

Chapitre 4 : Outils et méthodes de traitement de l'information

1- Bases de données et logiciels spécialisés

Les bases de données et les logiciels spécialisés sont des outils informatiques qui permettent de stocker, gérer et analyser des données. Une base de données est un gros ensemble d'informations structurées mémorisées sur un support permanent. Les logiciels spécialisés, quant à eux, sont des applications conçues pour des tâches spécifiques comme la conception graphique ou l'analyse de données.

Exemples de bases de données :

- **Bases de données bibliographiques** : Contiennent des informations sur des ouvrages publiés.
- **Bases de données textuelles** : Incluent des textes intégraux de documents.
- **Bases de données multimédias** : Stockent des images, des sons et des vidéos.
- **Bases de données vectorielles** : Utilisées pour les applications d'apprentissage automatique.

Exemples de logiciels spécialisés :

- **R et Python** : Langages de programmation très utilisés pour l'analyse statistique, la fouille de données et le texte.
- **Gephi** : Permet l'analyse de réseau et la création de cartographies.

2- Techniques de data science et intelligence artificielle

L'intelligence artificielle, ou IA, désigne des machines qui imitent les fonctions cognitives humaines en se montrant capables d'apprendre, de raisonner et de résoudre certains problèmes. La science des données est un domaine interdisciplinaire qui consiste à analyser des ensembles de données en combinant statistiques et informatique. Les techniques de data science et d'intelligence artificielle incluent des méthodes comme l'**apprentissage automatique** (Machine Learning) et le **traitement du langage naturel** (NLP), qui permettent aux ordinateurs d'apprendre à partir de données et de comprendre le langage humain. Elles sont appliquées dans divers domaines comme la **recherche de tendances** (Netflix) ou l'**optimisation** (UPS), pour accomplir des tâches humaines complexes à partir d'un grand volume de données.

Quelques techniques clés :

- **Apprentissage automatique (Machine Learning)** : Permet aux systèmes d'apprendre et de faire des prédictions à partir de données, sans être explicitement programmés.
- **Réseaux de neurones et apprentissage profond (Deep Learning)** : Modèles complexes inspirés du cerveau humain, performants pour des tâches complexes comme la reconnaissance d'images.
- **Traitement du langage naturel (NLP)** : Permet aux machines de comprendre, d'interpréter et de générer du langage humain.
- **Analyse statistique** : Un pilier de la data science pour comprendre et interpréter les données.
- **Visualisation de données** : Techniques pour présenter les données de manière visuelle (ex. : graphiques, tableaux de bord) afin de faciliter leur compréhension.

Applications : Ces techniques sont utilisées pour la maintenance prédictive, l'optimisation des itinéraires, l'analyse des sentiments des clients, la reconnaissance faciale, etc..

Exemple concret : Netflix utilise des techniques de data science et d'IA pour analyser les habitudes de visionnage et décider des séries à produire.

3- Cloud computing

Le cloud computing, ou « informatique en nuage », **fait référence à la fourniture de divers services informatiques – tels que serveurs, stockage, bases de données, mise en réseau, logiciels, analyse et intelligence artificielle – via Internet**, communément appelé « **cloud** ». Ce modèle permet un accès à des ressources informatiques à la demande, éliminant la nécessité de gérer une infrastructure informatique physique sur place.

- **Exemples de service de Cloud computing**

Google Drive : Permet aux utilisateurs de stocker, synchroniser et partager des fichiers en ligne.

Gmail : Permet d'envoyer et de recevoir des e-mails sans installer de logiciel de messagerie, l'infrastructure étant gérée par Google.

Netflix : Accès à une bibliothèque de films et de séries via Internet, sans avoir à télécharger ou gérer les fichiers vidéo.

Microsoft 365 : Une suite de productivité qui inclut des applications comme Word, Excel ou PowerPoint, utilisées en ligne.

4- Stratégies de recherche documentaire

La recherche documentaire est une démarche méthodique visant à identifier, collecter, évaluer et exploiter des informations pertinentes sur un sujet donné. Elle est essentielle dans les travaux académiques, les projets professionnels et la veille informationnelle.

