NSI · SECONDE / COURS

Les données structurées

Astrolab Cours · astrolabcours.fr

01 Le problème

Ton lycée enregistre des milliers d'élèves, de notes, d'absences. Comment ranger toutes ces informations pour retrouver en une seconde la moyenne d'un élève ou la liste des absents du lundi ?

01 L'idée : le fichier à fiches cartonnées

Avant les ordinateurs, les bibliothèques rangeaient leurs livres sur des fiches cartonnées. Chaque fiche avait les mêmes cases : titre, auteur, année, genre. Toutes les fiches ensemble formaient le fichier. On pouvait les trier par auteur ou sortir toutes celles d'une année. Une table de données, c'est ce fichier.

01 Données, descripteurs, collection

Une **collection** est un ensemble d'objets de même nature (les livres, les élèves). Chaque objet est décrit par des **descripteurs** (titre, auteur, année), qui prennent des **valeurs**. On range tout dans une **table** : une ligne par objet, une colonne par descripteur.

titre	auteur	année
Les Misérables	Hugo	1862
L'Étranger	Camus	1942

Le format CSV

Un fichier **CSV** est un simple fichier texte où chaque ligne est un objet, et les valeurs sont séparées par une virgule ou un point-virgule. La première ligne donne souvent le nom des descripteurs.

```
titre;auteur;annee
Les Misérables;Hugo;1862
L'Étranger;Camus;1942
```

Trier, filtrer, croiser

- **Trier** : ranger les lignes selon un descripteur (par année croissante).
- **Filtrer** : ne garder que les lignes qui vérifient une condition (les livres d'avant 1900).
- **Croiser** : rapprocher deux tables qui partagent un descripteur. Croiser des données apparemment anodines peut permettre d'identifier une personne.

Les métadonnées

© Astrolab Cours · astrolabcours.fr/nsi/seconde/donnees-structurees/ · Tous droits réservés : usage personnel uniquement, republication et usage commercial interdits

Les **métadonnées** sont des données qui décrivent d'autres données : la date et le lieu d'une photo, l'auteur d'un document, la taille d'un fichier.

Où sont stockées les données ?

Dans le **cloud**, les données sont stockées dans des **centres de données** : des bâtiments remplis de serveurs, qui consomment beaucoup d'électricité, surtout pour les refroidir. Stocker n'est donc pas sans impact sur l'environnement.

Exemple guidé

Voici le début du fichier `absences.csv` du lycée.

```
nom,classe,jour,heures
Sami,2nde3,lundi,2
Léa,2nde1,mardi,1
Inès,2nde3,lundi,4
Tom,2nde2,jeudi,2
```

1. **Les descripteurs** sont sur la première ligne : `nom`, `classe`, `jour`, `heures`. Chaque ligne suivante est une donnée, avec une valeur par descripteur.
2. **Filtrer** : on garde les lignes où `jour` vaut `lundi`. On obtient Sami et Inès.
3. **Trier** : par `heures` décroissantes, Inès (4) passe avant Sami (2).
4. **Croiser** : avec un second fichier des notes qui a aussi un descripteur `nom`, on rapproche absences et résultats d'un même élève.

Dans un tableur, le filtre se fait en deux clics sur la colonne `jour`.

Retour au problème

Ranger chaque information sous un descripteur précis rend la recherche immédiate, que le fichier ait quatre lignes ou des milliers.

Les pièges

- Une ligne = un objet, une colonne = un descripteur, pas l'inverse.
- Les métadonnées peuvent révéler des informations personnelles.
- Le cloud n'est pas immatériel : ce sont des machines bien réelles.

© Astrolab Cours · astrolabcours.fr/nsi/seconde/donnees-structurees/ · Tous droits réservés : usage personnel uniquement, republication et usage commercial interdits