

1 Theme

Tanagra add-on for OpenOffice 3.3 and LibreOffice 3.4.

The connection with spreadsheet applications is certainly a factor of success for Tanagra. It is easy to manipulate a dataset into **OpenOffice Calc** (up to version 3.2) and send it to Tanagra using the **TanagraLibrary.zip** extension for further analysis¹.

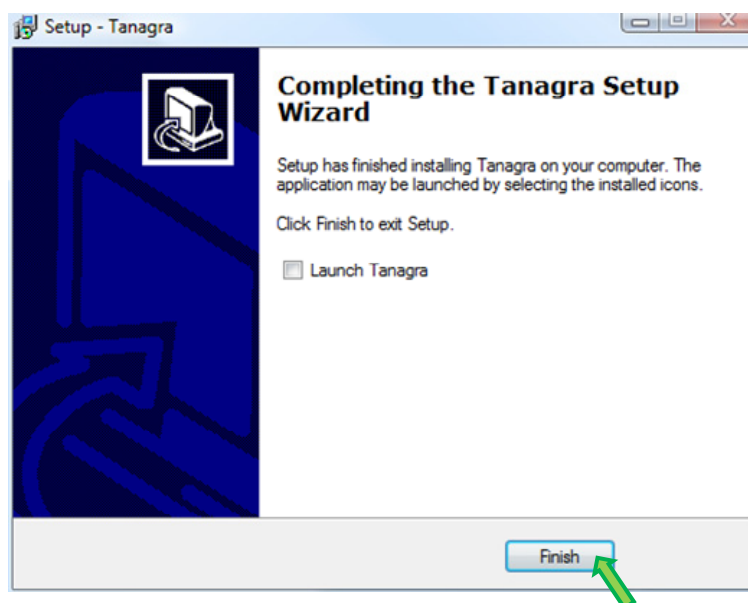
Recently, users have reported to me that the mechanism did not work with recent versions of OpenOffice² (version 3.3) and LibreOffice³ (version 3.4). I realized that, rather than a correction, it was more appropriate to elaborate a new module which meets the standard for managing extensions of these tools. The new library "**TanagraModule.oxt**" is now incorporated into the distribution.

This tutorial describes how to install and to use this add-on under **OpenOffice Calc 3.0**. The adaptation to **LibreOffice 3.4** is very easy.

2 Installing the add-on

2.1 Installing Tanagra

To install the add-on, we must install Tanagra on our system. We load and launch the "**setup_tanagra.exe**" setup file. There are not particular manipulations until the end of the installation process.



¹ See <http://data-mining-tutorials.blogspot.com/2008/10/ooocalc-file-handling-using-add-in.html> ; <http://data-mining-tutorials.blogspot.com/2009/04/launching-tanagra-from-ooocalc-under.html>

