

LES DONNÉES STRUCTURÉES

Très fortement inspiré de l'excellent site InfoForAll

I Stocker une donnée en informatique

Connaissance n°1 Le bit

L'unité de base du stockage de l'information en informatique est le **bit**. Il s'agit d'une case mémoire qui ne possède que deux états :

Vrai ou Faux, Haut ou Bas ou encore plus simple : 0 et 1 .

Remarque n°1.

Si on prend deux bits, on aura alors 4 possibilités différentes : 2^2

00 01 10 11

Si on prend trois bits, on aura alors 8 possibilités différentes : 2^3

000 001 010 011 100 101 110 111

etc.

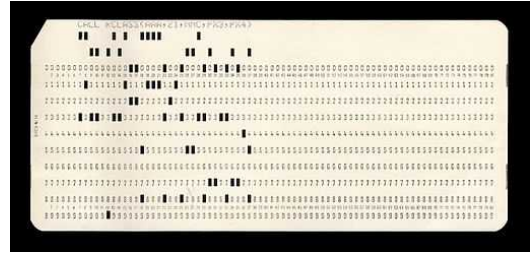
1 1 1 1 1 1 1 1 (dernière possibilité : 255)

LES DONNÉES STRUCTURÉES

Connaissance n°3

Les débuts

Au début de l'utilisation des systèmes automatisés, on utilisait des systèmes mécaniques et les données étaient codées sous forme de cartes perforées : on voyait bien les 0 et les 1 !



Carte perforée (Mutatis mutandis CC BY-SA sur Wikimedia)

Puis vint un autre système de stockage, plus rapide mais toujours fortement mécanique : la bande magnétique.

L'avantage par rapport aux fiches ? Plus de vitesse et la possibilité de modifier l'enregistrement sans sortir du scotch !



Bande magnétique (CC BY-SA sur Wikimedia)

En 1956, la société IBM crée un premier disque dur de 5 Mo. L'ensemble mesurait alors

1,52 x 1,73 x 0,74 m !

Le coût ? Environ 10 000 \$ le Mo !



L'IBM 3330 (source <https://www.cs.auckland.ac.nz>)

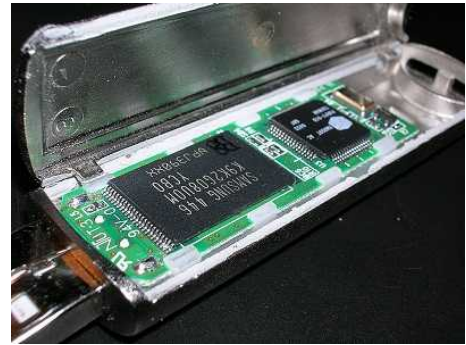
LES DONNÉES STRUCTURÉES

Connaissance n°4

Plus actuel

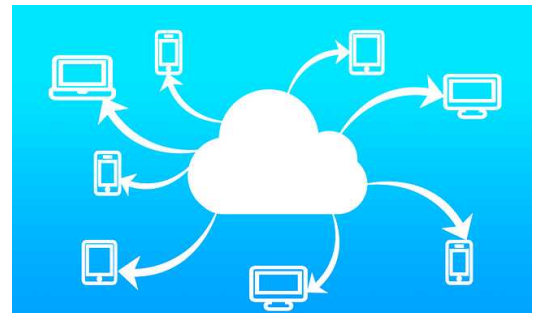
Dans les années 1984 sont apparus des systèmes de stockage de masse purement électroniques, sans pièce en mouvement : la carte SD et la clé USB.

Ce type de technologie remplace maintenant de plus en plus les disques durs classiques. Il s'agit des fameux disques durs SSD.



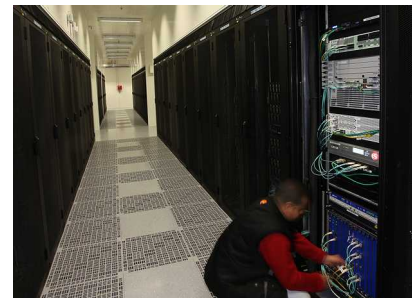
Intérieur d'une clé USB (image libre de droit, Wikipedia)

En 2007, le stockage a même l'air d'être devenu immatériel : on passe au stockage dans le "Cloud" avec Dropbox par exemple.



