

Kepler - Bug #3909

RExpression fails with certain port names

03/19/2009 12:08 PM - Oliver Soong

Status:	Resolved	Start date:	03/19/2009
Priority:	Normal	Due date:	
Assignee:	ben leinfelder	% Done:	0%
Category:	actors	Estimated time:	0.00 hour
Target version:	Unspecified	Spent time:	0.00 hour
Bugzilla-Id:	3909		

Description

XP Pro x64 SP2, Java 1.6.0_11, Kepler 1.0.0 from kepler-project.org

Port names that are not simple R names causes the RExpression actor to fail with the error message that R cannot be found. The error message is misleading, and the variable names could still be used through the assign function. Escaping the name with backquotes will work in nearly all cases, but I think the assign function is slightly more robust.

History

#1 - 03/26/2009 08:44 AM - ben leinfelder

can you give a snippet of R that demonstrates assign and back quote and the kinds of variable names that are causing errors when not escaped?
thanks!

#2 - 03/31/2009 10:16 AM - Oliver Soong

Here are three examples that should cover most bases. The first one shows off backticks for quoting otherwise invalid object names. Off the top of my head, I think it's safe to use backticks pretty much everywhere. The second shows how to use the assign and get operators, which is slightly more robust and can handle names with backticks in them. The third shows a limitation with assign and get, which is that you need to use a temporary variable to make changes with functions like names and dim. I used a local wrapper to keep the global environment tidy. From what I've seen, weird names can get passed into the RExpression actor through the port name (try an input port that's just a number) or through a record (e.g., EML 2 Dataset passes a record with %).

Backticks and `check.names = FALSE` are probably easier to implement and will cover the vast majority of edge cases.

```
`1 weird name` <- data.frame(`+` = 1, `^-` = 2, check.names = FALSE)
`1 weird name`
```

```
assign("` name", data.frame(`^` = 3, `^` = 4, check.names = FALSE))
get("` name")
```

```
assign("` name", local({
tmp <- get("` name")
names(tmp) <- c("`", "`")
tmp}))
get("` name")
```

#3 - 03/31/2009 02:31 PM - ben leinfelder

RExpression2 actor handles "1 weird name" port names. The caveat being that you have to escape (`) them when referenced in the script (i don't think there's a way around that).

#4 - 03/31/2009 03:09 PM - ben leinfelder

i'm going to mark this as fixed since it works with the "next generation" RExpression actor.

we can't really go back and fix the 1.0 release.

i suppose we could go back to the original RExpression actor and place all the port name assignments in backticks (`). But if we're moving away from that implementation as it is, I'm going to advocate not doing that.

#5 - 03/31/2009 03:25 PM - Oliver Soong

Ben, I don't seem to be getting the same results as you. Here's a workflow under Kepler 1.x dev build 17250 that shows two failures, one due to a bad port name and another due to bad column names in the incoming record. I think putting backticks around the names will handle nearly all problems. Anybody who uses backticks in variable names is just asking for trouble.

Oliver

#6 - 04/01/2009 11:14 AM - ben leinfelder

I was saying it works with the RExpression2 actor. The sample workflow you have uses the original RExpression actor.

#7 - 04/01/2009 05:43 PM - ben leinfelder

I tried this sample workflow, but using RExpression2 actors instead. Here are my findings:

the "-" and "" port names for the RecordAssembler worked fine when the downstream RExpression2 actor received the record token.

-head(input) did not work with the RExpression2 actor when trying to print out debug information. My thinking is that head() gives an ill-defined data structure that the JNI layer doesn't exactly understand.

#8 - 04/02/2009 10:32 AM - Oliver Soong

I think I'm more confused now. I can't seem to find an RExpression2 actor in the components list. I had assumed it was a replacement for the RExpression actor in the dev build, but that seems to have been wrong.

#9 - 03/27/2013 02:24 PM - Redmine Admin

Original Bugzilla ID was 3909