

# ADAMS

Advanced **D**ata mining **A**nd **M**achine learning **S**ystem

Module: adams-spreadsheet



Peter Reutemann

September 28, 2012

©2012



THE UNIVERSITY OF  
**WAIKATO**  
*Te Whare Wānanga o Waikato*



Except where otherwise noted, this work is licensed under  
<http://creativecommons.org/licenses/by-sa/3.0/>

# Contents

<b>1</b>	<b>Introduction</b>	<b>7</b>
<b>2</b>	<b>Flow</b>	<b>9</b>
<b>3</b>	<b>Tools</b>	<b>11</b>
	<b>Bibliography</b>	<b>13</b>



# List of Figures

3.1	Viewer for spreadsheet files. . . . .	11
-----	---------------------------------------	----



# Chapter 1

## Introduction

Tabular data is a very common data format, not only for machine learning. The *spreadsheet* module offers some basic spreadsheet support for reading/writing and some generic actors. Other modules, like the *odf* or *excel* one, offer other native readers and writers. The data read by these readers can be processed with the same actors.





## Chapter 2

# Flow

The following transformers are available:

- *SpreadSheetCell* – retrieves the value of a specific cell in the spreadsheet.<sup>1</sup>
- *SpreadSheetColumnIterator* – iterates over all the columns in the spreadsheet and outputs the names.<sup>2</sup>
- *SpreadSheetColumnsByName* – generates a new spreadsheet with only the columns that match a regular expression (inverting is possible as well).
- *SpreadSheetInfo* – generates basic information on the spreadsheet object.
- *SpreadSheetPlotGenerator* – turns a spreadsheet into plot containers to be displayed in the *SequencePlotter* sink.<sup>3</sup>
- *SpreadSheetReader* – for reading spreadsheet files.<sup>4</sup>
- *SpreadSheetReplaceCellValue* – replaces cell values that match a regular expression.
- *SpreadSheetStatistic* – calculates statistics using the data stored in the spreadsheet.<sup>5</sup>
- *SpreadSheetSubset* – for obtaining a subset of the spreadsheet object (subset of columns and/or rows).
- *StringToSpreadSheet* – parses a string in CSV format and turns it into a spreadsheet object.

The following sinks are available:

- *SpreadSheetDisplay* – for displaying spreadsheet objects in tabular form.<sup>6</sup>
- *SpreadSheetWriter* – writes a spreadsheet object to a file with the chosen writer class.

---

<sup>1</sup>adams-spreadsheet-output\_cells.flow

<sup>2</sup>adams-spreadsheet-iterate\_cols.flow

<sup>3</sup>adams-spreadsheet-spreadsheet\_plot1.flow,      adams-spreadsheet-spreadsheet\_plot2.flow,  
adams-spreadsheet-statistic.flow

<sup>4</sup>adams-spreadsheet-output\_cells.flow

<sup>5</sup>adams-spreadsheet-statistic.flow

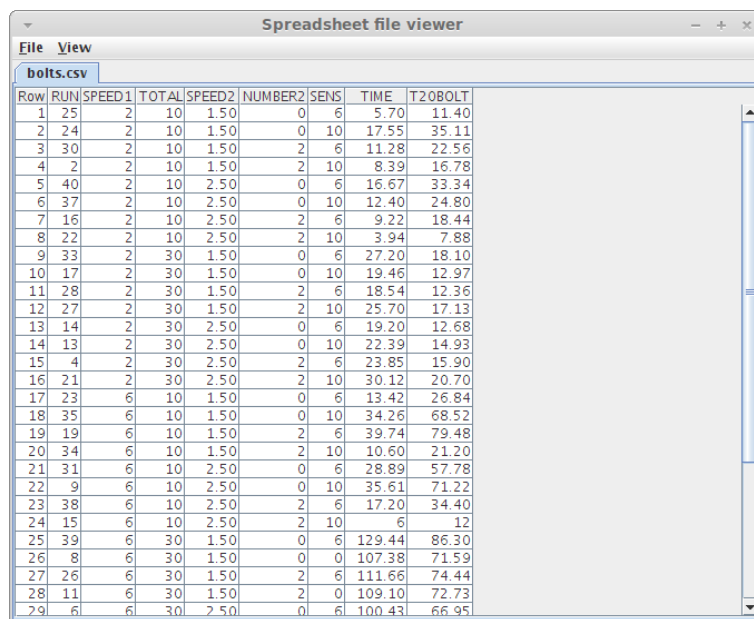
<sup>6</sup>adams-spreadsheet-display.flow



## Chapter 3

# Tools

The *Spreadsheet file viewer* is a simple tool for viewing all spreadsheet file formats that ADAMS supports. Figure 3.1 shows a dataset that was loaded from a CSV (comma-separated values) file.



The screenshot shows a window titled "Spreadsheet file viewer" with a menu bar containing "File" and "View". Below the menu bar is a tab labeled "bolts.csv". The main area displays a table with the following data:

Row	RUN	SPEED1	TOTAL	SPEED2	NUMBER2	SENS	TIME	T20BOLT
1	25	2	10	1.50	0	6	5.70	11.40
2	24	2	10	1.50	0	10	17.55	35.11
3	30	2	10	1.50	2	6	11.28	22.56
4	2	2	10	1.50	2	10	8.39	16.78
5	40	2	10	2.50	0	6	16.67	33.34
6	37	2	10	2.50	0	10	12.40	24.80
7	16	2	10	2.50	2	6	9.22	18.44
8	22	2	10	2.50	2	10	3.94	7.88
9	33	2	30	1.50	0	6	27.20	18.10
10	17	2	30	1.50	0	10	19.46	12.97
11	28	2	30	1.50	2	6	18.54	12.36
12	27	2	30	1.50	2	10	25.70	17.13
13	14	2	30	2.50	0	6	19.20	12.68
14	13	2	30	2.50	0	10	22.39	14.93
15	4	2	30	2.50	2	6	23.85	15.90
16	21	2	30	2.50	2	10	30.12	20.70
17	23	6	10	1.50	0	6	13.42	26.84
18	35	6	10	1.50	0	10	34.26	68.52
19	19	6	10	1.50	2	6	39.74	79.48
20	34	6	10	1.50	2	10	10.60	21.20
21	31	6	10	2.50	0	6	28.89	57.78
22	9	6	10	2.50	0	10	35.61	71.22
23	38	6	10	2.50	2	6	17.20	34.40
24	15	6	10	2.50	2	10	6	12
25	39	6	30	1.50	0	6	129.44	86.30
26	8	6	30	1.50	0	0	107.38	71.59
27	26	6	30	1.50	2	6	111.66	74.44
28	11	6	30	1.50	2	0	109.10	72.73
29	6	6	30	2.50	0	6	100.43	66.95

Figure 3.1: Viewer for spreadsheet files.

If there are more spreadsheet file formats registered, you can export the currently displayed spreadsheet to another format. Printing, of course, is always enabled. By default, the viewer displays each cell with as many digits after the decimal point as necessary. But you can also unify this and specify how many digits should be used for all floating point cells.



# Bibliography

- [1] *ADAMS* – Advanced Data mining and Machine learning System  
<https://adams.cms.waikato.ac.nz/>