

Kepler - Bug #3807

reserved symbols in record names passed to the RExpression actor generate a missing R error message

02/03/2009 04:08 PM - Oliver Soong

Status:	Resolved	Start date:	02/03/2009
Priority:	Normal	Due date:	
Assignee:	ben leinfelder	% Done:	0%
Category:	actors	Estimated time:	0.00 hour
Target version:	3.X.Y	Spent time:	0.00 hour
Bugzilla-Id:	3807		
Description			
XP Pro x64 SP2, Java 1.6.0_11-b03, Kepler 1.0.0, R 2.8.0			
The URL contains a bugged Kepler workflow. The data from KNB contains a column called %CC that causes an error when R is executing Kepler's RExpression initialization code. This causes Kepler to think that R is not found.			
A frustrating workaround is to disassemble the record and reassemble it, changing the name of the offending column.			

History

#1 - 02/03/2009 04:11 PM - ben leinfelder

will try this with the JRI implementation of R to see if that resolves the problem. RExpression2 needs more testing as it is!

#2 - 02/03/2009 04:51 PM - ben leinfelder

using RExpression2 (the JRI implementation) works!
of course this requires running from kepler-trunk and launching with java pointing to the appropriate native libraries in the "common" module....

#3 - 07/27/2009 12:39 PM - Oliver Soong

By way of update, Ben partially committed a patch for me that should mostly fix this. If I recall correctly, the remaining code should fix collisions with symbols in record names that are invalid in file names. Specifically, the remaining code addresses the cases where:

1. 2 ports have names that differ by reserved symbols
2. both ports cannot be converted to native tokens and so are saved to disk and passed as file names

The problem arises because the committed code converts almost all non-alphanumeric characters to underscores (_) to create valid file names, creating potential collisions. There is some code to avoid this, but it is implemented before files generated by the firing actor ports are processed. In other words, the existing code can only avoid collisions with ports belonging to different actors and not similarly named ports on a single actor.

#4 - 07/28/2009 11:53 AM - ben leinfelder

I believe the main part of this bug is fixed now. Certainly there are the potential for the temporary filename issues that are mentioned, but that can be avoided when constructing RExpression actors and their ports. For EML data sources that emit columns with reserved symbols (ie '%CC'), we can now handle that with back ticks (`) and check.names=FALSE.
Can we close this bug [and open a different one to address the port name/temp file issue] so that the bug tracking doesn't drift?

#5 - 07/28/2009 03:29 PM - Oliver Soong

The original summary of the bug has been addressed, so I'm closing this and will open a new one for the additional file collision issue.

#6 - 03/27/2013 02:24 PM - Redmine Admin

Original Bugzilla ID was 3807