

Présentation du didacticiel

Lors du développement de TANAGRA, l'accent a été mis sur les méthodes de data mining, et non sur comment importer les différents formats de données du marché (ce qui relève d'ailleurs plus d'un logiciel à vocation commerciale).

TANAGRA charge ses données à partir d'un fichier texte délimité par des tabulations, contenant les données suivantes :

- 1^{ère} ligne : noms des variables
- lignes suivantes : valeurs des variables pour les individus (une ligne par individu).

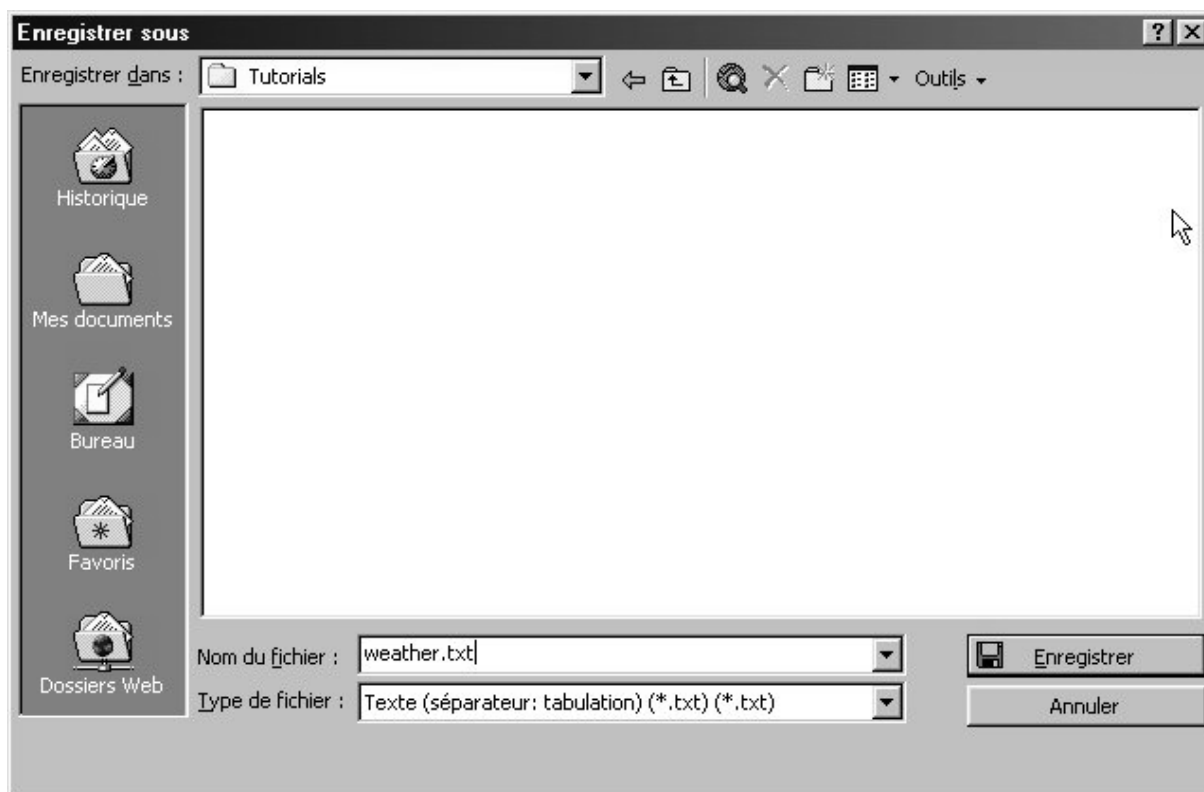
Dans ce didacticiel, nous montrons les étapes à suivre pour importer dans Tanagra les données provenant d'une feuille Excel. Nous avons choisi de montrer l'import depuis Excel, car il est généralement facile de copier/coller ses données dans ce tableur.

