

Les données structurées et leur traitement TP1

Traitement des données avec un tableur

TP décrit pour le tableur Excel.

Toutes les manipulations à faire sur tableur sont précédées du symbole •.

Contexte.

On dispose d'une base de données sur des chansons décrites par les huit *descripteurs* suivants : titre ; interprète ; auteur ; langue ; album ; genre ; année ; durée.

Le but de ce TP est d'apprendre à utiliser les fonctionnalités d'un tableur pour répondre à des *requêtes*.

Une requête est une demande d'informations sur la base de données.

Exemples.

La liste des chansons d'un album ou d'un interprète.

La liste des interprètes de la base de données.

La liste des chansons dont la durée ne dépasse pas 5 minutes

Etc...

Une fois que l'information répondant à la requête est extraite de la base de données, il s'agit de l'organiser de manière pertinente.

I. Demande de détails sur un objet ciblé.

Dans cette partie, on s'intéresse aux requêtes qui portent sur un objet ciblé :

Les chansons d'un album ou d'un interprète en particulier.

Les chansons portant un titre en particulier (reprises, versions live ou studio).

- Récupérer le fichier fourni : « mes_chansons.ods », puis ouvrez-le. On observe une table décrivant des chansons avec 8 *descripteurs* : titre ; interprète ; auteur ; langue ; album ; genre ; année ; durée.

1. Avec des tris : Menu Données / Trier

a. La liste des chansons de l'album « The dark side of the moon » de Pink Floyd.

- Commencez par renommer une feuille ou en insérer une nouvelle avec le nom : « Albums ».
- Retournez dans la feuille « base de données ».
- Dans le menu **Données**, choisissez **Trier**. (ou sur accueil trier/filtrer)
- Faire un **tri personnalisé** et choisir « album » dans **Trier par**.
- Retrouvez alors l'album souhaité, puis faites un Copier/Coller de la plage de données qui nous intéresse vers la feuille « Albums ». (Copiez les huit colonnes, pas seulement la colonne des titres).

Peut-être serait-il souhaitable de retrouver les descripteurs (en-têtes des colonnes) dans notre nouvelle feuille ?

- En insérant éventuellement une ligne au-dessus de la table et en utilisant le Copier/coller adéquat, faites apparaître ces descripteurs au-dessus de la table.

vous devez obtenir quelque chose qui ressemble à ça .

	A	B	C	D	E	F	G	H
1	titre	interprète	auteur	langue	album	genre	année	durée (mn : sec)
2	Time	Pink Floyd	David Gilmour / Nick Mason / Roger Waters / Richard Wright	Anglais	The Dark Side of the Moon	Pop/Rock	1973	06:53
3	On the Run	Pink Floyd	David Gilmour / Roger Waters	Anglais	The Dark Side of the Moon	Pop/Rock	1973	03:45
4	Speak to Me	Pink Floyd	Nick Mason	Anglais	The Dark Side of the Moon	Pop/Rock	1973	01:07
5	The Great Gig in the Sky	Pink Floyd	Clare Torry / Richard Wright	Anglais	The Dark Side of the Moon	Pop/Rock	1973	04:44
6	Breathe(in the air)	Pink Floyd	David Gilmour / Roger Waters / Richard Wright	Anglais	The Dark Side of the Moon	Pop/Rock	1973	02:49
7	Any Colour you Like	Pink Floyd	David Gilmour / Nick Mason / Richard Wright	Anglais	The Dark Side of the Moon	Pop/Rock	1973	03:26
8	Us and Them	Pink Floyd	Roger Waters / Richard Wright	Anglais	The Dark Side of the Moon	Pop/Rock	1973	07:49
9	Brain Damage	Pink Floyd	Roger Waters	Anglais	The Dark Side of the Moon	Pop/Rock	1973	03:46
10	Money	Pink Floyd	Roger Waters	Anglais	The Dark Side of the Moon	Pop/Rock	1973	06:23
11								

b. La liste des chansons de l'interprète « Barbara ».

- Procédez de la même manière, dans une nouvelle feuille nommée « interprètes », pour répondre à la requête suivante :
Les chansons de l'interprète « Barbara », triées selon la date de parution.

2. Avec des filtres : Données / Filtre / AutoFiltre

a. La liste des chansons de l'album « Puta's Fever » de la Mano Negra.

- Retournez dans la feuille « Base de données ».
- Dans le menu **Données**, choisissez **Filtres**
- Choisissez l'album souhaité dans la colonne album.
- Faites un Copier/Coller des données obtenues (en prenant aussi la ligne N°1 des descripteurs) vers la feuille « Albums », quelques lignes dessous de l'album précédent.
- De retour sur la feuille « Base de données », pour retrouver l'ensemble des données, cliquez sur la flèche à coté de « Album » puis sélectionner tout.

b. Listes des chansons pour un titre donné.

