Tutorial overview

While developing TANAGRA, the underlying objective was to give access to a lot of data mining methods, and not to manage with the numerous formats of dataset files (anyway, it is more the purpose of a commercial software).

TANAGRA loads data from text files with tab separator, built in the following way:

- 1st line : names of attributes

- next lines : values of the attributes for the sample (one line for each record).

N.B: TANAGRA does not manage missing data.

In this tutorial, we show the steps for data importation from an Excel spreadsheet towards TANAGRA. We choose Excel as the original source of data, because it is generally easy to copy the data you're working on, and paste it in an Excel sheet.
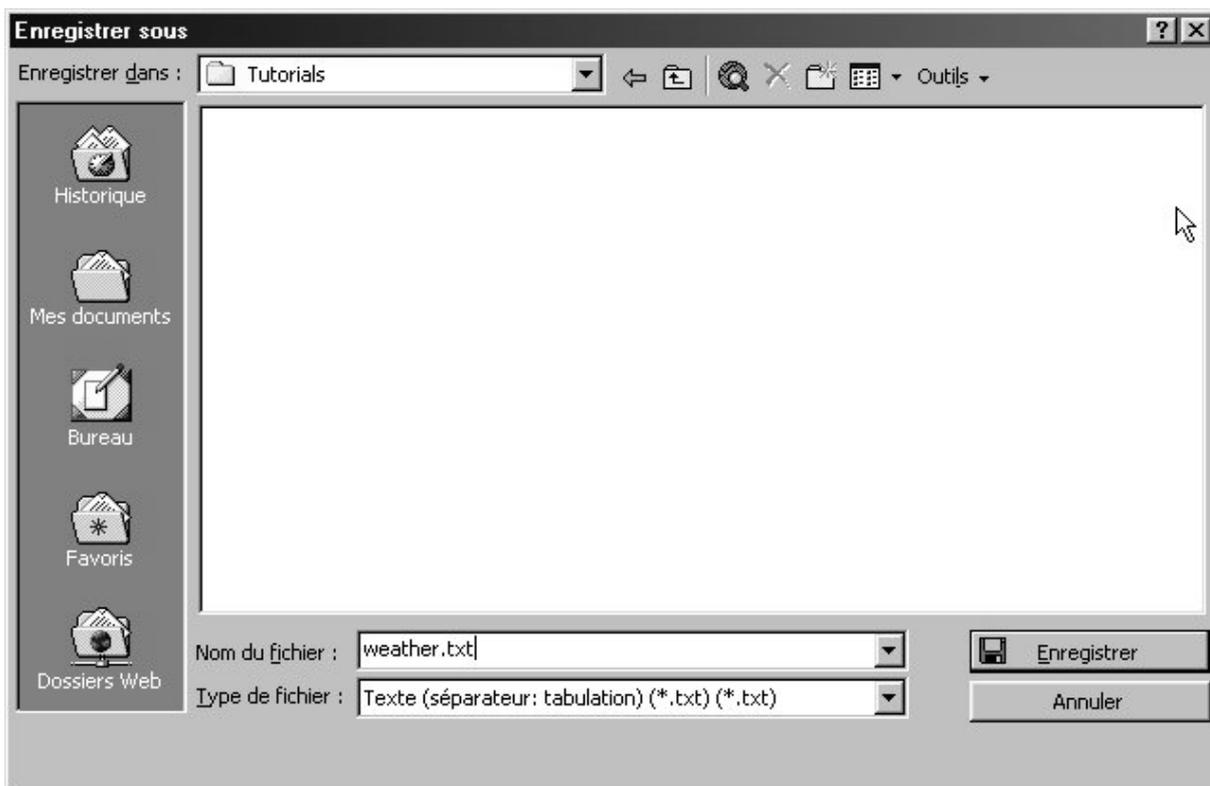
Saving an Excel sheet to TANAGRA format

1 – Launch Excel.
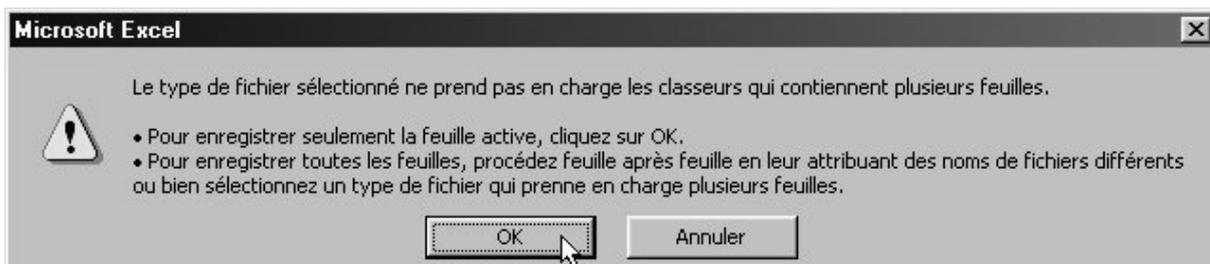
2 – Enter the example data as in the following screenshot.