Enregistrer une feuille Excel au format Tanagra

- 1 – Lancez Excel.
- 2 – Saisissez les données comme dans la copie d'écran ci-dessous.

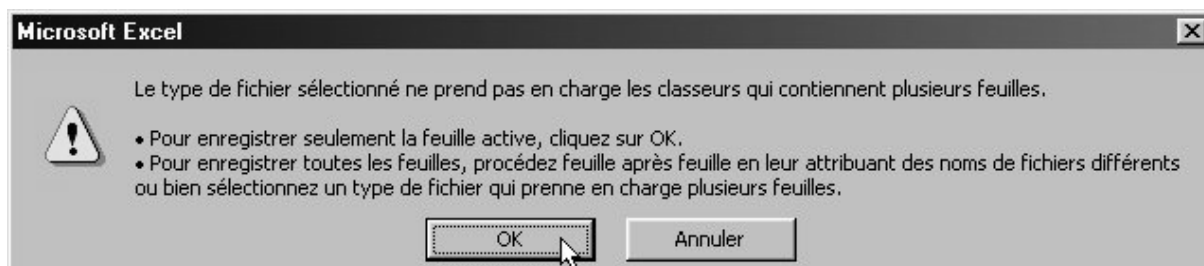
	A	B	C	D	E	F
1	outlook	temperature	humidity	windy	play	
2	sunny	85	85	FALSE	no	
3	sunny	80	90	TRUE	no	
4	overcast	83	86	FALSE	yes	
5	rainy	70	96	FALSE	yes	
6	rainy	68	80	FALSE	yes	
7	rainy	65	70	TRUE	no	
8	overcast	64	65	TRUE	yes	
9	sunny	72	95	FALSE	no	
10	sunny	69	70	FALSE	yes	
11	rainy	75	80	FALSE	yes	
12	sunny	75	70	TRUE	yes	
13	overcast	72	90	TRUE	yes	
14	overcast	81	75	FALSE	yes	
15	rainy	71	91	TRUE	no	
16						
17						

- 3 – Dans le menu *Fichier*, choisissez la commande *Enregistrer sous...*, la boîte de dialogue de sauvegarde apparaît. Dans la liste déroulante « Type de fichier », choisissez « Texte (séparateur : tabulation) ». Saisissez « weather.txt » comme nom du fichier.

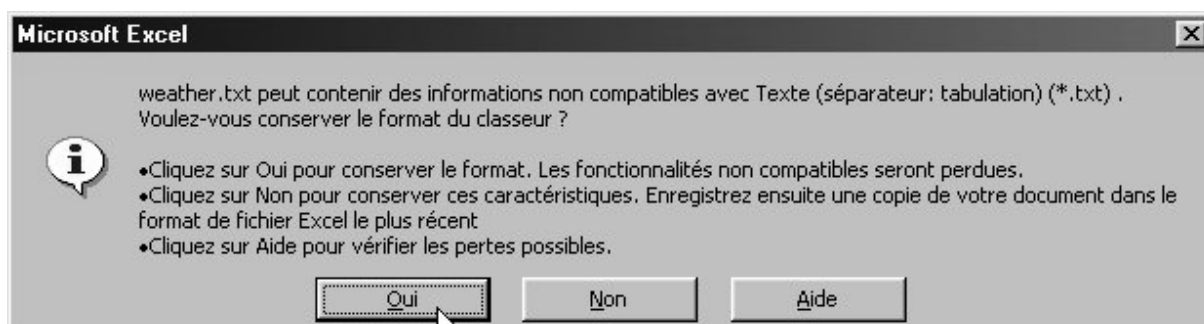


4 – Cliquez sur Enregistrer.

5 – Une fenêtre de dialogue vous avertit que « Le type de fichier sélectionné ne prend pas en charge les classeurs contenant plusieurs feuilles ». Cliquez sur OK.



