

République Algérienne Démocratique et Populaire
Ministère de l'enseignement Supérieur et de la Recherche Scientifique
Université des Frères Mentouri Constantine
Faculté des Sciences de la Nature et de la Vie
Département de Biologie Animale

Cours d'informatique

**Destiné aux étudiants en Licence 3 Entomologie, Génétique,
Immunologie et Toxicologie**



Enseignante : Dr.Aouatef CHAIB
ANNEE UNIVERSITAIRE 2020-2021

Sommaire

CHAPITRE 1 : INTRODUCTION EN INFORMATIQUE.....	4
1. Introduction	5
2. L'ordinateur	5
2.1 Familles d'ordinateurs	5
2.2 Constitution de l'ordinateur	5
3. Hardware	5
3.1 L'unité centrale :	6
La mémoire vive RAM.....	6
La mémoire morte ROM	6
Les unités de mesure de mémoire	6
4. Software	6
4.1 Logiciel.....	6
4.2 Types de logiciels.....	6
4.3 Programme.....	7
CHAPITRE 2 : LES BASES DE DONNEES ET LES BANQUES DE DONNEES BIOLOGIQUES ..	8
2.1 Introduction	9
2.2 Définitions	9
2.2.1 Qu'est-ce qu'une donnée ?	9
2.2.2 Les bases de données	9
2.2.3 Système de Gestion de Bases de Données SGBD.....	10
2.2.4 Banque de données	10
2.3 Banque et bases de données biologiques	11
2.3.1 Les banques de séquences généralistes.....	11
2.3.2 Les banques ou bases de données de séquences spécialisées.....	11

CHAPITRE 1 : INTRODUCTION EN INFORMATIQUE

1. Introduction

L'informatique est la science du traitement rationnel de l'information à l'aide d'une machine automatique.

Le terme informatique vient de la contraction des mots information et automatique. Il a été proposé en 1962 et accepté par l'académie française en 1966.

L'informatique est d'une importance capitale en science, dans l'industrie et dans l'administration.

2. L'ordinateur

Une machine électronique permettant de manipuler les données sous forme binaire (0,1)

2.1 Familles d'ordinateurs

- Mainframes (ordinateurs centraux)
- Grande puissance de calcul
- Effectue des traitements lourds
- Utiliser dans les grandes entreprises
- Les ordinateurs PC
- Desktop computers
- Laptop ou notebook
- Tablette PC (ardoise électronique)
- Les centres MultiMedia (media center)
- Les assistants personnels (Personal Digital Assistant)

2.2 Constitution de l'ordinateur

Dans un ordinateur on distingue deux aspects :

1. *Aspect abstrait* : qui désigne la partie logicielle (software) c'est une séquence d'instruction qui indique au matériel informatique les opérations qu'il doit effectuer sur les données.
2. *Aspect concret* : qui désigne la partie matérielle (hardware) c'est l'ensemble des dispositifs électronique constituant l'ordinateur.

3. Hardware

Un ordinateur est composé au minimum d'une unité centrale, d'un moniteur et d'un clavier, mais il est possible de connecter une grande diversité de périphérique : souris, imprimante, scanner, périphérique de stockage externe (hard disc, flash disc, cd...)

Ces périphériques sont divisés en 3 catégories

1. *Périphérique d'entrée* : les informations vont de l'utilisateur vers l'ordinateur ex : clavier, souris...

2. *Périphérique de sortie* : les informations sortent de l'ordinateur vers l'utilisateur ex : moniteur, imprimante...
3. *Périphérique d'entrée/sortie* : les informations circulent dans les deux sens ex : disque dur, CD...

3.1 L'unité centrale :

Contient tous les composants électroniques dont les plus importants sont :

- *Le processeur central CPU* : Le cerveau de l'ordinateur, Il permet de manipuler les informations numériques, il exécute les instructions stockées dans la mémoire, sa vitesse est conditionnée par la fréquence (Hz).
- *La mémoire* : Composant électronique capable de stocker temporairement des données. 2 types de mémoire: Mémoire vive (RAM), Mémoire morte (ROM)

La mémoire vive RAM

Elle permet de faire les 2 opérations de lecture et d'écriture, elle stocke temporairement les instructions en cours d'exécution, c'est une mémoire volatile.

La mémoire morte ROM

Elle permet seulement de faire l'opération de lecture. Elle stocke les informations de base nécessaire pour le fonctionnement de l'ordinateur (BIOS), c'est une mémoire non volatile.