1. Définir le sujet et formuler la problématique

La première étape consiste à cerner précisément le sujet de recherche. Cela implique de le formuler sous forme de question claire et ciblée. La méthode des 3QOCP (Qui ? Quoi ? Quand ? Où ? Comment ? Pourquoi ?) est souvent utilisée pour structurer cette réflexion.

2. Identifier les mots-clés et les concepts associés

Une fois le sujet défini, il est crucial de déterminer les mots-clés pertinents, y compris les synonymes et termes connexes. Cette étape facilite l'interrogation efficace des bases de données et des moteurs de recherche.

3. Sélectionner les sources d'information

Il est important de choisir des sources fiables et adaptées au sujet. Cela peut inclure des catalogues de bibliothèques, des bases de données spécialisées, des revues scientifiques, des ouvrages académiques et des sites institutionnels. L'utilisation de ressources telles que le SUDOC (système universitaire de documentation), Cairn.info ou Google Scholar est recommandée.

4. Utiliser les opérateurs de recherche

Les opérateurs booléens (ET, OU, SAUF) permettent de combiner ou d'exclure des termes pour affiner les résultats de recherche. Par exemple, utiliser "éducation ET numérique" restreint la recherche aux documents contenant les deux termes.

5- Évaluer la qualité et la pertinence des ressources

Il est important de pouvoir se fier à l'information que l'on a repérée, qu'elle soit imprimée ou électronique. On doit faire preuve de jugement critique pour en évaluer la qualité et il faut garder en tête que toute information dont on ignore la provenance devrait a priori être écartée.

Pour évaluer la qualité des sources, on se posera les questions suivantes :

Qui : Quelle est la provenance du document et qui en est l'auteur ?

C'est la **crédibilité** de la source qui est évaluée. Ce critère est souvent plus facile à établir lorsqu'il s'agit de documents imprimés que lorsqu'il s'agit d'informations trouvées sur le net.

Pourquoi : Quel est l'objectif poursuivi par l'auteur de l'information ?

On s'interroge ici sur la validité et plus particulièrement sur son **objectivité**.

Comment : Les documents sont-ils basés sur des recherches ou s'agit-il de textes d'opinion ?

C'est à nouveau la validité de l'information qui est ici évaluée ou plus particulièrement son exactitude.

Quand : A quelle date, époque les documents ont-ils été publiés, créés ?

C'est encore de la validité de l'information qu'il s'agit ou plus particulièrement de son actualité.

Où : Dans quel pays ou quelle région les documents ont-ils été publiés ?

On évalue la **validité** de l'information d'un point de **vue géographique**.

Quelques conseils : Pour évaluer la pertinence du document, il n'est pas nécessaire de lire tout le document. Une exploration rapide suffit pour se faire une idée du contenu du document. On sera attentif :

- au titre du document ;
- au résumé (abstract): on le trouve dans la plupart des notices bibliographiques tirées des bases de données, au début ou à la fin des articles de revues et souvent à l'endos des livres ;
- à la table des matières : elle permet de mieux apprécier le contenu et de bien repérer les chapitres qui peuvent être pertinents ;
- aux tableaux, graphiques, etc.: ils peuvent aider à la compréhension du sujet et être utiles pour le travail.

Enfin, pour évaluer le niveau de l'information, on s'interrogera sur les éléments suivants :

- la nature du document : déterminer s'il s'agit d'un document pédagogique, de recherche ou de vulgarisation. Il faut garder en mémoire qu'il s'agit d'effectuer un travail de niveau universitaire et savoir reconnaître la nature des revues consultées ;
- la qualité de l'information ;
- l'introduction et conclusion : consulter l'introduction et la conclusion du document. Une lecture rapide permet d'évaluer le niveau de l'information, de connaître quelle est la question de départ et quelles sont les conclusions que l'auteur en tire ;
- la spécialisation du vocabulaire : noter le niveau de spécialisation du vocabulaire utilisé.