

Les données structurées et leur traitement

Repères historiques

- 1930 : utilisation des cartes perforées, premier support de stockage de données
- 1956 : invention du disque dur permettant de stocker de plus grandes quantités de données, avec un accès de plus en plus rapide
- 1970 : invention du modèle relationnel (E. L. Codd) pour la structuration et l'indexation des bases de données
- 1979 : création du premier tableur, VisiCalc
- 2009 : Open Government Initiative du président Obama
- 2013 : charte du G8 pour l'ouverture des données publiques

Les machines

- Internes (disque dur ou SSD)
- Externes (disque, clé USB)
- Disques durs distants (cloud)
- Les fichiers de données sont stockés sur des supports de stockage
- Les grandes bases de données sont souvent implémentées sur des serveurs dédiés (machines puissantes avec une importante capacité de stockage sur disques). Ces centres de données doivent être alimentés en électricité et maintenus à des températures suffisamment basses pour fonctionner correctement.

Impacts sur les pratiques humaines

- Surabondance des données et au développement de nouveaux algorithmes capables de les exploiter.
- Exploitation de données massives (Big Data)
- Données ouvertes (OpenData)
- Marché de la donnée
 - Cadre juridique permettant de protéger les usagers : RGPD
 - Eau
 - Electricité
 - Métaux rares (fabrication)
 - Pollution

Les données et l'information

- Donnée : Une donnée est une valeur décrivant un objet, une personne, un événement digne d'intérêt pour celui qui choisit de la conserver. Par exemple, le numéro de téléphone d'un contact est une donnée.
- Descripteur : Plusieurs descripteurs peuvent être utiles pour décrire un même objet (par exemple des descripteurs permettant de caractériser un contact : nom, prénom, adresse et numéro de téléphone).
- Collection : Une collection regroupe des objets partageant les mêmes descripteurs (par exemple, la collection des contacts d'un carnet d'adresses).
- Table : La structure de table permet de présenter une collection : les objets en ligne, les descripteurs en colonne et les données à l'intersection. Les données sont alors dites structurées.
- Métadonnées : Pour assurer la persistance des données, ces dernières sont stockées dans des fichiers. Le format CSV (Comma Separated Values, les données avec des séparateurs) est un format de fichier simple permettant d'enregistrer une table. À tout fichier sont associées des métadonnées qui permettent d'en décrire le contenu. Ces métadonnées varient selon le type de fichier (date et coordonnées de géolocalisation d'une photographie, auteur et titre d'un fichier texte, etc.).
- Base de données : Une base de données regroupe plusieurs collections de données reliées entre elles. Par exemple, la base de données d'une bibliothèque conserve les données sur les livres, les abonnés et les emprunts effectués.

les algorithmes et les programmes

- Recherche dans des données structurées :
 - Indexation préalable faite par l'homme.
 - Automatisation de l'indexation à partir de textes, d'images ou de sons par des algorithmes
- Opérations sur une table de données
 - Rechercher une information précise dans la collection
 - Trier la collection sur une ou plusieurs propriétés
 - Filter la collection selon un ou plusieurs tests sur les valeurs des descripteurs
 - Effectuer des calculs
 - Mettre en forme les informations produites