(image licence Pixabay, Image par Pete)

Mais en réalité, sous cette belle image de nuage tout mignon, se cache des Data Centers, c'est à dire un entassement d'ordinateurs. Et ils chauffent, quoi qu'on en dise...



(image récupérée sur le site du lycée Brassens)

Cela pourrait s'améliorer : [vidéo Orange sur le data center](#)

LES DONNÉES STRUCTURÉES

Remarque n°2.

les prix ont baissé de façon drastique :

- en 1958 : 10000 \$ le Mo
- en 2000 : 0.02 \$ le Mo...
- en 2010 : 0.08 \$ le Go, 8 cents pour mille fois plus de mémoire !
- et ça continue à baisser...

Connaissance n°5

Les sous-unités de l'octet

| Sous-unité | Valeur en octet | Equivalent au stockage |
|-----------------------------------|--------------------------|-------------------------------------------------------------|
| Un kilo (k) représente 10^3 | 1000 | Un petit texte |
| Un Méga (M) représente 10^6 | 1000000 | un livre ou une image HD |
| Un Giga (G) représente 10^9 | 1000000000 | un film |
| Un Téra (T) représente 10^{12} | 1 puis douze zéros | |
| Un Péta (P) représente 10^{15} | 1 puis quinze zéros | capacité mémoire d'un cerceau humain selon certaines études |
| Un Exa (E) représente 10^{18} | 1 puis dix-huit zéros | |
| Un Zetta (E) représente 10^{21} | 1 puis vingt-et-un zéros | Production annuelle d'informations par Internet |

LES DONNÉES STRUCTURÉES A01

EXERCICE N°1

Fouillez dans la configuration de votre Smartphone ou de votre ordinateur pour trouver la place mémoire totale.

LES DONNÉES STRUCTURÉES

EXERCICE N°2

Les mémoires des smartphones sont-elles composées de disques durs classiques ?

LES DONNÉES STRUCTURÉES

Les mémoires des smartphones sont-elles composées de disques durs classiques ?

Non, ces mémoires sont des mémoires purement électroniques, sans composant mobile. Comme les cartes et disques durs SSD.

LES DONNÉES STRUCTURÉES

Connaissance n°6

En réalité, on ne stocke pas juste des données.

Donner un rapport de 30 000 pages écrits à un système informatique ne sert pas à grand chose. Pour être efficace, il faut fournir au système les informations dans un format bien défini.

C'est ce qu'on nomme des **données structurées**.

LES DONNÉES STRUCTURÉES

EXERCICE N°3 *Une donnée non structurée*

Une donnée non structurée est une donnée présente mais qu'un système automatisé aura bien du mal à retrouver.

Voici un texte :

C'est une voiture de sport de couleur rouge et conduite par un moustachu qui porte des chemises à fleurs.

L'un de ses amis possède un bel hélicoptère. Celui-ci est plutôt marron mais possède également quelques lignes colorées dans des tons rouge, bleu et jaune.

- 1) Quel est le type et la couleur du premier véhicule ?
- 2) Quel est le type et la couleur du second véhicule ?

LES DONNÉES STRUCTURÉES

C'est une voiture de sport de couleur rouge et conduite par un moustachu qui porte des chemises à fleurs.

L'un de ses amis possède un bel hélicoptère. Celui-ci est plutôt marron mais possède également quelques lignes colorées dans des tons rouge, bleu et jaune.

1) Quel est le type et la couleur du premier véhicule ?

Véhicule : Voiture de sport --- Couleur : rouge

2) Quel est le type et la couleur du second véhicule ?

Véhicule : Hélicoptère --- Couleur : marron

LES DONNÉES STRUCTURÉES

Vous venez d'utiliser un centre de traitement de l'information super performant : votre cerveau.

Il existe une branche de la recherche informatique qui traite de cette capacité de donner du sens à un texte brut : C'est l'Intelligence Artificielle (Vous connaissez tous ChatGPT).

Mais on peut s'en passer au prix d'un peu d'organisation...

LES DONNÉES STRUCTURÉES

Nous allons **structurer** les données.

Connaissance n°7 Les données structurées

Une **donnée structurée** est une donnée présentée selon un format connu et fixe.

Il peut s'agir par exemple d'un tableau à deux entrées.

Exemple n°1.

| Identifiant | Type | Utilisation | Couleur |
|-------------|-------------|-------------|---------|
| 0 | Voiture | Sport | Rouge |
| 1 | Hélicoptère | Tourisme | Marron |
| 2 | Voiture | Sport | Rouge |

Remarque n°3.

