

L'intelligence artificielle

Séance 2 : Stocker, identifier et traiter des données numérisées

Quelle est la différence entre donnée et programme ? Quel est l'espace mémoire utilisé pour stocker différents types de données ?




Différents supports d'information au cours du temps

Des moyens de stockage en perpétuelle évolution



??? Envisager les problèmes éthiques et environnementaux que pose la montée en puissance du stockage des données sur le cloud, proposer des solutions qui permettraient de les limiter.

Différents types et tailles de données

Type	Extension	Taille	Exemples d'exécutables utilisables
Son 	WAV, MP3, FLAC	Quelques Mo pour un morceau de musique de quelques minutes	Itunes, Windows media player
Image 	JPG, BMP, PNG, GIF	Quelques centaines de Ko à plusieurs Mo	Gimp, Apple Photos, Picasa, Paint, Adobe photoshop
Vidéo 	AVI, MOV, MP4	La taille monte très vite en fonction de la durée. De quelques dizaines de Mo jusqu'à plusieurs Go	Windows media player, Quicktime, VLC

DOC Diversité et taille des données multimédia. Le premier type d'informations que les systèmes étaient capables de manipuler étaient des nombres, puis des textes, jusqu'à l'arrivée dans les années 1980 des systèmes multimédia, c'est-à-dire capables de manipuler divers types d'informations: images, sons, vidéos... Les instructions d'un programme informatique sont également des informations et sont stockées dans le même type de mémoire que les données.

L'acquisition d'une donnée informatique

De par sa conception, un ordinateur ne peut manipuler que des données numériques. Dans certains cas, leur acquisition passe par une étape de **numérisation**, au cours de laquelle les grandeurs mesurées par un capteur (scanner, appareil photo, caméra, microphone, etc.) vont être converties en valeurs numériques binaires sous la forme d'une suite de bits symbolisés par 0 et 1 (contraction anglo-saxonne de l'expression *binary digit*, signifiant chiffre binaire).

Par exemple, dans le cas d'un texte, chaque caractère est codé par un octet, c'est-à-dire une valeur numérique comportant 8 bits. Selon la norme ASCII, la lettre « A » sera codée par l'octet 01000001, la lettre « a » par 01100001, etc.

Avec 16 bits par valeur, l'ordinateur sera capable de distinguer 65 536 niveaux sonores différents et plus de 16 millions de couleurs peuvent être codées sur 24 bits (tableau ci-contre).

Information à numériser	Encodage (en bit)	Nombre de valeurs possibles
Texte	8	$2^8 = 256$
Son	16	$2^{16} = 65\,536$
Image (pixel*)	Codage RVB = 8 bits par couleur primaire Rouge, Vert, Bleu soit 24 bits	$2^{24} = 16\,777\,216$

■ Plus le nombre de bits est élevé, plus le nombre de valeurs possibles est important et plus l'encodage est fidèle à la réalité.

??? Pour stocker un livre au format numérique, expliquer s'il vaut mieux filmer les pages, les scanner ou les avoir dans un format texte.

Cas particulier : Calculer la taille en octets d'une page de texte

Dans le codage ASCII, chaque caractère, y compris les espaces et la ponctuation, nécessite un octet pour être stocké en mémoire. Ainsi, un fichier texte contenant 2 000 caractères a besoin de 2 ko de mémoire pour être stocké.

1 octet contient 8 bits et permet donc de stocker un nombre entre 0 à 255. Le codage ASCII sur 1 octet permet donc le codage de 255 caractères différents, suffisants pour un alphabet latin. Il existe cependant d'autres types de codage afin de représenter l'ensemble des alphabets non latins.

1. Rechercher la signification de l'acronyme ASCII.
2. Une page de texte contient environ 2500 signes. En ne considérant que des caractères présents dans la table ASCII, indiquer la taille en mémoire du fichier considéré.
3. Justifier la taille du fichier **sortie.txt** du document ci-contre. (On pourra créer un fichier **donnees.txt** contenant dix fois la lettre A sur une même ligne pour comparer la taille des deux et conclure)

PROTOCOLE

- Ouvrir BlocNote
- Inscrire les deux lignes suivantes:
set /p lettre=Entrez une lettre:
for /l %%i in (1, 1, 10) do (echo
%lettre%>>sortie.txt)
- Enregistrer sous : modifier le type (afficher « Tous les types »)
- Donner le nom **executable.bat** et cliquer sur Enregistrer
- Le fichier obtenu est exécutable, double cliquer dessus pour vous en rendre compte.
- Entrer une lettre. L'exécutable fabrique un fichier **sortie.txt** qui contient dix fois la lettre choisie (une par ligne).
- Noter la taille du fichier **sortie.txt**.

DOC **Construire un fichier exécutable sous Windows.** Voici un protocole pour écrire un programme dans le langage Batch de Windows. Ce programme est exécutable directement sur les machines Windows grâce à un interpréteur présent dans le système d'exploitation : cmd.exe.

Différents processeurs pour les exécuter

À partir d'un problème à résoudre par l'informatique, la première étape est de décomposer ce problème en une suite d'opérations simples décrites en français. C'est l'**algorithme**. Cet algorithme peut ensuite être traduit dans un langage de programmation compréhensible par le programmeur comme Scratch ou Python pour obtenir un **programme**. Le **fichier exécutable** est le fichier qui contient le programme sous une forme compréhensible par l'ordinateur (langage machine) ou interprétable directement par l'ordinateur, c'est à dire traduit en langage machine à la volée au fur et à mesure de l'exécution (programme Python avec le bon interpréteur ou bytecode java par exemple). Le langage machine est constitué d'instructions très simples spécifiques au processeur et compréhensibles directement par l'unité de commande du processeur. Les **données** sont les informations fournies au programme pour qu'il puisse travailler. Un fichier exécutable est une donnée pour un autre programme : le système d'exploitation.

DOC Quelques définitions utiles.

Voici une liste d'extensions de fichiers exécutables avec les systèmes d'exploitation correspondants. Les plus courants sous Windows sont .exe et .msi mais il en existe d'autres comme le .bat par exemple. Ces fichiers sont exécutables et contiennent donc des opérations réalisées par le système d'exploitation. Si, en général, ces fichiers réalisent le programme souhaité, ils peuvent être parfois détournés pour réaliser des actions malveillantes. Il existe d'autres extensions à risque, et certains fichiers textes ou feuilles de calcul peuvent aussi contenir du code exécutable.

Extension	Système d'exploitation
.apk	Android
.app	MacOS
.bat	Windows
.bin	Windows, macOS, Linux
.exe	Windows
.ipa	iOS
.sh	MacOS, Linux

DOC Quelques extensions d'exécutables.

??? Les fichiers **sortie.txt** et **executable.bat** ont été construits avec un même logiciel (BlocNote). Expliquer la différence entre ces deux fichiers.

POINT COURS :

► Les ordinateurs manipulent divers types de données numérisées : images, sons, vidéos... Les programmes sont des informations stockées dans le même type de mémoire que les données. Un programme écrit dans un langage de programmation de haut niveau, comme Python, peut être traduit en instructions spécifiques à chaque type de processeur. Ainsi, un **fichier exécutable** sur une machine n'est pas obligatoirement exécutable sur une autre.

Def :

- **Fichier exécutable** : Fichier qui contient le programme sous une forme qui peut être traduite en instructions spécifiques à chaque type de processeur.