Les unités de mesure de mémoire

- Bit: la plus petite unité de mesure de la mémoire, elle peut contenir soit 0 soit 1.
- Octet (mot) =8 bits
- Kilo octet = 2¹⁰ octets
- Méga octet= 2²⁰ octets
- Giga octet= 2³⁰ octets
- Tétra octet= 2⁴⁰ octets

4. Software

C'est l'aspect abstrait qui désigne la partie logicielle d'un ordinateur.

4.1 Logiciel

est un ensemble des programmes qui permet à un ordinateur ou à un système informatique d'assurer une tâche ou une fonction particulière (logiciel de gestion, logiciel de comptabilité...)

Un logiciel est un ensemble d'informations relatives à des traitements effectués automatiquement par un appareil informatique (ordinateur).

4.2 Types de logiciels

Il existe plusieurs types de logiciels

1. *BIOS* : logiciel de bas niveau qui permettent de faire fonctionner la carte mère.
2. *Système d'exploitation* : Sert d'interface entre le matériel et les logiciels applicatifs, c'est un ensemble de logiciels système central qui contrôle l'utilisation de l'ordinateur par les autres logiciels, MS-DOS, WINDOWS, LINUX ...
3. *Logiciels applicatifs* : le type de logiciel le plus courant, aussi appelé application informatique, ce type de logiciels sont destinés à assister un utilisateur dans une de ces activités, logiciels de traitement de texte, logiciels de calcul, les jeux...

4.3 Programme

Un programme est une séquence d'instructions qui spécifient étape par étape les opérations à effectuer pour obtenir un résultat, il est exprimé sous une forme qui permet de l'utiliser avec une machine comme ordinateur pour exécuter les instructions. Un programme est la forme électronique et numérique d'un algorithme exprimé dans un langage de programmation.

Exemple d'un programme : Calculer la relation: $A*B+C$

1. Donner les valeurs de : A, B, C
2. Calculer D $A*B$
3. Calculer E $D+C$
4. Afficher E sur l'écran.

CHAPITRE 2 : LES BASES DE DONNEES ET LES BANQUES DE DONNEES BIOLOGIQUES

2.1 Introduction

A cause de l'explosion des données biologiques (séquences et structures) les bases de données (BD) ont joué un rôle important dès le début dans la bioinformatique.

Collectionner les données dans un BD permet l'analyse de ces données pour trouver des similarités.

Parce que les données sont importantes pour l'analyse et parce que les données sont disponibles pour tout le monde, on doit s'assurer de la qualité des données

2.2 Définitions

2.2.1 Qu'est-ce qu'une donnée ?

C'est une information quelconque comme, par exemple : voici une personne, elle s'appelle Mohamed. C'est aussi une relation entre des informations : Mohamed *enseigne* les bases de données. Des relations de ce genre définissent des *structures*

2.2.2 Les bases de données

Une base de données est un ensemble structuré de données enregistrées dans un ordinateur et accessibles de façon sélective par plusieurs utilisateurs.

Une Base de données est un gros ensemble d'informations structurées, mémorisées sur un support permanent et pouvant être restituées au biologiste de manière opportune.

Il existe plusieurs types de base de données.

1. *BD Hiérarchiques* : les plus anciennes fondées sur une modélisation arborescente des données.
3. *BD DédDUCTives* : organisation de données sous forme de table et exploitation à l'aide d'un langage logique.
4. *BD Objets* : organisation des données sous forme d'instances de classes hiérarchisées qui possèdent leurs propres méthodes d'exploitation.
5. *BD Relationnelles* : organisation des données sous forme de tables et exploitation à l'aide d'un langage déclaratif (ex: Oracle, mySQL, Access). notre étude va se porter sur ce type des bases de données

Une base de données biologique peut contenir des informations sur des protéines, alors qu'une autre peut contenir des informations sur des gènes, des plasmides, etc.

Il existe plusieurs bases biologiques que vous pouvez consulter sur Internet tels que : nucleotide sequence, proteomics data, structural data, RNA sequence/structure...

Le logiciel qui permet l'interaction avec une base de données s'appelle un système de gestion de base de données (SGBD).

2.2.3 Système de Gestion de Bases de Données SGBD

Un Système de Gestion de Bases de Données (SGBD) est un logiciel de haut niveau qui permet de manipuler les informations stockées dans une base de données.

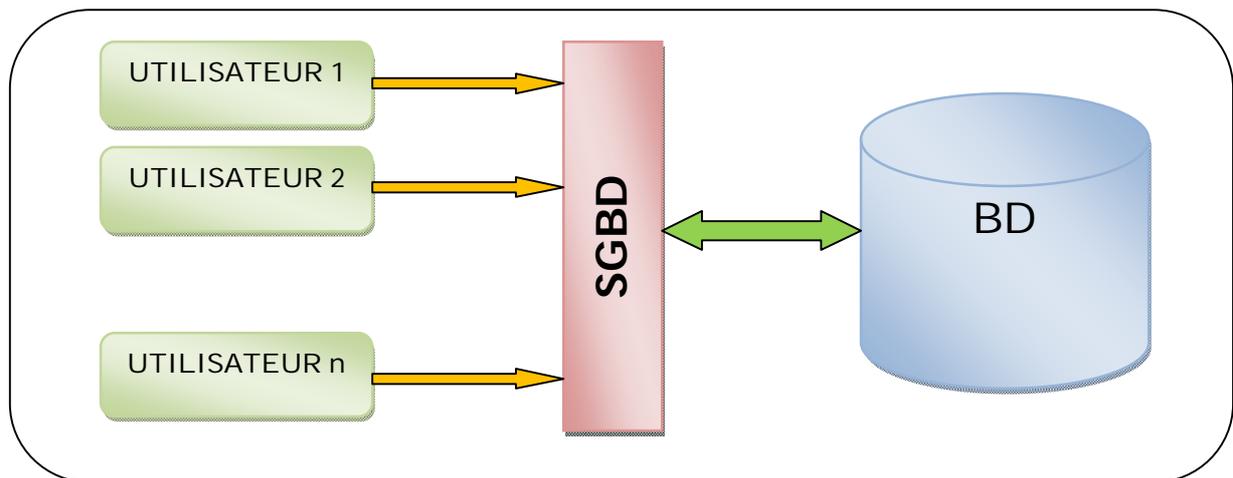


Figure 2.1 : Système de Gestion de Bases de Données

Le SGBD joue le rôle d'interface entre les utilisateurs et la base de données comme montré dans la figure 2.1. Les principales fonctionnalités que peut offrir les SGBDs sont :

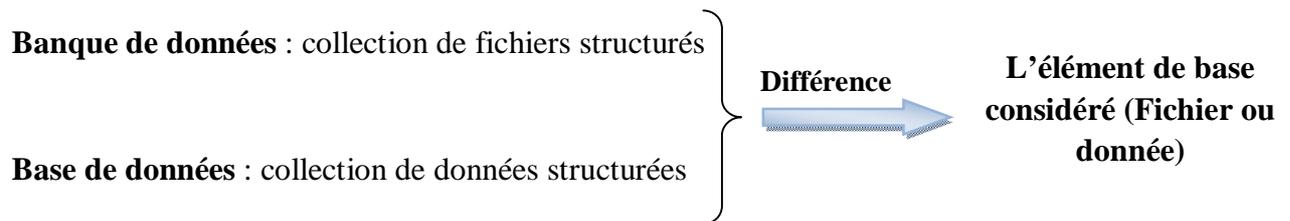
- La Description des données qui seront stockées dans la base.
- La Manipulation de ces données (ajouter, modifier, supprimer des informations)
- La Consultation des données et traiter les informations obtenues (sélectionner, trier, calculer, agréger,...)
- La Définition des protections d'accès (mots de passe, autorisations,...)
- Le Résous des problèmes d'accès multiples aux données (blocages, interblocages)

Exemples de SGBD: ORACLE, MySQL, PostgreSQL, ...

2.2.4 Banque de données

Souvent les termes de banque ou base sont utilisés sans distinction particulière. Mais il existe une différence non seulement pour l'utilisateur mais aussi pour l'implantation informatique de ces dernières.

On définit la banque de données comme un ensemble de données relatif à un domaine défini des connaissances et organisé pour être offert aux consultations d'utilisateurs. Alors que la base de données est un ensemble de données organisé en vue de son utilisation par des programmes correspondant à des applications distinctes et de manière à faciliter l'évolution indépendante des données et des programmes.



2.3 Banque et bases de données biologiques

Il existe un grand nombre de banques ou bases de données d'intérêt biologique. La séquence est l'élément central autour duquel les banques de données se sont constituées. Les séquences biologiques, dès qu'elles ont pu être établies, ont très tôt fait l'objet d'une compilation dans les banques de données. La première compilation de protéines est publiée en 1965 par Margaret Dayhoff : c'est l'Atlas of Protein Sequences qui contient 50 entrées. Nous distinguerons deux types de banques.

2.3.1 Les banques de séquences généralistes

C'est au début des années 80 que les premières banques de séquences sont apparues sous l'initiative de quelques équipes dont la première à l'initiative de Grantham et C. Gautier à Lyon. Très rapidement avec les évolutions techniques du séquençage, la collecte et la gestion des données ont nécessité une organisation plus conséquente. Ainsi, plusieurs organismes ont pris en charge la production de telles bases de données.

Il existe deux types de banques de séquences généralistes tels qu'EMBL, GenBank, DDBJ, les banques de séquences généralistes nucléiques et banques de séquences généralistes protéiques telles que PIR-NBRF, SwissProt...

2.3.2 Les banques ou bases de données de séquences spécialisées

Pour des besoins spécifiques, de nombreuses bases de données spécialisées ont été créées, certaines sont pérennes et continuent d'être développées et mises à jour, d'autres sont laissées à l'abandon et enfin d'autres ont disparu. La nature ainsi que la quantité d'informations sont très variable. Nous avons des banques nucléiques spécialisées, des banques protéiques spécialisées, des banques immunologiques et des banques Structure 2D ou 3D

La figure 2.2 montre la classification des banques des données biologiques, elle résume les différents types des banques et des bases de données : les banques généralistes et les banques spécialisées avec des exemples des banques qui existent.

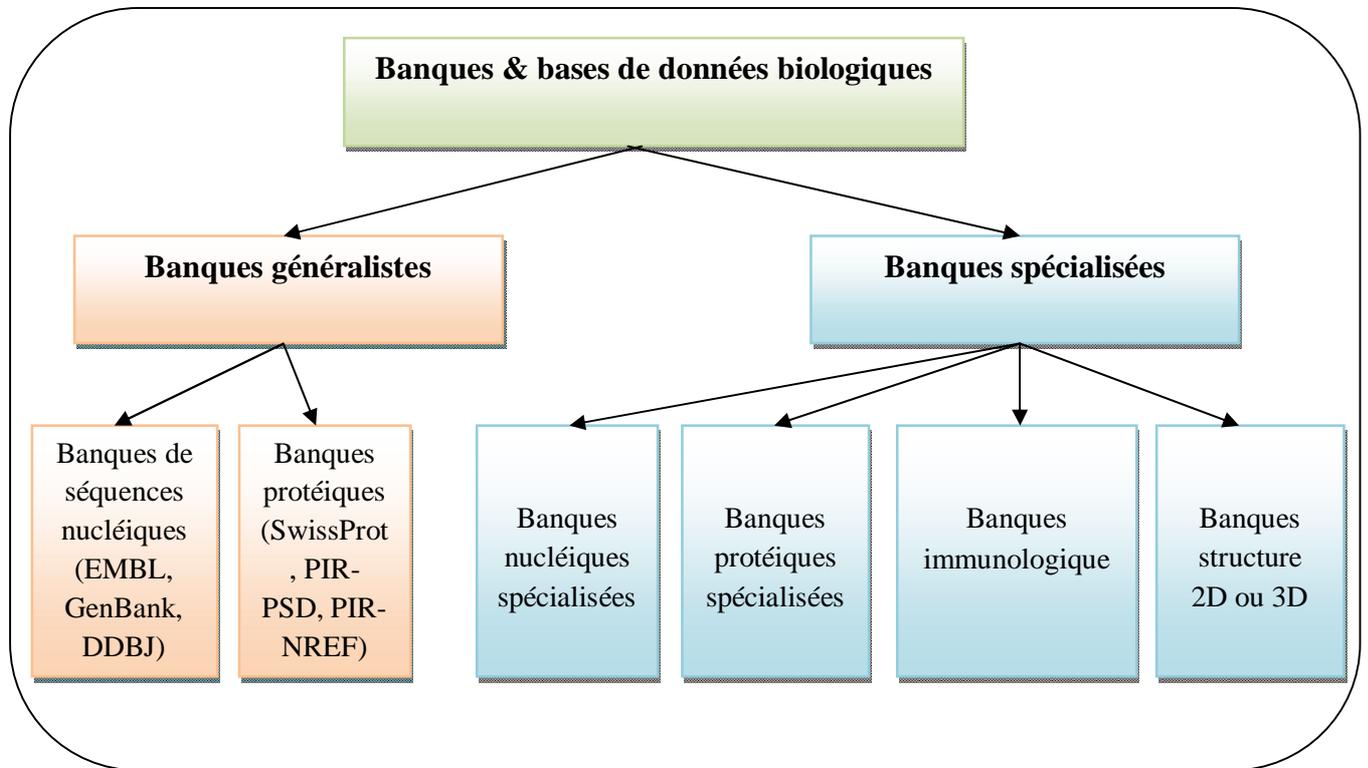


Figure 2.2 : Classification des banques des données biologiques