3 – In the *File* menu, choose the "*Save as…"* command. A dialog box opens, in the "Type of file" list, choose « Text (separator: tabulation) ». Enter « weather.txt » as the name of the file.
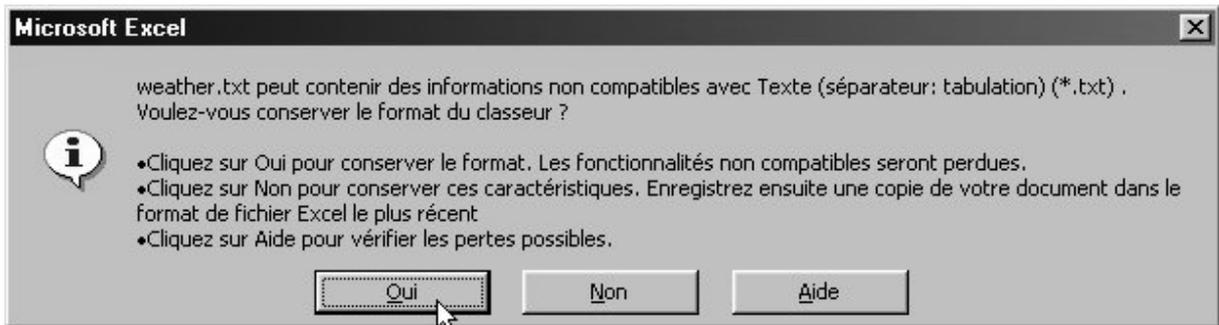

4 – Click on the Save button.

5 – A dialog box advertises you that, in textfile format, only one sheet can be saved at the same time. Click the OK button.
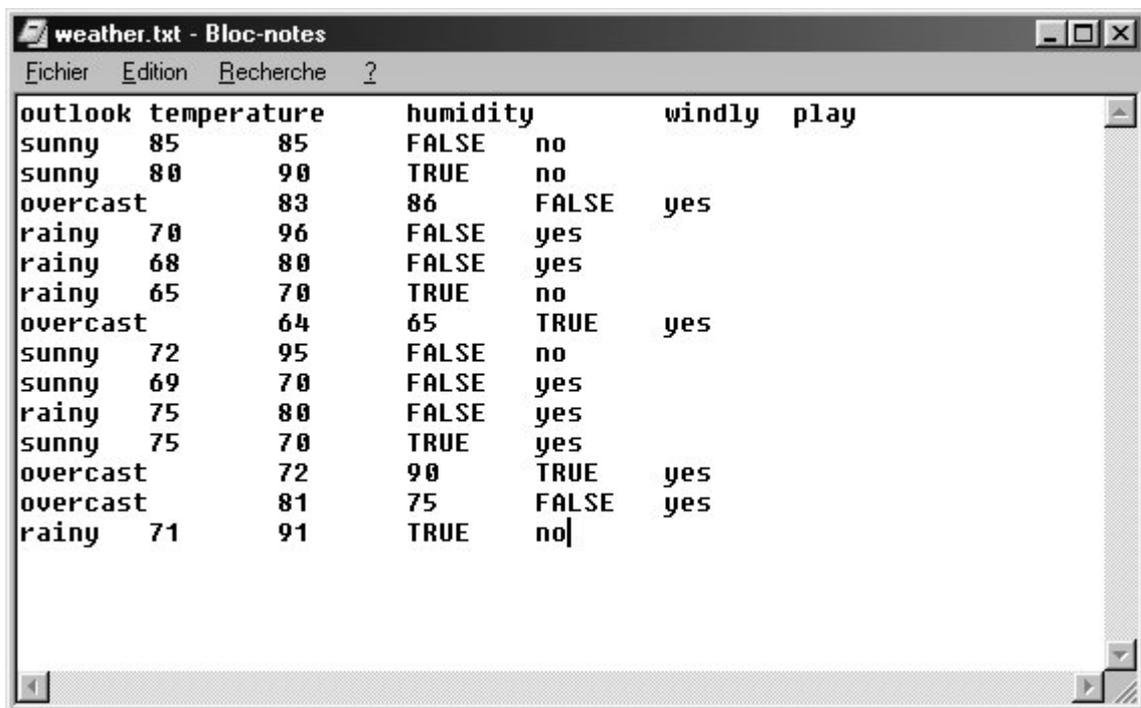

6 – Another dialog box tells you that « weather.txt » can contain data that are incompatible with the text file format. Click on Yes.

