



Fiche d'accompagnement

Parcours : Prérequis - Qu'est-ce que la donnée ?

Séance 0 : Séance introductive

Niveaux : cycles 2 et 3
Âge : 7 à 12 ans

Avec



Sommaire

Sommaire	2
Objectifs	3
Savoirs	3
Savoir-être	3
Savoir-faire	3
DÉROULEMENT DE LA SÉANCE	4
Introduction	4
Etape 1 : Introduction	4
Emergence des représentations	4
Etape 2 : À votre avis, qu'est-ce qu'une donnée ?	4
Approfondissement	5
Etape 3 : Vidéo "Le big data fait-il grossir"	5
Etape 4 : Dans la vidéo, qu'apprend-on en plus sur les données ?	5
Pour alimenter le débat	5
Construction de la définition	6
Manipulation	7
Etape 5 : Manipuler une base de données avec DataDecode	7
Question finale	8
Etape 6 : À votre avis que pourrait-on faire d'autre avec un algorithme qui a accès à de nombreuses données ?	8
Pour alimenter le débat	8
Conclusion	9
Etape 7 : Messages-clés	9
Pour aller plus loin :	10

Objectifs

- Comprendre les principales mécaniques régissant le circuit de la donnée (récolte, stockage, traitement)
- Découvrir le fonctionnement de DataDecode

Savoirs

- Découvrir ce qu'est un serveur, un algorithme, une requête
- Découvrir le fonctionnement d'une base de données et de tags
- Découvrir le langage de programmation.

Savoir-être

- Comprendre le fonctionnement des algorithmes pour mieux se les approprier

Savoir-faire

- Récolter et trier des données
- Créer un algorithme pour le manipuler
- Utiliser un système de tags pour générer des résultats ciblés

Thèmes abordés

- Stockage, traitement de la donnée. Données personnelles.

Ressources utilisées

- Vidéo « Le big data fait-il grossir ? »
- Activité Data-Decode : Générateur

Durée de l'atelier

- 45 minutes

Modalité

- Activité alternant manipulations en groupe et en binômes,
- Activité Data Decode à réaliser sur le site internet code-decode.net

Matériel nécessaire

- Un ordinateur connecté
- Un vidéoprojecteur ou TNI
- Un ordinateur pour deux enfants

DÉROULEMENT DE LA SÉANCE

Introduction

Etape 1 : Introduction

L'objectif de ce parcours est de définir ce qu'est une donnée en s'intéressant aux mécaniques fondamentales qui la régissent.

En récoltant et manipulant des données de la classe, on pourra aborder le sujet de la donnée personnelle.

Les sujets suivants seront abordés et pourront faire l'objet de séances d'approfondissement

- récolte et stockage de la donnée / big data
- algorithme / traitement
- donnée personnelle

Emergence des représentations

Etape 2 : À votre avis, qu'est-ce qu'une donnée ?

Objectif	Éléments de réponse
À votre avis, c'est quoi une donnée ?	
Faire émerger les représentations de chacun pour pouvoir ensuite co-construire une première définition de la donnée avant de la confronter aux informations apportées par le média de l'étape 3.	<p>Pour alimenter la réflexion, on peut se demander :</p> <ul style="list-style-type: none">- Est-ce qu'une vidéo ou une musique est une donnée ?- Est-ce qu'un site internet ou un mail est une donnée ? <p>Les deux réponses sont oui, cependant le site internet ou le mail ne sont pas des données stockées sur notre ordinateur, dans ce cas où sont-elles ?</p>

Approfondissement

Etape 3 : Vidéo "Le big data fait-il grossir"

Visionnage de la vidéo en collectif.

Etape 4 : Dans la vidéo, qu'apprend-on en plus sur les données ?

Pour alimenter le débat

- Où est-ce que je peux trouver une donnée ?
- Est-ce que je peux stocker une donnée sur mon ordinateur ou un cloud ?
- Où se trouve la donnée que je partage ?
- En quoi cela me concerne-t-il ?