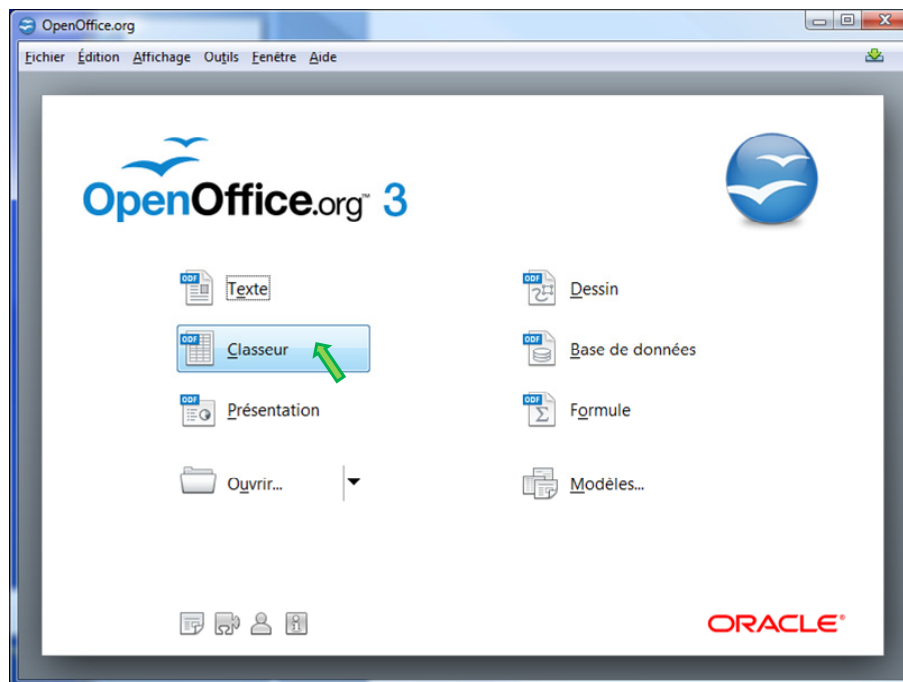
The equivalent exists for the **Excel** spreadsheet with the **Tanagra.xla** add-on: <http://data-mining-tutorials.blogspot.com/2008/10/excel-file-handling-using-add-in.html> (up to Excel 2003 version) ; <http://data-mining-tutorials.blogspot.com/2010/08/tanagra-add-in-for-office-2007-and.html> (Excel 2007 and 2010).

² <http://www.openoffice.org/>

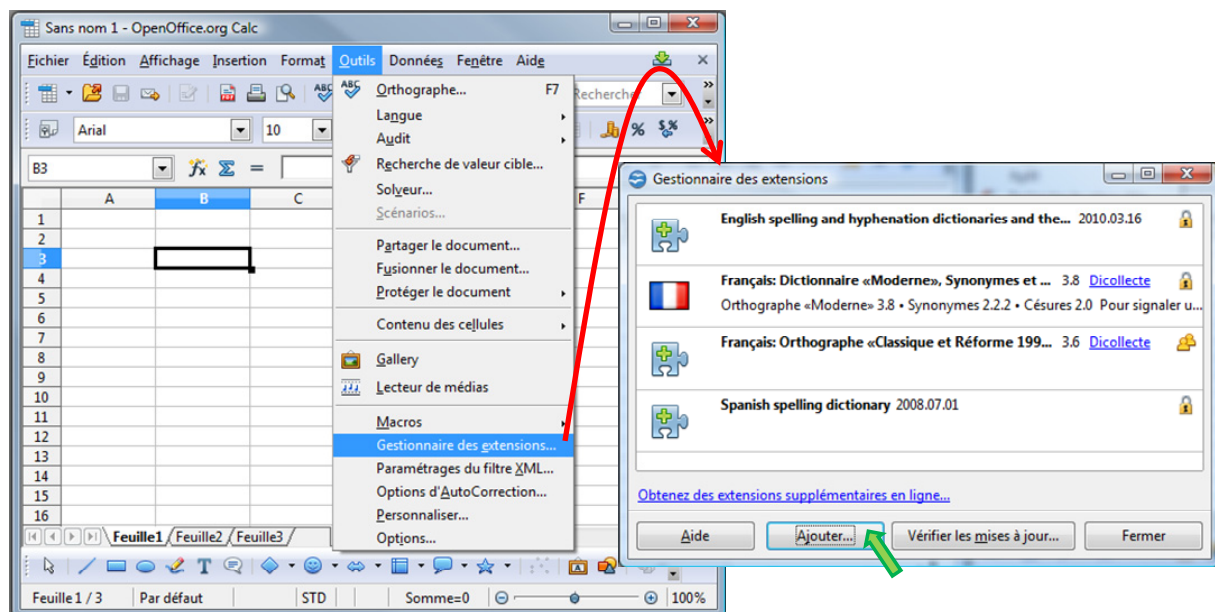
³ <http://www.libreoffice.org/>

2.2 Installing the add-on into OpenOffice Calc

Thereafter, we launch OpenOffice and we select the Calc module⁴.