Microsoft Excel ☒

ⓘ weather.txt peut contenir des informations non compatibles avec Texte (séparateur: tabulation) (*.txt) . Voulez-vous conserver le format du classeur ?

•Cliquez sur Oui pour conserver le format. Les fonctionnalités non compatibles seront perdues.
•Cliquez sur Non pour conserver ces caractéristiques. Enregistrez ensuite une copie de votre document dans le format de fichier Excel le plus récent
•Cliquez sur Aide pour vérifier les pertes possibles.

| Oui | Non | Aide |

7 – The data are ready to be imported into TANAGRA. Close Excel.

8 – You can visualize the file you've just saved under any text editor. Here is this file « weather.txt », opened in the Windows notepad.

```
weather.txt - Bloc-notes                          _ □ ×
Fichier  Edition  Recherche  ?
outlook  temperature    humidity      windly  play
sunny    85     85       FALSE   no
sunny    80     90       TRUE    no
overcast        83       86      FALSE   yes
rainy    70     96       FALSE   yes
rainy    68     80       FALSE   yes
rainy    65     70       TRUE    no
overcast        64       65      TRUE    yes
sunny    72     95       FALSE   no
sunny    69     70       FALSE   yes
rainy    75     80       FALSE   yes
sunny    75     70       TRUE    yes
overcast        72       90      TRUE    yes
overcast        81       75      FALSE   yes
rainy    71     91       TRUE    no
```
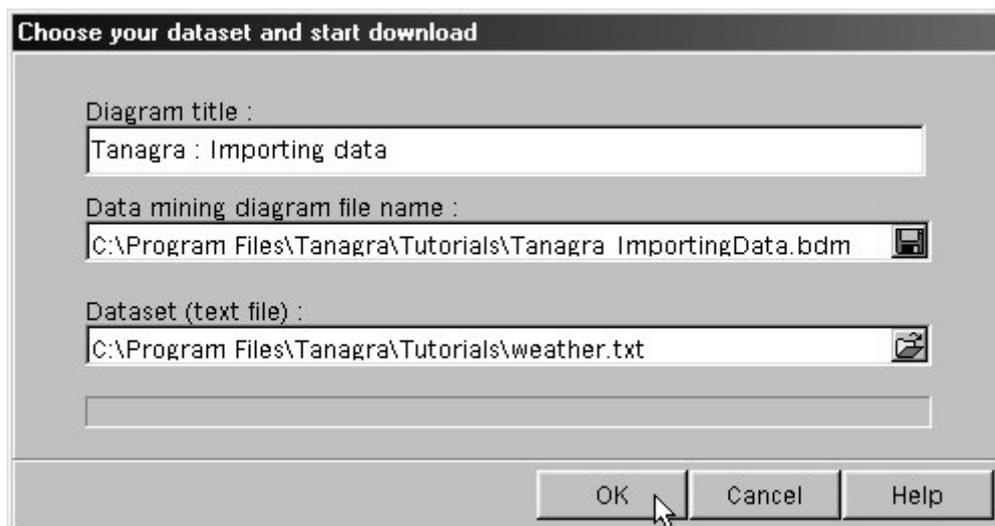
Importing your data in TANAGRA