Objectif	Éléments de réponse	Info + / exemple
Alimenter le débat : Où est-ce que je peux trouver une donnée ?		
Définir le contexte de la donnée numérique, Établir un lien entre les éléments qui nous paraissent être virtuels et leur existence physique sur un ordinateur.	Tout ce qui est enregistré sur mon ordinateur est données. Internet est constitué d'ordinateurs reliés les uns aux autres. Internet représente pour nous un océan de données (d'où le terme naviguer)	Exemples de données auxquelles on a accès grâce à internet : <ul style="list-style-type: none"> - mails - radios - podcasts - tv - streaming - vidéos - musiques - blogs - articles - forums de discussion
Alimenter le débat : Est-ce que je peux stocker une donnée sur mon ordinateur ou un cloud ?		
Définir le format de la donnée et déterminer son existence physique en s'intéressant aux différentes formes de stockage.	Un cloud est un espace de stockage proposé par un ordinateur distant auquel je peux me connecter.	Un mail est de la donnée que je peux consulter sur un ordinateur distant appelé serveur. Un site internet est un ensemble de donnée hébergé sur un serveur. Les films, musiques, photos sur mon ordinateur sont de la donnée.
Alimenter le débat : Où se trouve la donnée que je partage ?		
Définir l'existence physique de données accessibles uniquement virtuellement.	Parfois une donnée partagée se trouve sur tous les ordinateurs qui la consultent, parfois aucun. Quand j'envoie un fichier par mail, il se trouve à la fois sur mon ordinateur, sur celui du destinataire mais aussi sur les serveurs qui permettent au mail de circuler. Quand je consulte un article de journal sur internet, l'article n'est ni chez moi ni chez les autres personnes qui le consultent, il est sur les ordinateurs du journal qui leur permettent d'"héberger" leurs articles.	Les vidéos sur Youtube sont des vidéos contenue sur leurs serveurs et on y accède avec le nôtre via leur site. La moindre lettre écrite sur un blog existe physiquement sur une machine connectée à internet.

Alimenter le débat : En quoi cela me concerne ?

Déterminer les différents producteurs de data et les différentes façon d'en produire.

Lorsque j'écris un message, un article, que je produis une vidéo, que je crée une musique, tout ce que je filme, enregistre ou prend en photo, devient de la donnée numérique une fois enregistrée sur un ordinateur.

Quand je navigue sur internet je laisse aussi des traces, à chaque fois que je remplis un formulaire pour m'inscrire à un site internet, ces informations sont enregistrées et associée à un profil.

Il y a aussi des données que je génère sans le vouloir. Les sites que je consulte enregistrent des informations sur ma navigation, appelées méta-données. Cela leur permet par exemple de savoir si j'ai déjà consulté le site une première fois.

SMS
Formulaires
Stories
Murs des réseaux sociaux
Réseaux sociaux photos ou vidéos
Cookies
Meta-données

Construction de la définition

Objectif	Éléments de réponse	Info + / exemple
Construire une définition commune.	<p>Une donnée est enregistrée sur un ordinateur. Elle peut être partagée et se diffuser sur plusieurs ordinateurs. Tout le monde en génère en utilisant l'outil informatique (téléphone, ordinateur, etc.)</p> <p>Les données peuvent être étudiées par des experts, plus ils en ont, plus les résultats de leur étude seront fins.</p>	




Manipulation


Etape 5 : Manipuler une base de données avec DataDecode

Objectif : Découverte de Data decode et de la notion de base de donnée.

Manipulation : Alimenter un générateur d'histoires en récoltant et ordonnant des mots dans une base de données.

Consignes :

- À chaque fois que tu cliques sur le bouton  la création se jouera et affichera une phrase différente. Clique sur le bouton  quand tu as bien testé et sur ce bouton,  pour afficher de nouveau les consignes.

- Dans la zone de texte de l'onglet  sont définis quatre mots-clés : **SUJET, ACTION, OBJET, LIEU**. Si on observe la collection de mots de la base de données, on observe que chacun des mots possède un tag correspondant à ces mots-clés. Pour chaque mot-clé, le code cherche dans la base de donnée un mot, au hasard du même tag et remplace le mot-clé par le résultat de cette recherche. Cela permet d'obtenir une phrase différente à chaque fois. Passe à l'étape suivante pour ajouter des possibilités à notre phrase.
- Pour ajouter des possibilités et transformer encore plus cette phrase, rien de plus simple : Ajoute un mot dans la base de données et donne lui un tag en fonction du sens de ce mot dans la phrase. Pour éviter les incohérences, chaque mot ne peut avoir qu'un seul tag

Exemple :

Le tracteur roule dans le champs

- tracteur est **sujet**

Le fermier monte dans le tracteur

- tracteur est un **lieu**

Le fermier répare le tracteur à la ferme

- le tracteur est un **objet**

- Pour augmenter le nombre de résultats absurdes quels groupes de tags pourrait-on appliquer à un même mot ou plusieurs?

