

Thème 4 : les données structurées et leur traitement

Activité 1 : données structurées et format de données

Contenus	Capacités attendues
Données	Identifier les principaux formats et représentations de données.
Données structurées	Identifier les différents descripteurs d'un objet. Distinguer la valeur d'une donnée de son descripteur. Utiliser un site de données ouvertes, pour sélectionner et récupérer des données.
Métadonnées	Retrouver les métadonnées d'un fichier personnel.

1. Introduction

Les données constituent la matière première de toute activité numérique. Afin de permettre leur réutilisation, il est nécessaire de les conserver de manière persistante. Les structurer correctement garantit que l'on puisse les exploiter facilement pour produire de l'information. Cependant, les données non structurées peuvent aussi être exploitées, par exemple par les moteurs de recherche.

2. Les formats de données

Répondez aux questions ci-dessous en vous aidant du documents ressources « Ressources-Formats_données.pdf ».

Quels sont les trois principaux formats de données utilisés pour la construction de bases de données ?

.....

.....

Retrouvez l'extension de chacun des extraits des fichiers ci-dessous :

```
Nom,Prénom,Ville,Département,Date de naissance,Temps passé sur des écrans dans la journée en %
Durand,Jean-pierre,L'Aigle,61,23/05/85,"20,00 %"
Tertra,Henry,Verneuil sur avres,27,15/12/67,"10,00 %"
Dupont,Christophe,Rai,61,12/06/78,"5,00 %"
```

Fichier :

```
},
{
  "Nom": "Tertra",
  "Prénom": "Henry",
  "Ville": "Verneuil sur avres",
  "Département": 27,
  "Date de naissance": "15/12/67",
  "Temps passé sur des écrans dans la
  journée en %": "10,00 %"
},
```

Fichier :

```
<row>
  <Nom>Tertra</Nom>
  <Prénom>Henry</Prénom>
  <Ville>Verneuil sur avres</Ville>
  <Département>27</Département>
  <Date de naissance>15/12/67</Date de naissance>
  <Temps passé sur des écrans dans la journée en %>10,00 %</Temps passé sur des
  écrans dans la journée en %>
</row>
```

Fichier :

Observez le fichier ci-dessous.

```
BEGIN:VCARD
VERSION:3.0
FN:Arnaud NEAUNO
N:NEAUNO;Philippe;;;
item1.EMAIL;TYPE=INTERNET:ar.neauno@phrii.fr
item1.X-ABLabel:Domicile
TEL;TYPE=CELL:06 06 00 00 00
TEL;TYPE=HOME:01 00 00 00 78
PHOTO:https://k2.org/support/uploads/
monthly_2019_04/
icon.png.8adf148bdc489d80467cf450c6d37d22.png
END:VCARD
```

Quel est le format de ce fichier ?

Rajoutez l'extension qui termine le nom du fichier.

Arnaud.neauno.....

Quelle est l'utilité de ce type de fichier ?

.....

.....

.....

.....

3. Etude de cas : « les Youtube awards »

Vous devez établir les YouTube Awards des meilleures chaînes YouTube. Pour cela, vous disposez d'un fichier tableur dans lequel vous devez classer les chaînes selon différents critères.

a) Le fichier contenant les données brutes « *BaseDonneesYoutubeurs.txt* » est à aller chercher sur le réseau et copier dans votre répertoire personnel.

b) Visualiser des données au format .txt.

Ouvrir le fichier avec le bloc-notes. Les données sont-elles faciles à lire ?

.....
Quels descripteurs contient ce fichier (ce sont les informations de la 1ère ligne) ?

.....
.....
.....
c) Visualiser les données au format .csv.

Enregistrez le fichier au format « csv » et ouvrez le avec LibreOffice. Les données sont-elles plus faciles à lire ?

.....
.....
.....
Pour chaque descripteur donnez un exemple de valeur associée :

Descripteur	Exemples de valeurs associées

d) Compléter la fiche « Youtube AWARDS » (page 3).

Vous disposez de 4 fiches d'aide sur les opérations de filtrage nécessaires pour définir le classement pour chaque item analysé.

Youtube AWARDS

PLUS GRAND NOMBRE D'ABONNÉS MONDIAUX (MANUEL)

Vous allez vous chronométrer pour chercher chacun des éléments suivants.

« 3 Plus grand nombre d'abonnés » :

- 1^{er} mondial :
- 2^{ème} mondial :
- 3^{ème} mondial :

Temps mis pour trouver :
..... secondes

PLUS GRAND NOMBRE D'ABONNÉS MONDIAUX (TRI AUTOMATIQUE) (fiche 1)

Vous allez vous chronométrer pour chercher chacun éléments suivants, à l'aide du tri automatique et de la fiche 1.

« 3 Plus grand nombre d'abonnés » :

- 1^{er} mondial :
- 2^{ème} mondial :
- 3^{ème} mondial :

Temps mis pour trouver :
..... secondes

Est-ce qu'on a bien retrouvé les mêmes résultats qu'en manuel ?

Et que pensez-vous du temps mis pour trouver les bonnes informations grâce au tri des données ?

LISTE DES PAYS NOMINÉS : (fiche 2)

- Pays 1 :
- Pays 2 :
- Pays 3 :
- Pays 4 :
- Pays 5 :
- Pays 6 :

CHAÎNES FRANÇAISES PAR NOMBRE DE VUES : (fiche 3)

- 1^{er} :
- 2^{ème} :
- 3^{ème} :

PLUS VIEILLES CHÂÎNES DU CLASSEMENT : (fiche4)

- 1^{er} :
- 2^{ème} :
- 3^{ème} :

PLUS GRAND NOMBRE DE VUES PAR ABONNÉS :

- 1^{er} :
- 2^{ème} :
- 3^{ème} :

PLUS GRAND NOMBRE D'ABONNÉS EN MOYENNE PAR AN :

- 1^{er} :
- 2^{ème} :
- 3^{ème} :

CHAÎNES DE MUSIQUE AYANT LE PLUS GRAND NOMBRE DE VUES PAR ABONNÉS :

- 1^{er} :
- 2^{ème} :
- 3^{ème} :

CHAÎNES DE MUSIQUE INDIENNES POSSÉDANT LE PLUS GRAND NOMBRE DE VUES PAR AN :

- 1^{er} :
- 2^{ème} :
- 3^{ème} :