- Procédez de la même manière, dans une nouvelle feuille nommée « titres », pour répondre aux deux requêtes suivantes :
Les chansons dont le titre est « Ma plus belle histoire d'amour ».
Les chansons dont le titre est « Eleanor Rigby ».

Vous séparerez bien les deux tables par quelques lignes.

3. Les redondances.

**Une redondance est une répétition inutile d'une même information.
Il est important d'éliminer les redondances.**

Dans la table de la feuille « Titres », on observe des redondances pour les descripteurs :
Titre, auteur, langue.

Un titre est donc décrit par trois descripteurs : nomTitre, Auteur, langue.

On peut donc supprimer les colonnes correspondantes et ne faire apparaître ces informations qu'une seule fois en en-tête de la table.

- reprenez les opérations nécessaires pour obtenir dans la feuille « Titres » quelque chose qui, avec un petit travail de mise en forme, ressemble à ça.

	A	B	C	D	E	F
1	titre	Ma plus belle histoire d'amour	auteur	Barbara	langue	Français
2						
3		interprète	album	genre	année	durée
4		Marc Ducret	Monk to bach	Jazz	1996	06:42
5		Barabara	Bobino 1967	variété	1967	04:02
6		Barabara	Ma plus belle histoire d'amour	variété	1967	04:50
7						
8						
9	titre	Eleanor Rigby	auteur	Paul McCartney / John Lennon	langue	Anglais
10						
11		interprète	album	genre	année	durée
12		The Beatles	Revolver	Pop/Rock	1966	02:06
13		Joan baez	Joan	Folk	1967	02:18
14		Sarah Vaughan	Songs of the Beatles	R&B	1981	03:48
15		Aretha Franklin	This girl's in love with you	soul	1970	02:38
16		Joe Jackson	Summer in the city	Pop/Rock	2000	03:12
17		Thrice	If we could only see us now	Pop/Rock	2005	03:51
18						

- Procédez de la même manière avec la feuille « Albums » pour identifier les descripteurs d'un album, et réorganiser les tables en conséquence.

II. Demande d'une liste d'objets

Dans la partie précédente, les requêtes demandaient une liste de chansons répondant à certains critères.

Dans cette partie, on s'intéresse à des requêtes qui ne demandent pas une liste de chansons, mais une liste d'objets présents dans la base de données :

Liste des albums ou liste des titres présents dans la base de données par exemple.

1. La liste des albums.

- Insérez une nouvelle feuille que vous nommerez « listeAlbums ».

On souhaite y créer les listes des albums et des titres avec uniquement leurs descripteurs.

- Dans la base de données, sélectionnez les colonnes Album, date, interprète.
Pour sélectionner des plages de données non adjacentes, appuyez sur la touche ctrl à chaque nouvelle sélection.
- Faites alors un Copier/Coller vers votre nouveau tableur.

On obtient bien la liste des albums avec leurs descripteurs, mais le résultat n'est pas satisfaisant à cause de la présence de **doublons**.

**Dans une table de données, des doublons sont des lignes identiques
Il est important de supprimer les doublons d'une table.**

Suppression des doublons

- Sélectionnez les données de la colonne Albums puis dans le menu Données choisissez :
Filtrer ► Avancé
Cochez « extraction sans doublon »
Cliquez sur OK.
- Renommez et réorganisez les colonnes pour avoir dans l'ordre :
NomAlbum, Interprète, genre, année.

vous devez obtenir quelque chose qui ressemble à ça .

	A	B	C	D
1	Nom Album	interprète	genre	année
2	This girl's in love with you	Aretha Franklin	soul	1970
3	Barbara chante Brassens	Barbara	variété	1960
4	Bobino 1967	Barbara	variété	1967
5	Ma plus belle histoire d'amour	Barbara	variété	1967
6	Le Vent	Georges Brassens	variété	1953
7	non Homologué	J.J. Goldman	Pop/Rock	1985
8	Joan	Joan baez	Folk	1967
9	Summer in the city	Joe Jackson	Pop/Rock	2000
10	Putas's Fever	Mano Negra	Rock alternatif	1989
11	Monk to bach	Marc Ducret	Jazz	1996
12	nina simone sings the blues	nina simone	blues	1967
13	The Dark Side of the Moon	Pink Floyd	Pop/Rock	1973
14	Wish you where here	Pink Floyd	Pop/Rock	1975
15	Songs of the Beatles	Sarah Vaughan	R&B	1981
16	1993...J'appuie sur la gâchette	Suprême NTM	Rap	1993
17	Revolver	The Beatles	Pop/Rock	1966
18	If we could only see us now	Thrice	Pop/Rock	2005
19				

2. Liste des titres.

- Procédez de la même manière pour obtenir :
Une feuille « listeTitres » avec, dans l'ordre, les descripteurs : NomTitre, Auteur, langue.

