SYSTÈMES D'INFORMATION DE GESTION · TERMINALE STMG

# Structurer et indexer les documents

6 notions, l'essentiel à retenir et les pièges à éviter.

## 1 Numériser, structurer, indexer

- Standardiser : suivre des règles communes pour que les documents circulent entre personnes, logiciels et organisations.
- Numérisation : scanner puis OCR pour obtenir un texte exploitable.
- Structurer : découper le contenu en parties identifiées. Indexer : le décrire par des métadonnées.

## 2 Le langage XML

- Élément : balise ouvrante, contenu, balise fermante.
- Attribut : nom="valeur", dans la balise ouvrante.
- Arborescence : éléments parents et enfants, sous une racine unique.

## 3 Bien formé, valide

**valide = bien formé + conforme à la DTD ou au schéma XML**

- Bien formé : une racine, balises fermées et bien imbriquées, même casse, attributs entre guillemets.
- DTD ou schéma XML : éléments permis, ordre, attributs obligatoires ; le schéma contrôle aussi les types.

## 4 HTML et CSS

- head : titre, encodage, métadonnées, lien vers le CSS. body : titres h1 à h6, paragraphes p, liens a.
- HTML décrit le contenu, CSS la présentation : une feuille de style met en forme tout le site.

## 5 Métadonnées, RSS, référencement

- Métadonnées : titre, auteur, date, mots-clés, description.
- Flux RSS : fichier XML des dernières publications, lu par un agrégateur (syndication).
- Référencement : titre et description précis, contenu structuré, liens entrants.

## 6 La GED

- Cycle de vie : capture, indexation, stockage, diffusion, archivage.
- Moteur de recherche, droits d'accès par profil, gestion des versions (traçabilité).

### Pièges à éviter

- Confondre bien formé et valide : un document valide est d'abord bien formé.
- Oublier que XML distingue majuscules et minuscules.
- Écrire un attribut dans la balise fermante.
- Mettre la présentation dans le HTML au lieu du CSS.
- Confondre stockage et archivage.