6 – Une autre fenêtre vous avertit que « weather.txt peut contenir des informations non compatibles avec Texte (séparateur : tabulation) ». Cliquez sur Oui.



7 - Votre fichier est prêt à être importé sous Tanagra. Fermez Excel.

8 – Vous pouvez visualiser le fichier que vous venez d'enregistrer sous n'importe quel éditeur texte. Voici par exemple le fichier « weather.txt » que vous venez de créer, dans le bloc-notes Windows.


```

weather.txt - Bloc-notes
Fichier  Edition  Recherche  ?
outlook temperature  humidity  windy  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
    
```

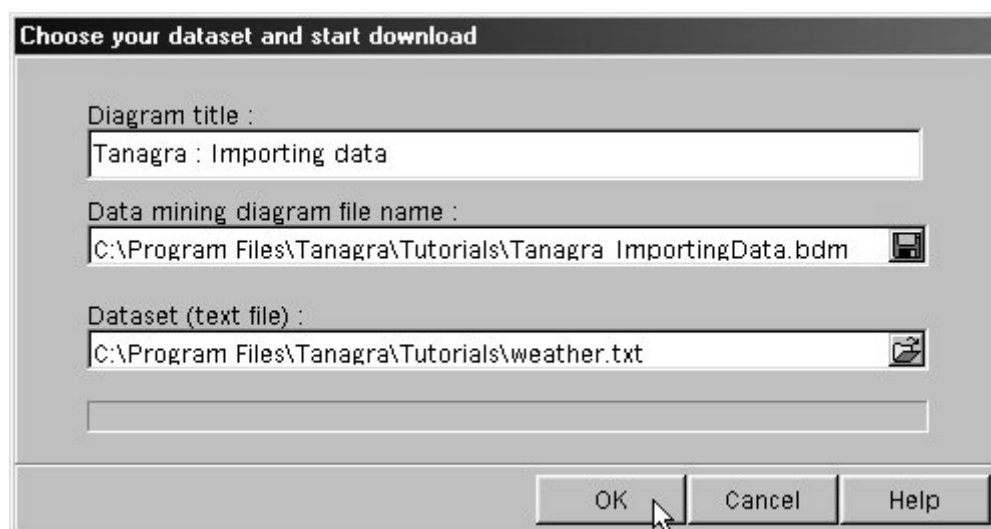
Importer ses données dans Tanagra

1 – Choisissez *File/New...* dans le menu de Tanagra.