III. Une nouvelle structure de données

On cherche ici à restructurer notre base de données.

Dans le sens où on cherche à éviter au maximum les redondances d'information, on voit que notre structure initiale (le contenu du fichier initial) n'est pas satisfaisante car pleine de redondances. En effet, les informations sur l'année et l'interprète d'un album pourraient n'apparaître qu'une seule fois.

Tout comme les informations sur l'auteur et la langue d'un titre donné.

- On va travailler ici dans un nouveau classeur que vous nommerez « nouvelle_structure ». Dans le menu **Fichier**, choisissez **Nouveau ► Classeur**. Sauvegardez tout de suite le classeur pour le renommer.

Puisque l'objectif est d'éliminer les redondances d'information, nous allons partir des deux dernières tables : « listeAlbums » et « listeTitres », qui ne contiennent justement aucune redondance.

- Copiez les données des deux feuilles « listeAlbums » et « listeTitres » et copiez-les dans deux feuilles de notre nouveau classeur que vous nommerez « Albums » et « Titres ».

On souhaite maintenant nettoyer la liste des chansons de la base initiale de toutes les informations déjà contenues dans ces deux tables.

- Insérez une nouvelle feuille nommée : « listeChansons ».

Pour retrouver une chanson, on n'a besoin que de son **titre** et de son **album**.

De plus, la seule information sur une chanson qui ne soit pas déjà dans les deux tables est sa **durée**.

- Dans la base de données initiale, sélectionnez les colonnes titres, album, durée. Faites alors un Copier/Coller vers la feuille « listeChansons » du classeur « nouvelle_structure ».

Avec ces trois tables, on obtient une nouvelle structure de notre base de données.

QUESTIONS.

1. Pas de perte d'information.

Justifiez que si on se donne une chanson, identifiée par son titre et son album, on peut retrouver toutes les informations que nous donnait la table initiale.

2. Elimination des redondances.

Avec la base de données initiale, on avait 130 chansons et 8 descripteurs soit $8 \times 130 = 1040$ valeurs à stocker.

Combien de valeurs contient notre nouvelle structure ?

IV – Pour aller plus loin : le format csv

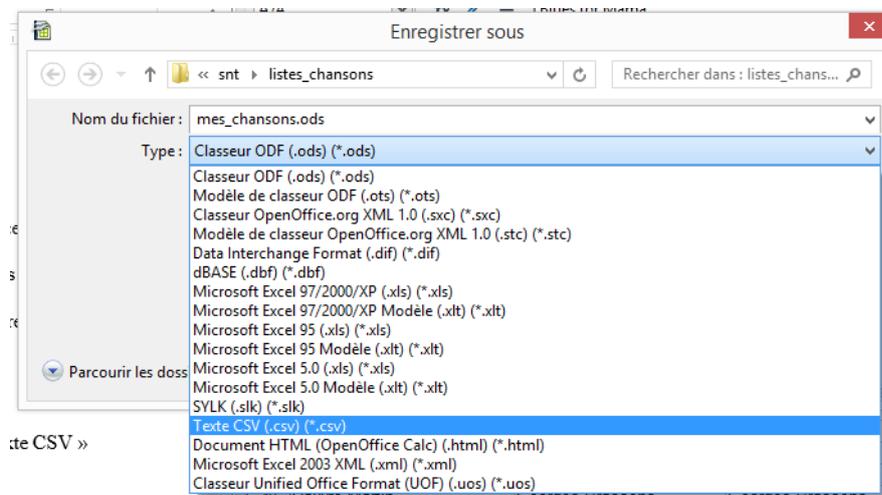
Un fichier au format .ods ou xls est une feuille de calcul qui n'est lisible que par des logiciels spécifiquement dédiés.

En particulier, la plupart des langages de programmation ne permettent pas de lire ce type de fichier.

En revanche, il existe un format respectant une norme universelle, lisible à travers tous les langages de programmation : le format .csv.

Il s'agit d'un fichier texte conçu pour stocker des données. Chaque donnée étant séparée par un virgule (un point-virgule dans certains cas).

- Ouvrez le fichier « mes_chansons.ods ».
- Dans le menu « Fichier » choisissez « enregistrer sous ».
- Dans la fenêtre qui s'affiche dans le menu déroulant « Type », choisissez « CSV séparateur point virgule »



• Cliquez sur « Enregistrer ». Vous devez maintenant avoir le fichier « mes_chansons.csv ». Il peut s'ouvrir avec Excel et vous obtiendrez une feuille de calcul habituelle.

- Avec un clic droit sur le fichier, choisissez « Ouvrir avec » puis l'éditeur de texte qui vous est proposé : « Bloc-note » très probablement.
On a bien un fichier texte dans lequel on retrouve les mêmes lignes que dans la feuille de calcul, mais dans chaque ligne, les valeurs sont séparées par des virgules.
- Vous pouvez fermer votre fichier.