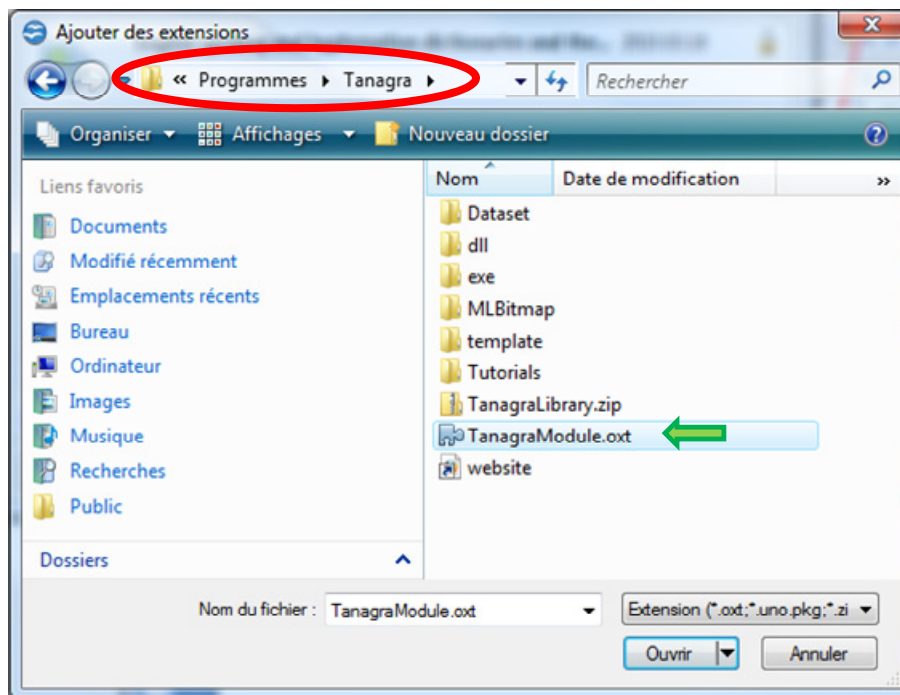


To install the add-on, we activate the Extensions Manager menu (TOOLS menu). Into the dialog settings, we click on the ADD button.

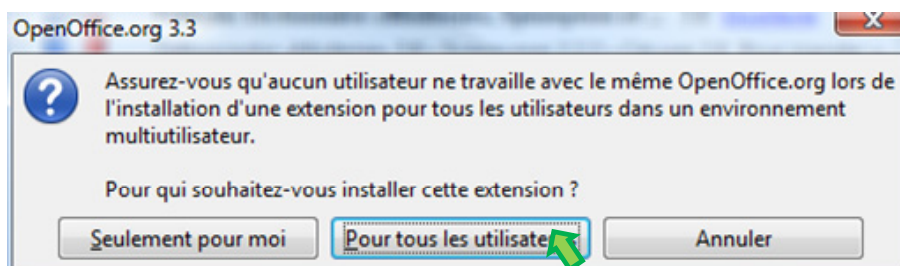


We set the installation directory of Tanagra (usually `C:\Programmes\tanagra` for French Windows Vista, but actually, *it depends on your system configuration*).