Cette fois, on voit qu'on pourra demander facilement (sans IA) à l'ordinateur de renvoyer les couleurs des véhicules : il suffit de trouver la bonne colonne. C'est beaucoup plus facile.

LES DONNÉES STRUCTURÉES

Connaissance n°8

- Les colonnes sont nommées les **descripteurs**.

Toutes les données présentes dans ce tableau partagent les mêmes descripteurs et les fournissent dans le même ordre :

ID - Type - Catégorie - Couleur.

- Chaque ligne se nomme un **enregistrement**

La seule ligne qui ne soit pas un enregistrement est la toute première, celle qui est grisée. Il s'agit de l'en-tête (header en anglais). Elle permet de connaître le nom et l'ordre des descripteurs.

- La **collection** est le nom donné à l'ensemble des enregistrements, donc au tableau.

Exemple n°2.

| Collection | Attributs | | | |
|-----------------|-------------|-------------|-------------|---------|
| | Identifiant | Type | Utilisation | Couleur |
| Enregistrements | 0 | Voiture | Sport | Rouge |
| | 1 | Hélicoptère | Tourisme | Marron |
| | 2 | Voiture | Sport | Rouge |

LES DONNÉES STRUCTURÉES

EXERCICE N°4

| Collection | Attributs | | | |
|-----------------|-------------|-------------|-------------|---------|
| | Identifiant | Type | Utilisation | Couleur |
| Enregistrements | 0 | Voiture | Sport | Rouge |
| | 1 | Hélicoptère | Tourisme | Marron |
| | 2 | Voiture | Sport | Rouge |

- 1) Quels sont les descripteurs de la collection précédente ?
- 2) Combien y-a-t-il d'enregistrements dans cette collection ?
- 3) Quelle est la valeur qui correspond au descripteur d'indice 2 et à l'enregistrement d'indice 0 ?
- 4) Quel est le seul attribut qui permette de distinguer le premier et le dernier enregistrement ?

LES DONNÉES STRUCTURÉES

| Collection | Attributs | | | |
|-----------------|-------------|-------------|-------------|---------|
| | Identifiant | Type | Utilisation | Couleur |
| Enregistrements | 0 | Voiture | Sport | Rouge |
| | 1 | Hélicoptère | Tourisme | Marron |
| | 2 | Voiture | Sport | Rouge |

1) Quels sont les descripteurs de la collection précédente ?

Il y a 4 descripteurs :

- Indice 0 : Identifiant
- Indice 1 : Type
- Indice 2 : Utilisation
- Indice 3 : Couleur

2) Combien y-a-t-il d'enregistrements dans cette collection ?

Il y a 3 enregistrements d'indice 0, 1 et 2.

3) Quelle est la valeur qui correspond au descripteur d'indice 2 et à l'enregistrement d'indice 0 ?

La valeur d'indice 2 en descripteur et d'indice 0 en enregistrement correspond à **sport**

4) Quel est le seul attribut qui permette de distinguer le premier et le dernier enregistrement ?

Il s'agit de **l'identifiant**. On rajoute TOUJOURS un identifiant qu'on définit soi-même. De cette façon, nous sommes certains de pouvoir distinguer deux enregistrements même si les contenus sont similaires.

LES DONNÉES STRUCTURÉES

EXERCICE N°5

Peut-on séparer deux individus uniquement à l'aide du nom et du prénom ? A quoi sert le numéro de sécurité sociale ?

LES DONNÉES STRUCTURÉES

Peut-on séparer deux individus uniquement à l'aide du nom et du prénom ? A quoi sert le numéro de sécurité sociale ?

Si chaque élève de la classe tape son nom et son prénom dans un moteur alors il est extrêmement probable que qu'au moins l'un d'eux se trouve un homonyme sur le net...

Afin de pouvoir différencier deux individus portant éventuellement les mêmes noms et prénoms, on attribue à chaque personne un numéro unique : c'est le numéro de sécurité sociale.

