Kepler - Bug #2050

EMLDataSource output as Ptolemy records

03/22/2005 04:01 PM - Dan Higgins

Status:	Resolved		**Start date:**	03/22/2005
Priority:	Normal		**Due date:**	
Assignee:	Jing Tao		**% Done:**	0%
Category:	general		**Estimated time:**	0.00 hour
Target version:	Unspecified		**Spent time:**	0.00 hour
Bugzilla-Id:	2050			

Description

Currently, when an eml2 datasource is dragged to the graph display, a port appears for each column in the datatable. Each port is given a name corresponding to the column name in the data table. A sequence of tokens is then output when the workflow runs, one for each row. The length of the sequence must be set by 'hand' either in the Director or somewhere else.

In many cases, it would seem appropriate to not output a sequence of tokens, but rather to include a single token which contains the column name, the number of rows, and the individual cell data. In Ptolemy/Kepler this can be done with a record token. The record name is the name of the column, the record contents is an array which contains the contents of the cells. The array length is the length of the column.

History

#1 - 04/07/2005 08:56 AM - Dan Higgins

Upon consideration, there are two options that should be considered.

1) Show each column as a port, as is currently done in the default, but output a single ARRAY from each port where the length of the array is the number of rows in the column. Some arrays would be doubles, some integers, and some strings, depending on the content of the column. The number of rows would be implicit in the vector lenght. (call this "ColumnVector" output?)

2) Optionally, have a single record port where named column arrays are the items in the record (ie col name is the name of record item, with all columns included). This might be called a "Column-based Record" view and represents an entire data table with a single record token.

#2 - 04/07/2005 09:33 AM - Shawn Bowers

I don't think it is correct to have one port per attribute. This approach looses the information that the ports are actually dependent. This assumes a particular domain functionality (i.e., that the director knows the dependency); but the constraint cannot be captured in Ptolemy's constraint language.

From a modeling perspective, it is more appropriate in Ptolemy to use a single port that outputs a tuple (i.e., a record). Of course, one could always connect an array deconstructor after the data set if desired. This approach (of outputing tuples instead of individual values) follows more closely the standard approach used in database systems and makes collection-oriented dataflows/programming much easier.

#3 - 04/08/2005 03:26 PM - Jing Tao

In EML2DataSource actor, two options - output as column vector or as column based Record, were added. If user chose the first one - as column vector as output, the output will be a ArrayToken which contains a array with entire column data for this field. If user chose the second one - as column based record, the port will output whole table as an array of columnDataArray.

#4 - 03/27/2013 02:18 PM - Redmine Admin

Original Bugzilla ID was 2050