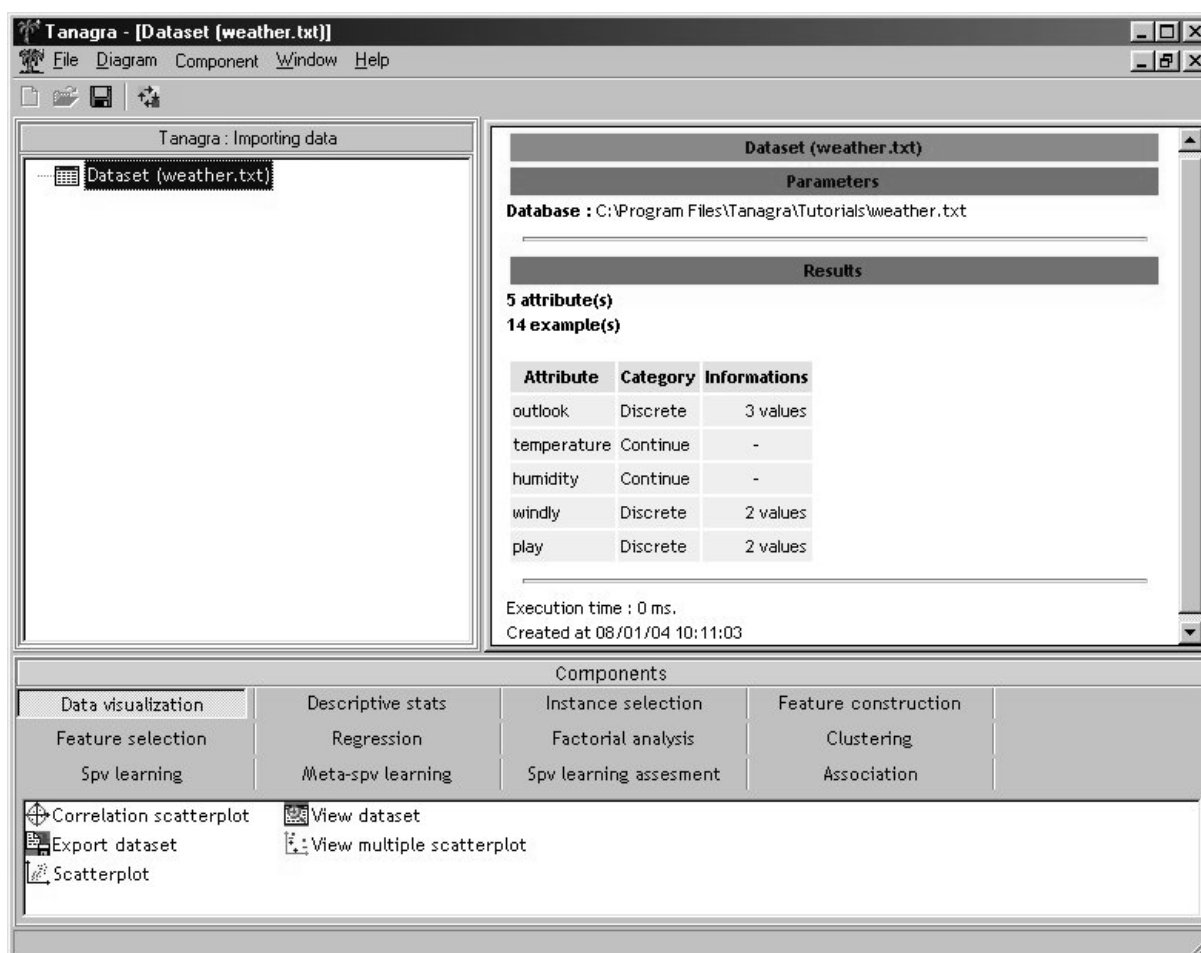
2 – Entrez un titre pour le diagramme : « Tanagra : Importing Data ».

3 – Entrez le nom du fichier associé dans lequel votre travail sera sauvegardé (« Tanagra_ImportingData.bdm »). Cliquez avant sur le bouton  pour parcourir le disque dur et vous placer dans le répertoire « ...\Tanagra\Tutorials ».

4 – Allez chercher le fichier weather.txt que vous venez de créer en cliquant sur l'icône



5 – Cliquez sur OK pour débiter l'importation.



Un nouveau diagramme a été créé, sur la base du fichier « weather.txt ». Dans le cadre de droite, vous voyez sa description.

Note importante sur les formats de sauvegarde des diagrammes

Dans la fenêtre de création d'un nouveau diagramme, nous avons sauvegardé sous le nom « Tanagra_ImportingData.bdm ».

BDM (Binary Diagram) correspond à un fichier contenant, outre la description du diagramme, les données importées (ici celles du fichier « weather.txt »). On s'affranchit ainsi de la connexion au fichier texte, qu'on peut modifier ou supprimer sans affecter le travail effectué sous Tanagra.

TDM (Text Diagram), l'autre format proposé, correspond à un fichier ne contenant que la description du diagramme et une **référence** sur le fichier de données. De sorte que si vous mettez à jour le fichier texte source (ici « weather.txt »), vous aurez des résultats différents quand vous réouvrirez votre diagramme dans Tanagra ! Mais vous ne devez pas supprimer ce fichier texte si vous voulez pouvoir rouvrir le diagramme.