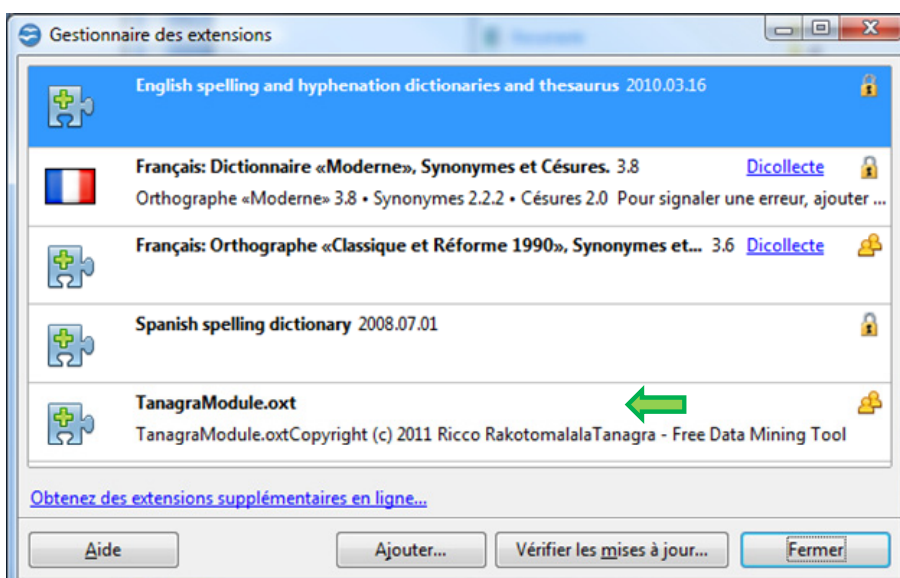
⁴ The screenshots show menus in French, but I hope you will find easily the corresponding menus on your system.



We select "TanagraModule.oxt". A dialog box appears, we activate the add-on for all the users.



Now, the add-on is displayed into the list of extensions.

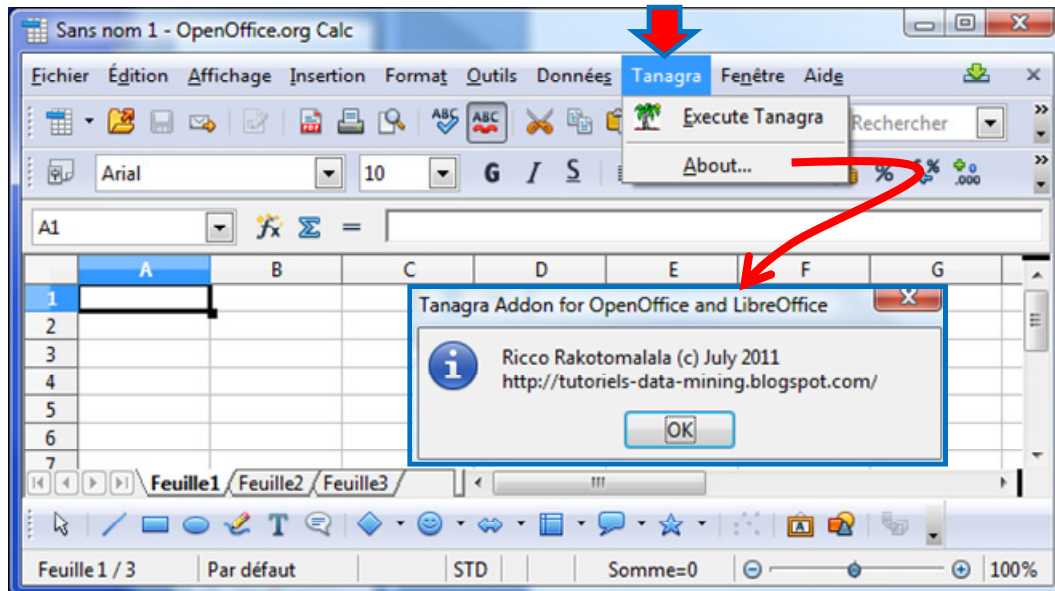


Last, we close the extension manager by clicking on the CLOSE button.

2.3 Activating the add-on

The add-on is not immediately activated after the installation. To do this, **we must restart OpenOffice Calc** i.e. we close, and then we launch again the application. The novelty is that an additional item TANAGRA is available into the menu bar.