Exemple :

Le prince et la princesse peuvent êtres taggués à la fois 'sujet' et 'objet'

Question finale

Etape 6 : À votre avis que pourrait-on faire d'autre avec un algorithme qui a accès à de nombreuses données ?

On a vu qu'on peut stocker toutes sortes d'informations dans une base de données et qu'à l'aide d'algorithmes il est possible de les manipuler pour, par exemple, écrire une histoire aléatoirement.

Pour alimenter le débat

- Quel intérêt à accumuler autant de données ?
- Si ce sont des données personnelles peut on identifier quelqu'un ou connaître ses habitudes ?
- Que pourrait-on apprendre avec des données publiques ?

Objectif	Éléments de réponse	Info + / exemple
Quel intérêt à accumuler autant de données ?		
Définir en quoi il peut être intéressant d'accumuler de nombreuses données à propos d'un sujet.	<p>Pour se rappeler d'une pensée ou revoir un film on a tous eu envie d'enregistrer quelque chose.</p> <p>Mais les algorithmes ont aussi le pouvoir de traiter toutes ces données rapidement pour aboutir à des conclusions.</p> <p>Plus on a de données sur un sujet plus cela permet de le traiter avec précision et plus les conclusions des algorithmes seront précises.</p>	<p>Lorsque des scientifiques veulent prouver une expérience, ils doivent la reproduire plusieurs fois. Plus cette expérience réussit plus ils pourront valider leur théorie.</p> <p>Dans un système de votes ou de statistiques, plus on a de votant plus le résultat sera précis. Exemple : plus il y a de personnes pour attribuer la même note à un produit plus elle est fiable. Si on cherche la meilleure recette de gâteau au chocolat sur Marmiton, on choisira plutôt une recette bien notée par beaucoup de personnes. Comparer 1 app 5 étoiles mais avec un seul vote à 1 app 4 étoiles mais avec plus de vote, sur le playstore, laquelle semble être mieux?</p>
Si ce sont des données personnelles peut on identifier quelqu'un ou connaître ses habitudes ?		
Définir les processus simples qui permettent de déduire nos habitudes de vie.	<p>Lorsque ce sont des données personnelles qui sont récoltées comme le nom prénom, profession, âge, nombre de personnes dans la famille, plus on a de critères (tags) pour définir une personne, plus il sera simple de la retrouver dans la base.</p> <p>Ces données peuvent aussi être associées à des données de géolocalisation, ce qui permet très</p>	<p>Par exemple, il n'est écrit nulle part le nom de ton école par contre, dans les bases de données on pourra retrouver que tous les matins de toute l'année, sauf pendant les vacances et les weekend, tu t'arrêtes pendant la journée au niveau d'un établissement scolaire. Donc on peut savoir où tu étudies, si tu fais une activité, laquelle, à quelle</p>

	rapidement de déterminer un trajet quotidien.	fréquence.
Que pourrait-on apprendre avec des données publiques ?		
Présenter l'open data, montrer que c'est un sujet accessible à tous.	<p>L'open data c'est rendre accessible une base de données au grand public.</p> <p>Ce sont des données que nous pouvons consulter et utiliser pour nos algorithmes.</p>	<p>Les bases de données publiques existent sous de nombreuses formes.</p> <p>Recensement des prénoms par année, cartes géographiques, flux météorologiques.</p> <p>Wikipedia (encyclopédie ouverte)</p> <p>Git (plateforme de travail ouverte)</p> <p>Dans ces deux cas, plus de personnes se regroupent sur un article ou projet plus celui-ci peut être considéré comme fiable.</p>

Conclusion

Etape 7 : Messages-clés

- La donnée c'est tout ce que je peux conserver dans le disque dur de mon ordinateur.
- Toutes les données ne sont pas sur mon ordinateur, grâce à internet je peux me connecter à des ordinateurs distants, appelés serveurs, pour en consulter d'autres, les données que je consulte sont alors présentes dans les disques durs de ces ordinateurs.
- Tout le monde génère de la donnée y compris sans le vouloir en navigant sur internet quelque soit la technologie utilisée.
- On peut utiliser un algorithme pour manipuler les données et obtenir un résultat.
- On peut identifier quelqu'un à partir de ses données personnelles.

Pour aller plus loin :

- Faire un parallèle entre notre base de données d'histoires et une base de données vidéos comme Youtube (peut-on faire une recherche, filtrer, accéder à du contenu, etc..)
- Questionner le groupe sur ce qui paraît possible de faire avec les data
- Profiling, statistiques, partage de culture, IA, ...