LES DONNÉES STRUCTURÉES

EXERCICE N°6

Nous allons travailler sur un jeu de données stocké dans un fichier dit CSV et qu'un tableur (comme excel ou calc) sait ouvrir et lire. Il contient les caractéristiques des Pokemons. En voici un extrait :

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|---|---------------------------|--------|--------|-------|----|--------|---------|---------|---------|-------|------------|-----------|---|
| 1 | # | Name | Type 1 | Type 2 | Total | HP | Attack | Defense | Sp. Atk | Sp. Def | Speed | Generation | Legendary | |
| 2 | 1 | Bulbasaur | Grass | Poison | 318 | 45 | 49 | 49 | 65 | 65 | 45 | 1 | False | |
| 3 | 2 | Ivysaur | Grass | Poison | 405 | 60 | 62 | 63 | 80 | 80 | 60 | 1 | False | |
| 4 | 3 | Venusaur | Grass | Poison | 525 | 80 | 82 | 83 | 100 | 100 | 80 | 1 | False | |
| 5 | 3 | VenusaurMega Venusaur | Grass | Poison | 625 | 80 | 100 | 123 | 122 | 120 | 80 | 1 | False | |
| 6 | 4 | Charmander | Fire | | 309 | 39 | 52 | 43 | 60 | 50 | 65 | 1 | False | |
| 7 | 5 | Charmeleon | Fire | | 405 | 58 | 64 | 58 | 80 | 65 | 80 | 1 | False | |
| 8 | 6 | Charizard | Fire | Flying | 534 | 78 | 84 | 78 | 109 | 85 | 100 | 1 | False | |
| 9 | 6 | CharizardMega Charizard X | Fire | Dragon | 634 | 78 | 130 | 111 | 130 | 85 | 100 | 1 | False | |
| 10 | 6 | CharizardMega Charizard Y | Fire | Flying | 634 | 78 | 104 | 78 | 159 | 115 | 100 | 1 | False | |

- 1) Combien d'enregistrement sont-ils visibles sur cette photo d'écran ?
- 2) Quelle est la lettre caractérisant le descripteur points de vie (HP pour Hit Points en anglais) ?
- 3) La cellule G4 sélectionné fait référence à 82. Quel est le descripteur correspondant ?
- 4) Trouver le ou les Pokemons de type Feu qui possède une attaque supérieure à 80 et une défense supérieure à 80.

LES DONNÉES STRUCTURÉES

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|---|---------------------------|--------|--------|-------|----|--------|---------|---------|---------|-------|------------|-----------|---|
| 1 | # | Name | Type 1 | Type 2 | Total | HP | Attack | Defense | Sp. Atk | Sp. Def | Speed | Generation | Legendary | |
| 2 | 1 | Bulbasaur | Grass | Poison | 318 | 45 | 49 | 49 | 65 | 65 | 45 | 1 | False | |
| 3 | 2 | Ivysaur | Grass | Poison | 405 | 60 | 62 | 63 | 80 | 80 | 60 | 1 | False | |
| 4 | 3 | Venusaur | Grass | Poison | 525 | 80 | 82 | 83 | 100 | 100 | 80 | 1 | False | |
| 5 | 3 | VenusaurMega Venusaur | Grass | Poison | 625 | 80 | 100 | 123 | 122 | 120 | 80 | 1 | False | |
| 6 | 4 | Charmander | Fire | | 309 | 39 | 52 | 43 | 60 | 50 | 65 | 1 | False | |
| 7 | 5 | Charmeleon | Fire | | 405 | 58 | 64 | 58 | 80 | 65 | 80 | 1 | False | |
| 8 | 6 | Charizard | Fire | Flying | 534 | 78 | 84 | 78 | 109 | 85 | 100 | 1 | False | |
| 9 | 6 | CharizardMega Charizard X | Fire | Dragon | 634 | 78 | 130 | 111 | 130 | 85 | 100 | 1 | False | |
| 10 | 6 | CharizardMega Charizard Y | Fire | Flying | 634 | 78 | 104 | 78 | 159 | 115 | 100 | 1 | False | |

1) Combien d'enregistrement sont-ils visibles sur cette photo d'écran ?

Il y a 9 enregistrements (9 ligne) sur cette photo d'écran.

2) Quelle est la lettre caractérisant le descripteur points de vie (HP pour Hit Points en anglais) ?

C'est la lettre F.

3) La cellule G4 sélectionné fait référence à 82. Quel est le descripteur correspondant ?

Le descripteur correspondant est : Attack

4) Trouver le ou les Pokemons de type Feu qui possède une attaque supérieure à 80 et une défense supérieure à 80.

C'est le pokemon : CharizardMega Charizard X