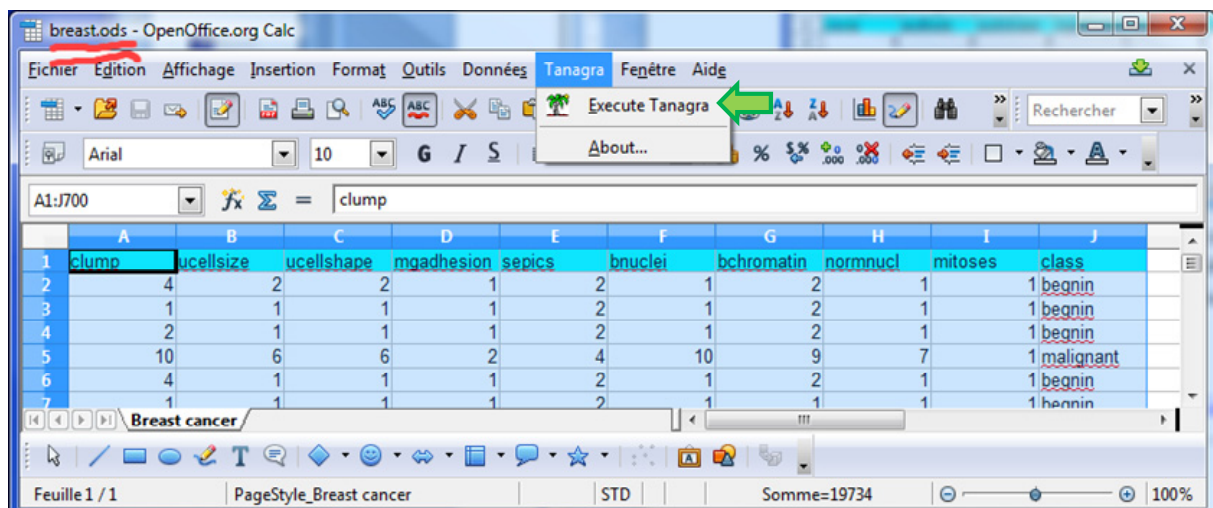
By clicking on the TANAGRA / ABOUT menu, we get the about box of the add-on.



3 Using the add-on

To describe the utilization of the add-on, we use [breast.ods](http://breast.ods.org)⁵. We load the dataset into OpenOffice.

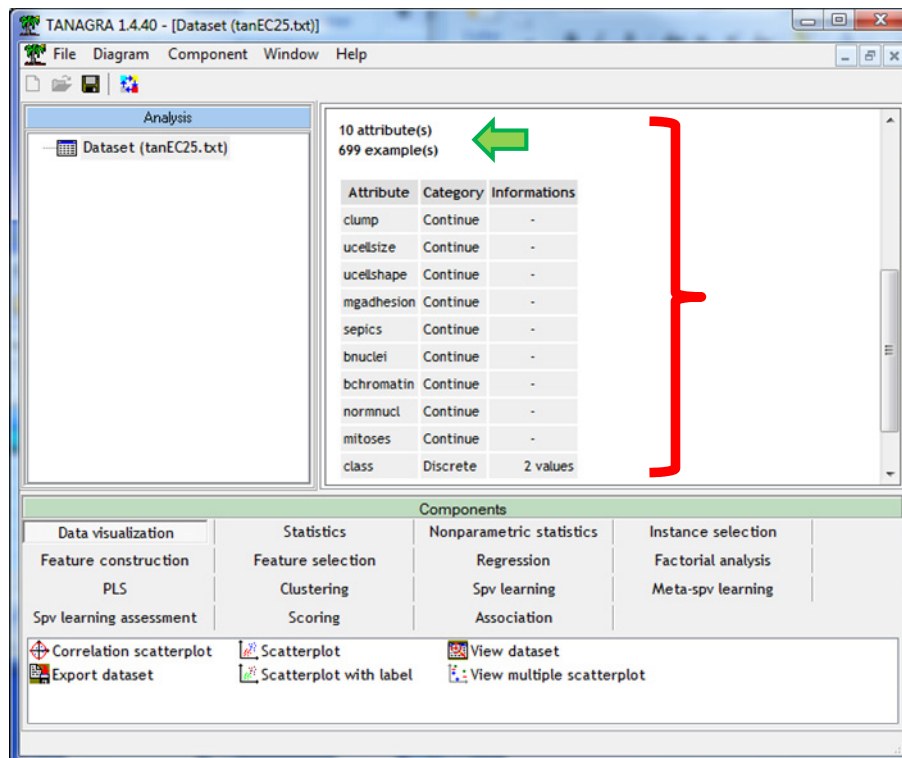
After we select the data range, including the first row corresponding to the name of the variables, we activate the TANAGRA / EXECUTE TANAGRA menu. TANAGRA is automatically launched and the dataset is loaded.