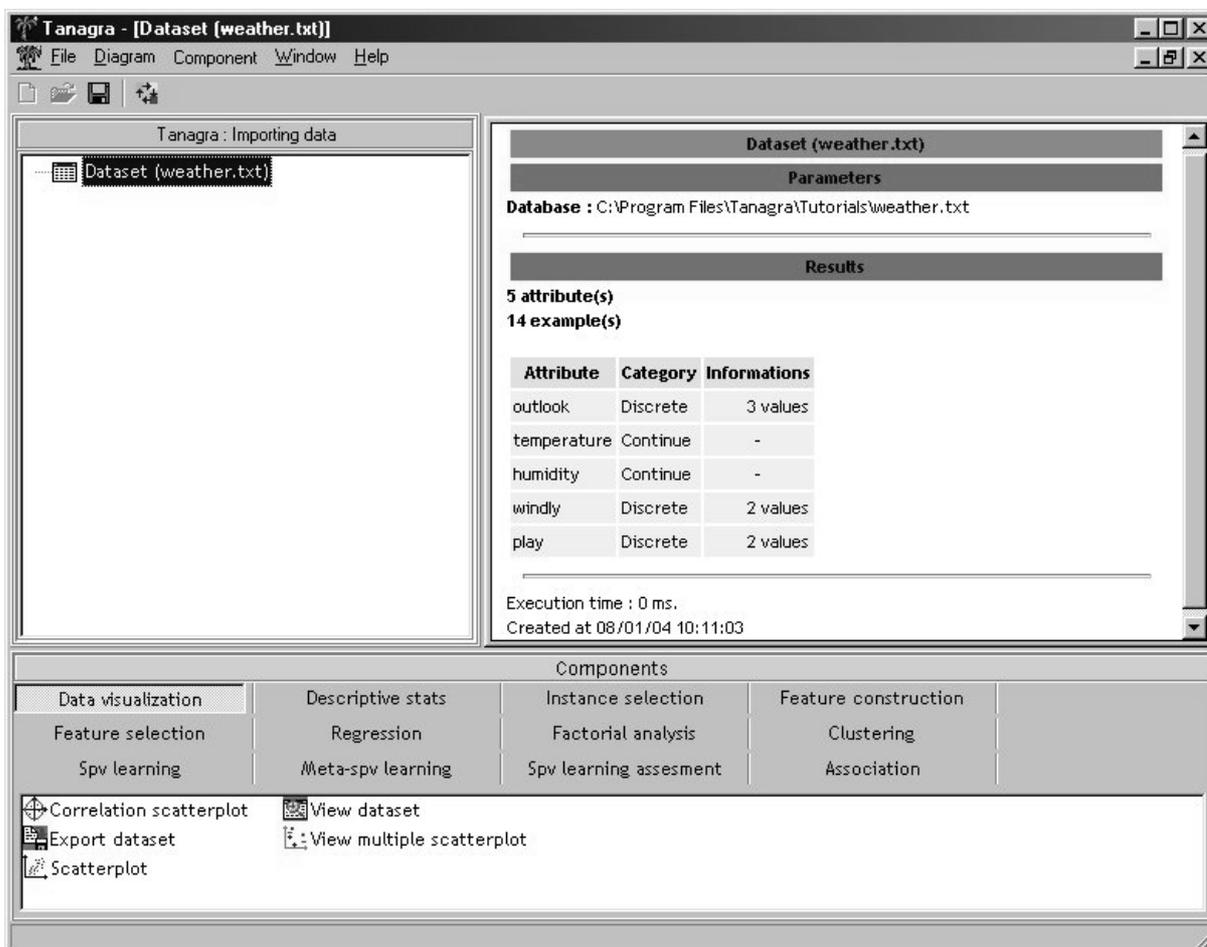1 – Choose "*File/New…*" in the main menu of TANAGRA.

2 – Enter a title for the diagram: « TANAGRA : Importing Data ».

3 – Enter the name of the associated file in which you will save your work (« TANAGRA_ImportingData.bdm   »). Before click on this button 🖫   to run through the hard disk and place yourself in the directory « …\TANAGRA\Tutorials ».

4 – Click on the 📂 icon to seek the file you have created "weather.txt".

5 – Validate with OK to start data importation.

A new diagram is created, based on the file « weather.txt ». You can see the description of its contents in the right frame.

Important precisions on diagrams saving formats

While creating a new diagram, we have used the extension "BDM" for the name of the diagram file (« TANAGRA_ImportingData.bdm »).

BDM (Binary Diagram) is the extension for a TANAGRA file containing, besides the diagram description, the imported dataset (here the data from the file « weather.txt »). So we loose the association to the original data textfile, which can be modified or deleted without impacting the work we done under TANAGRA.

TDM (Text Diagram), the other TANAGRA file format for diagrams, is the one used if, besides the diagram description, you only want to save a **reference** to the data file. So that if you make any change in the data file (here « weather.txt »), you will have different results the next time you execute the sequences of operations in TANAGRA! But be careful not to delete the source file if you want to be able to reopen the diagram.