

LES DONNÉES STRUCTURÉES ET LEUR TRAITEMENT - VOCABULAIRE

Contenus et capacités

Contenus	Capacités attendues
Données	Définir une donnée personnelle. 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.

Note d'intention

Avant de parler traitement de données ou structure de données, il nous semble pertinent de présenter le vocabulaire de la gestion de données : objet, donnée, collections, type de donnée, format d'une donnée, donnée personnelle, descripteur, données structurées, structure en table, format .csv, base de données, ...

Exercices

Exercice 1 - tweet

1. Choisir un livre au CDI.
2. Résumer en 1 tweet l'objet que vous avez entre les mains afin qu'un de vos camarades puisse l'identifier parmi tous les objets à partir de la lecture des 280 caractères.

Commentaires

Cet exercice à faire seul ou en binôme, permet d'introduire le mot de vocabulaire objet.

Exercice 2 - identifier un objet

1. Après avoir récupéré le tweet d'un de vos camarades, identifier l'objet décrit par ce tweet dans toute la collection d'objet.

Commentaires

Cet exercice permet de présenter le terme de collection d'objet et pose les préalables à la recherche d'un objet dans une collection.

Exercice 3 - analyse des descriptions

1. Quelles données ont systématiquement été décrites dans les tweets et de quels types sont-elles ?
2. Quelles données ont quelques fois été décrites dans les tweets et de quels types sont-elles ?
3. Tous les nombres sont-ils écrits pareils ?
4. Quelle(s) donnée(s) permet(tent) d'identifier le livre avec certitude ?

Pour une personne physique, ce genre de données est appelé « données personnelles ».

Commentaires

Ici les élèves où les groupes d'élèves sont invités à consulter le plus possible de tweets écrits par la classe.

Il est alors possible d'introduire le vocabulaire de donnée et faire émerger le terme générique de descripteurs d'une donnée (titre, auteur, date de parution, nombre de pages, ...).

Une fois cette formalisation faite, il est possible d'extraire tous les descripteurs auxquels, globalement, les élèves ont pensé même si ce n'est qu'une seule fois. Pour prolonger, les inviter à en trouver d'autres.

A l'issue de cela, les élèves sont amenés à s'intéresser aux différents types d'une donnée et proposer les trois classifications : texte, nombre, booléen.

En s'intéressant maintenant aux données de type nombre, il est possible de s'interroger aux différents formats permettant d'écrire ces nombres (date, taille, entier, décimal, ...).

Pour finir, les élèves sont amenés à réfléchir à la question des données permettant, avec certitude, d'identifier le livre : le numéro ISBN (sous réserve que nous n'ayons pas deux fois le même livre dans la collection). Il est possible à ce moment-là d'évoquer la notion de données personnelles en précisant que cette notion est réservée aux personnes physiques.

Exercice 3 - et vous ?

1. Choisir un objet, dont il est possible de faire la collection. En énoncer une liste de descripteurs.

Retrouvez éducol sur :



Exercice 4 - descripteurs humains

1. Énoncer des descripteurs pouvant s'appliquer à un être humain. Préciser leur type.
2. Énoncer des descripteurs correspondant à une donnée personnelle pour un être humain.

Commentaires

Deux exercices d'application. L'exercice 4 permet d'aborder ce qu'est une donnée personnelle : numéro sécurité sociale, ...

Exercice 4 bis - mise en commun

1. Écrire son tweet dans le pad indiqué par votre enseignant.

Commentaires

Le pad est un bloc note collaboratif, tous les élèves peuvent s'y connecter en même temps. Cela peut se faire avec les téléphones des élèves en l'absence d'ordinateur. La majorité des ENT intègrent un tel pad.

Si un pad n'est pas disponible ou si l'on n'est pas à l'aise avec cet outil, on peut récupérer en fin de séance tous les tweets écrits des élèves et les intégrer en dehors de la classe dans un unique fichier .txt ou .odt pour la séance suivante. Le faire faire individuellement en classe par les élèves prend un certain temps.

Exercice 5 - travail sur le document unique

À partir du pad élaboré collaborativement, résumer la collection en un seul fichier. Ce fichier devra :

- être compréhensible par quelqu'un d'extérieur à la classe;
- être pratique à consulter;
- permettre de rajouter facilement un nouvel objet à la collection;
- permettre de supprimer facilement un objet de la collection;
- être synthétique.

Commentaires

Naturellement, une majorité d'élèves (voire tous) vont s'orienter vers le tableur. La première ligne contiendra les descripteurs et chaque autre ligne correspondra à un objet. Quelques-uns auront un fichier symétrique : les descripteurs seront dans la première colonne et chaque colonne correspondra à un objet.

Ils ont naturellement créé un fichier .csv!

Retrouvez éducol sur :



Exercice 6 - données personnelles

Déterminer la quantité de données possédées par Facebook et/ou GOOGLE sur soi. Préciser l'unité utilisée pour mesurer cette quantité.

Commentaires

Une recherche sur un moteur de recherche permet de trouver rapidement où télécharger une archive des données que possède GOOGLE et FACEBOOK. La construction de l'archive peut cependant prendre plusieurs jours.

L'unité est l'octet, ou le kilo octet, ou le méga-octet, ... mais bien souvent c'est du giga-octet.

Exercice 7 - comprendre la structure du fichier .csv

1. Nommer le logiciel utilisé par le système d'exploitation pour ouvrir le fichier [budget_parisv1.csv](#).
2. Forcer l'ouverture de ce fichier par un bloc note (wordpad).
3. Décrire comment sont structurés les fichiers .csv.
4. Déterminer le caractère séparateur des fichiers :
 - [budget_parisv2.csv](#)
 - [budget_parisv3.csv](#)
 - [budget_parisv4.csv](#)
 - [budget_parisv5.csv](#)

Commentaires

Même si la visualisation est naturelle avec un tableur, un fichier .csv n'est rien d'autre qu'un fichier texte. On y trouve un objet par ligne, la première ligne correspondant aux descripteurs. Les données sont séparées par un caractère séparateur (, ; tabulation espace).

Retrouvez éduscol sur :

