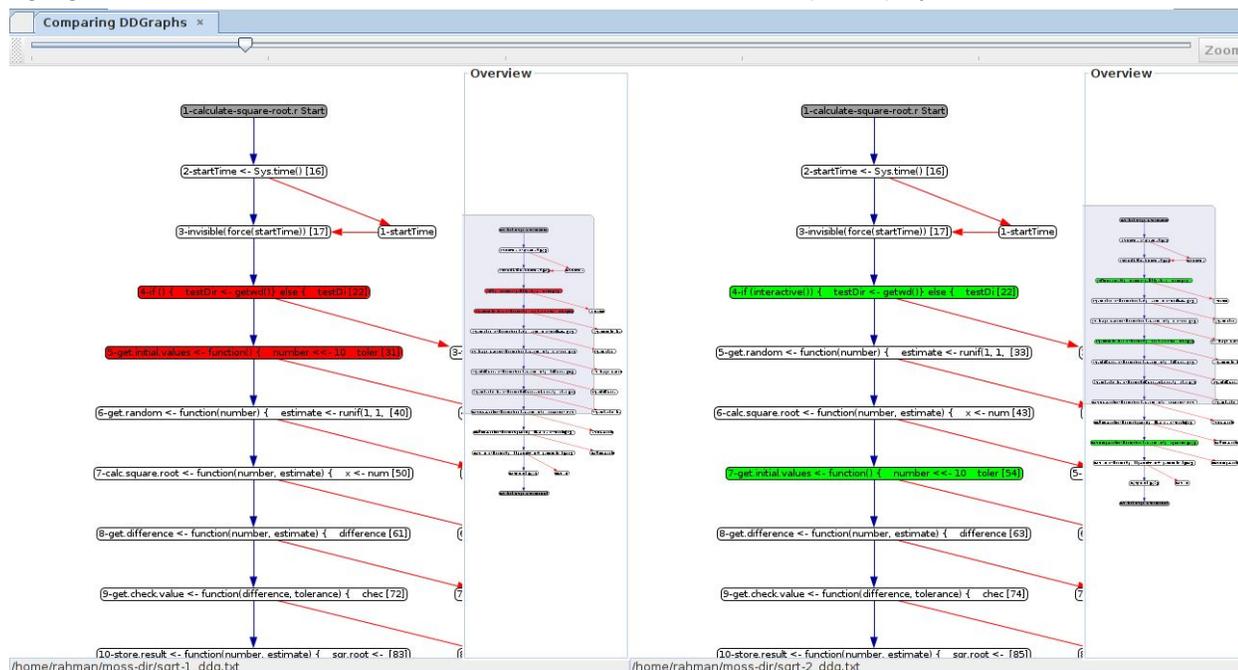


Compare DDGs

1. Introduction:

This comparison feature in DDG-Explorer compares 2 Data Derivation Graphs(DDGs) and then highlights the nodes which have been identified as a mismatch. A sample display from this feature is as follows:



This feature in DDGExplorer has unified scrolling i.e, both the DDGs are scrolled simultaneously.

a. Color Coding explained:

The comparison feature in DDGExplorer uses 'Red' and 'Green' colors to highlight the differences between the 2 graphs.

Red: The red nodes in the left ddg denote the nodes which are present in the left ddg but missing from the right ddg.

Green: The nodes in green represent the ones which are present in the right ddg but missing from the left ddg.

2. Technical Dependencies:

The feature makes use of the 'Linux and Unix diff' tool to evaluate the comparison result. The dependencies for a few of the popular OS distributions are as follows:

Linux/Mac : Any linux distribution/Mac OS comes with the 'diff' tool in-built.

Windows: 'diff' is not present by default. But this could be done by installing 'Cygwin', which is a Unix-style terminal for Windows. The installation and configuration details are as follows:

a. Install Cygwin:

Available at <https://cygwin.com/install.html>

b. Add Cygwin Path to Windows Environment Variable :

To make Cygwin work in your normal Windows command prompt, you need to add the location of Cygwin bin to the Windows Environment Variables.

(Detailed Instructions available at :

<http://www.howtogeek.com/howto/41382/how-to-use-linux-commands-in-windows-with-cygwin/>)

3. Working Explained:

- a. Once the left and the right DDGs have been uploaded by the user (from the option 'Select from File' to open files from the system or from the option 'Select from Database' to open the DDGs from the Jena database), 2 temporary files get created.
- b. The temporary file contains the reverse-engineered R script which created this DDG. More specifically, it contains the list of all procedural nodes present in a DDG.
- c. These 2 temporary files (each one for the 2 DDGs being compared) are now compared using the 'Unix Diff' tool.
- d. A sample output from the diff operation of 2 scripts is as follows:

```

calculate-square-root.rStart
startTime<-Sys.time()
invisible(force(startTime))
if(){testDir<-getwd()}else{testDi
get.initial.values<-function(){number<<-10toler
get.random<-function(number){estimate<-runif(1,1,
calc.square.root<-function(number,estimate){x<-num
|
calculate-square-root.rStart
startTime<-Sys.time()
invisible(force(startTime))
if(interactive()){testDir<-getwd()}else{testDi
get.random<-function(number){estimate<-runif(1,1,
calc.square.root<-function(number,estimate){x<-num
>
get.initial.values<-function(){number<<-10toler
get.difference<-function(number,estimate){difference
get.check.value<-function(difference,tolerance){chec
get.difference<-function(number,estimate){difference
get.check.value<-function(difference,tolerance){chec
store.result<-function(number,estimate){sqr.root<-
write.result<-function(sqr.root){file.name<-"result
store.result<-function(number,estimate){sqr.root<-
write.result<-function(sqr.root){file.name<-"result
>
storing.result<-function(number,estimate){sqr.root<
main<-function(){ddg.start("main")}get.initial.val
tryCatch
main<-function(){ddg.start("main")}get.initial.val
tryCatch
calculate-square-root.rFinish
calculate-square-root.rFinish

```

- e. The diff tool works in such a way that it tries to match the 2 files to the maximum extent possible on a line-by-line basis.

'|' - Denotes lines which are aligned wrt to lines in both the files but differ in content.

'>' - Denotes a completely unique line found in the right file.

'<' - Denotes a completely unique line found in the left file.

- f. '|' between the 2 nodes obtained from the left and right ddgs denote that the node on the left has been removed and the node on the right has been added. So, the left node is colored red and the right node is in green.

Similarly, '>' implies that there is a new node which was present in the right ddg and hence should be colored in 'Green' and '<' implies that this node in the left ddg is missing and there is no matching node for this in the right ddg and hence is colored in 'Red'.

4. Example:

Consider we have 2 versions of the same script 'Script-1.R':

Version 1 of script :

```
a <- 1
b <- 2
if (b == 1) {
  cc <- 3
} else {
  d <- 4
}
for (e in 1:5) {
  f <- e + 1
}
j <- switch (b, g = 7, h = 8, i = 9)
```

Version 2 of the same script:

```
a <- 1
b <- 2
c <- 3
j <- switch (b, g = 7, h = 8, i = 9)
```

The diff result of the 2 versions of the script obtained by the 'Unix diff':

script-1.RStart		script-1.RStart
a<-1		a<-1
b<-2		b<-2
if(b==1){cc<-3}else{d<-4}		c<-3
for(ein1:5){f<-e+1}	<	
j<-switch(b,g=7,h=8,i=9)		j<-switch(b,g=7,h=8,i=9)
script-1.RFinish		script-1.RFinish

The display of the diff result in DDGExplorer highlighting the differences between the 2 graphs:

