

**Département d'Informatique
L3 – Systèmes Informatiques**

**Données Semi-Structurées
Chapitre 01
Introduction**

Plan

- Données structurées,
- Données non-structurées (format libre)
- Internet et Web
 - Documents HTML
- Données semi-structurées

Plan

- **Données structurées,**
- Données non-structurées (format libre)
- Internet et Web
 - Documents HTML
- Données semi-structurées

Données structurées

- Les bases de données
 - Une Base de données est un gros ensemble d'informations structurées mémorisées sur un support permanent

Données structurées

- Les bases de données
 - Contrer les inconvénients de l'utilisation des fichiers,
 - Architecture centralisée,
 - Grande « quantité » de données,
 - Partie statique d'un système d'information,
 - Utilisation via SGBD

Données structurées

- Les systèmes de gestion de bases de données
 - 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,

Données structurées

- Les systèmes de gestion de bases de données
 - Contrairement au niveau fichier (SGF) :
 - Ouvrir/Fermer
 - Lire (tableau binaire)/Ecrire
 - Un SGBD offre :
 - LMD : Langage de Manipulation de Données,
 - LDD : Langage de Définition de Données,
 - Grant/Revoke

Données structurées

- Les systèmes de gestion de bases de données
 - Stockage basé sur les enregistrements
 - Une structure de données pour chaque table (fichier),
 - Fichiers binaires lisibles seulement par le SGBD,
 - Et l'échange des données ?
 - Dump ?

Données non-structurée

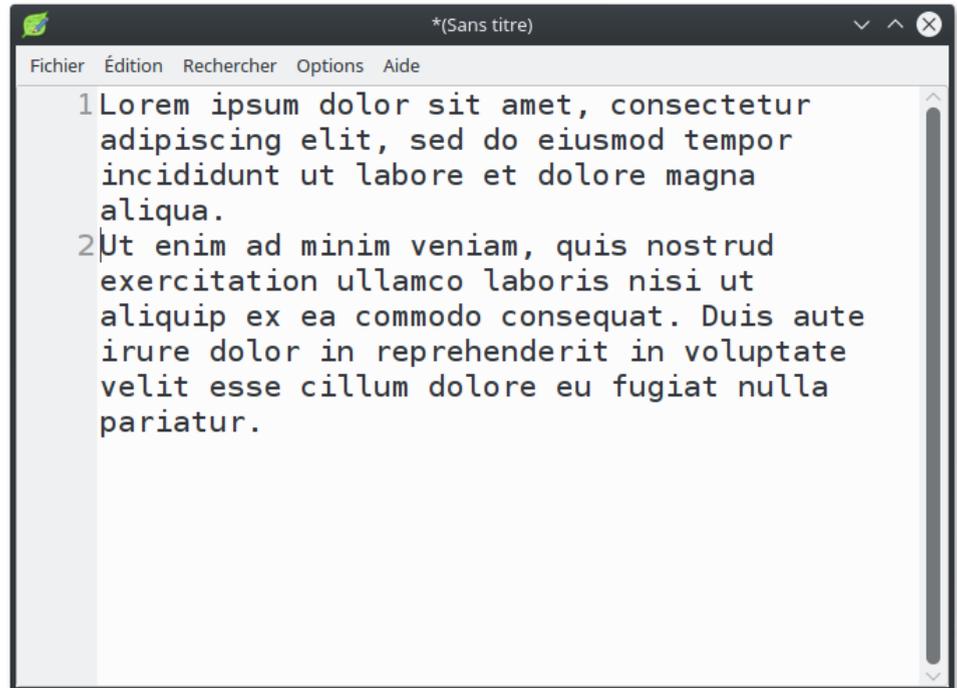
- Données au format libre :
 - Fichiers textes (format la plus utilisée) :

Plan

- Données structurées,
- **Données non-structurées (format libre)**
- Internet et Web
 - Documents HTML
- Données semi-structurées

Données non-structurée

- Données au format libre :
 - Fichiers textes (format la plus utilisée) :



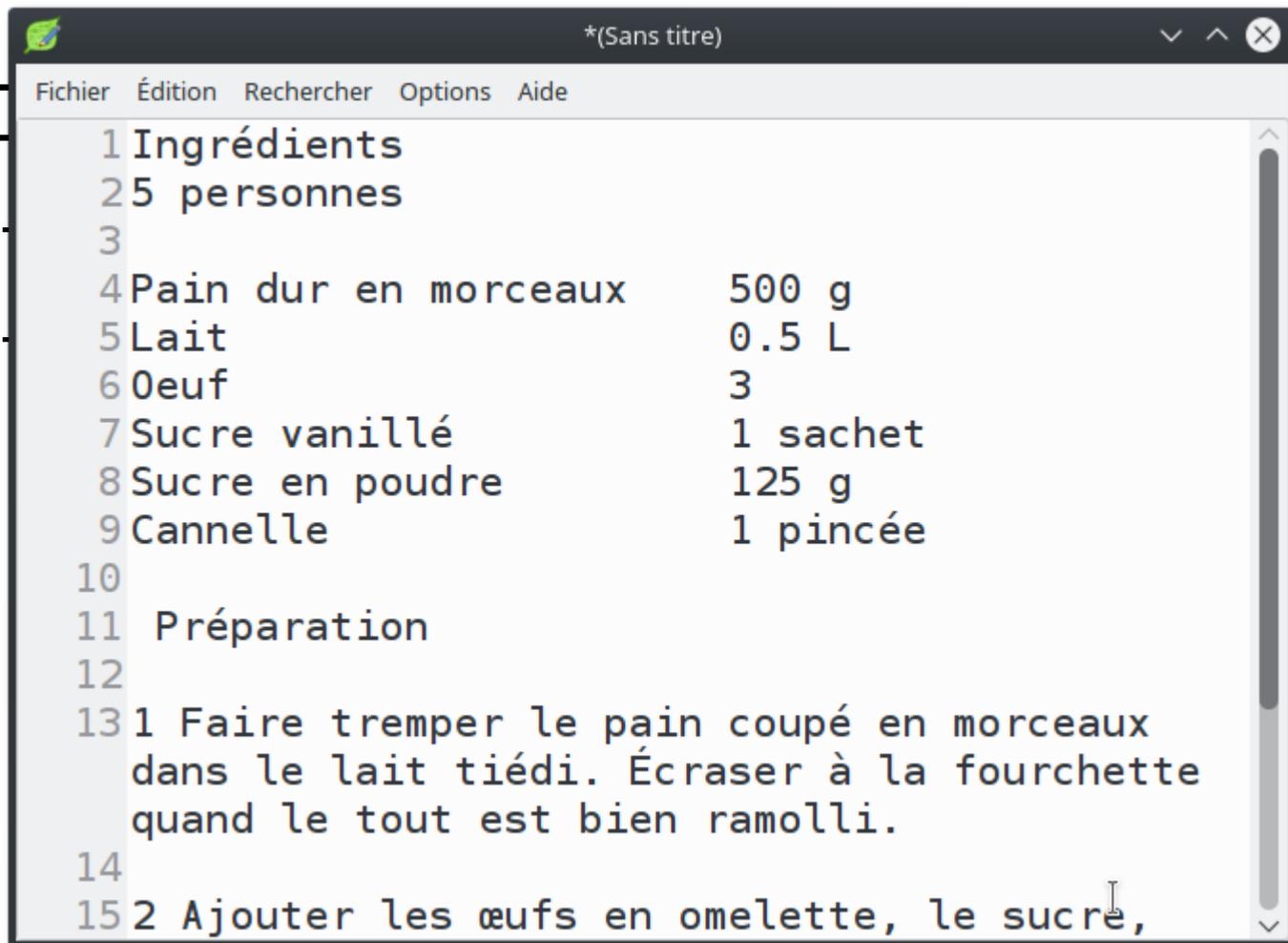
A screenshot of a text editor window titled "(Sans titre)". The window has a menu bar with "Fichier", "Édition", "Rechercher", "Options", and "Aide". The text area contains two paragraphs of Lorem Ipsum text, numbered 1 and 2. Paragraph 1: "Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua." Paragraph 2: "Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur." The text is displayed in a monospaced font.

Données non-structurée

- Données au format libre :
 - Lisible par les humains,
 - Difficile à lire et à traiter par les machines,

Données non-structurée

-

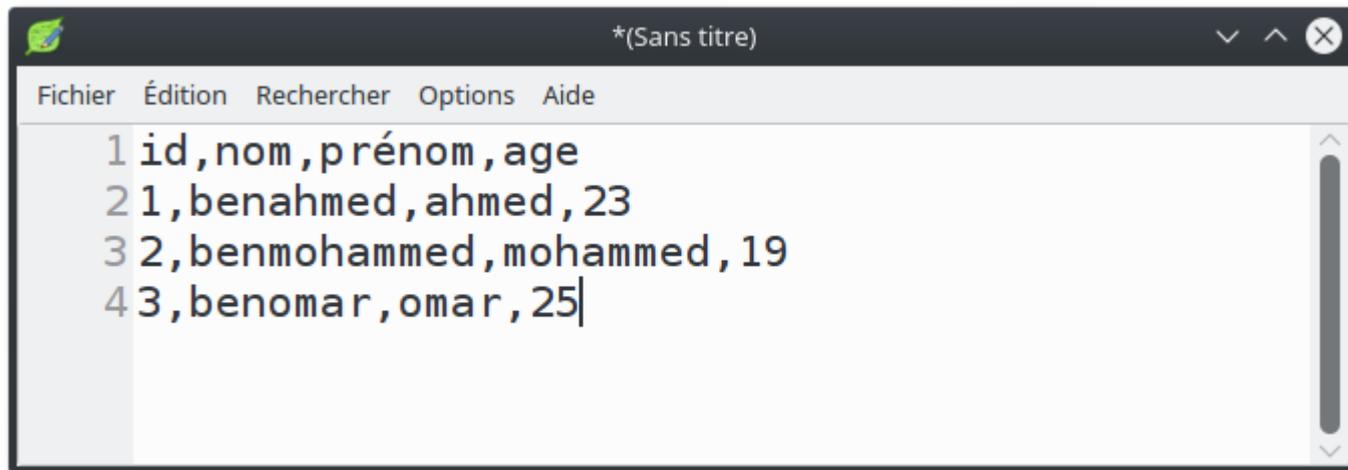


```
1 Ingrédients
2 5 personnes
3
4 Pain dur en morceaux      500 g
5 Lait                      0.5 L
6 Oeuf                      3
7 Sucre vanillé             1 sachet
8 Sucre en poudre          125 g
9 Cannelle                  1 pincée
10
11 Préparation
12
13 1 Faire tremper le pain coupé en morceaux
   dans le lait tiédi. Écraser à la fourchette
   quand le tout est bien ramolli.
14
15 2 Ajouter les œufs en omelette, le sucre,
```

es,

Données non-structurée

- Données au format libre :
 - Fichiers CSV
 - Comma Separated Values
 - Structuration ?



A screenshot of a text editor window titled "(Sans titre)". The window has a menu bar with "Fichier", "Édition", "Rechercher", "Options", and "Aide". The main text area contains the following CSV data:

```
1 id,nom,prénom,age
2 1,benahmed,ahmed,23
3 2,benmohammed,mohammed,19
4 3,benomar,omar,25|
```

Plan

- Données structurées,
- Données non-structurées (format libre)
- **Internet et Web**
 - Documents HTML
- Données semi-structurées

Internet et Web

- Internet
 - Réseau mondial,
 - Réseau des réseaux,
 - Infrastructure de communication :
 - Transmission de paquets,
 - Routage,
 - Accusé de réception,
 - TCP/IP

Internet et Web

- Web
 - Un service d'Internet,
 - Protocole : HTTP
 - Hyper Text Transfert Protocol
 - Langage : HTML
 - Basé sur SGML (1986)
 - Standard Generalized Markup Language
 - Transfert par Document.

Internet et Web

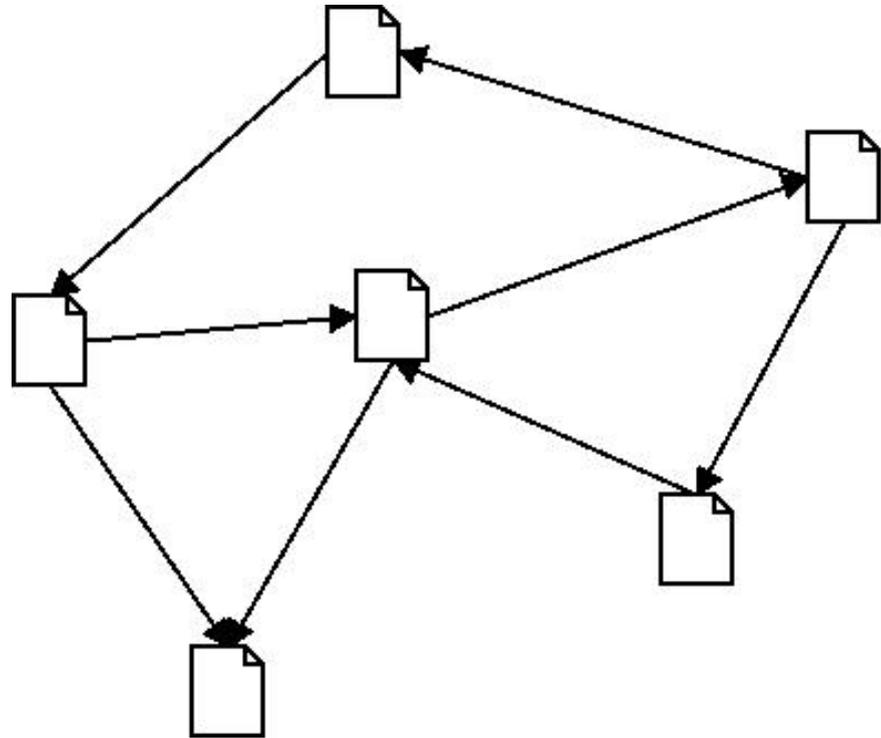
- Web
 - SGML :
 - Langage de description,
 - Séparer la structure logique de la représentation
 - Permet de visualiser le document sur différents médias
 - Introduction de la notion des feuilles de styles.

Internet et Web

- Web
 - HTML :
 - Même principes (balise, feuilles de styles, ...)
 - Création des documents pour visualisation dans le navigateur,
 - Support de l'hypertexte
 - « a database (text) format in which information related to that on a display can be accessed directly from the display »
 - Enrichi pour supporter l'hypermédia

Internet et Web

- Web
 - Hypertext :
 - Non-linéaire :



Plan

- Données structurées,
- Données non-structurées (format libre)
- Internet et Web
 - **Documents HTML**
- Données semi-structurées

Documents HTML

- `<!DOCTYPE html>`
- `<html>`
 - `<head>`
 - `<title>Exemple</title>`
 - `</head>`
 - `<body>`
 - `<p>Ceci est un paragraphe.</p>`
 - `</body>`
- `</html>`

Plan

- Données structurées,
- Données non-structurées (format libre)
- Internet et Web
 - Documents HTML
- **Données semi-structurées**

Données semi-structurées

- Les données semi-structurées sont des données qui ne respectent pas une structure fixe (comme les tables d'une base de données relationnelles) mais qui contiennent des balises qui séparent les différents éléments sémantiques.
- Au milieu entre les données structurées et non-structurées.

Données Semi-Structurées

Chapitre 01

Introduction