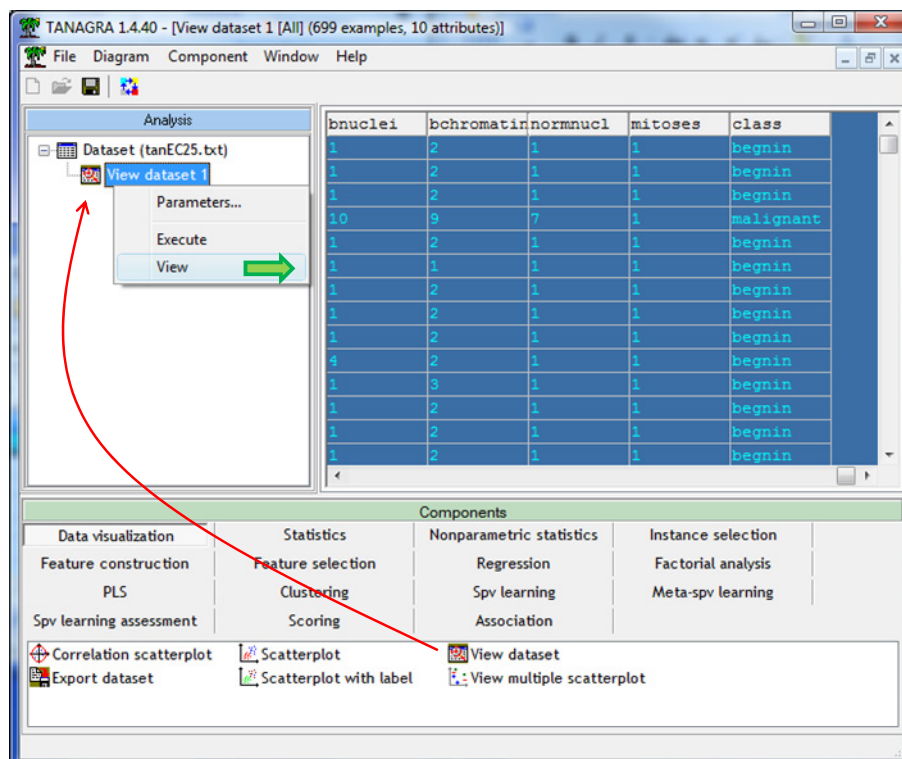
We check the data importation by visualizing the description of variables: we have 9 continuous attributes and 1 discrete attribute.

⁵ Source : <http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29>

There are 699 instances into the data file.



We can visualize the values by using the VIEW DATASET component (DATA VISUALIZATION tab). We insert it into the diagram, and we click on the VIEW contextual menu.



From now, we can launch any kind of data analysis. We have many examples of treatments on our tutorials website <http://data-mining-tutorials.blogspot.com/>.

4 Conclusion

The spreadsheet is a favored tool for data mining. Its capacities for data management are widely recognized⁶. It is one of the tools most used by data miners⁷. The increased capabilities of recent versions (since version 3.0 of OpenOffice Calc: 1,048,576 rows and 1,024 columns) makes it operational in many situations, including in a professional use context.

⁶ <http://www.kdnuggets.com/polls/2008/tools-languages-used-data-cleaning.htm>

⁷ <http://www.kdnuggets.com/polls/2011/tools-analytics-data-mining.html> ; Excel is named, but by extension we can include any kind of spreadsheet applications, of which OpenOffice and LibreOffice.