

Introduction

Damien Nouvel



Plan

1. Généralités
2. Document, textes et connaissances

Séances et modalités de contrôle

- ▶ **Un semestre**
 - Structuration et présentation de documents
- ▶ **Séances**
 - 12 séances
 - Chaque séance : **cours** et **exercices**
- ▶ **Modalités de contrôle** (semestre)
 - Un **projet** (50%)
 - Un **examen final** (50%)

Contenu du cours

▸ Progression

- Introduction générale
- Formats et codages des documents
- Documents pour le Web (HTML)
- Structuration du document (XML)
- Génération de documents

Plan

1. Généralités
2. Document, textes et connaissances

Définir le document

▸ Définitions

- **Enseignement**, oral ou écrit, **transmis** par une pers. [CNRTL]
- **Support** et **information** enregistrée (**persistante** ?)
[Wikipedia]
- Pièce **écrite** servant d'**information** (preuve ?) [Larousse]
- Emprunt au latin *documentum* (enseignement), dérivé de *docere* (enseigner, informer) [Wiktionary]

⇒ Importance de la notion d'**enregistrement** (résiste au temps)

⇒ Lien entre **information** et **medium**

Diversité des documents

- ▶ Historiquement : **papier**, **écrit**, administratif ...mais aussi
 - Oral (document **audio** ou **vidéo**)
 - Support **numérique** (digital)
 - Contenus non-langagiers : partitions, schémas, peintures
 - Toutes utilisations : historique, gestion de projet, scolaire, etc.

 - ▶ **Accessibilité** du document pas **immédiate**
 - **Disponibilité** (bibliothèques, librairies, en ligne)
 - **Support** +/- lisible : papier, photo, bande, CD-ROM, fichier
 - **Langue** et **dialecte** d'écriture (système typographique)
 - Contenu +/- difficile (lecture, effort **cognitif**)
- ⇒ Pour ce cours : **écrits** et **numériques**
- ⇒ Quels **langages**, quels **codes**

Enjeux de la représentation des contenus

▶ Du **caractère** au **livre**

- Caractère / lettre (morphème, phonème)
- Mot (token, syntagme)
- Phrase (énoncé, proposition)
- Paragraphe (segment thématique, argument)
- Section, chapitre, livre, volume ...encyclopédie

⇒ Structure **linéaire** explicite (exceptions : index, thesaurus)

- ▶ Documents numériques : même organisation **par défaut**, mais peuvent aussi être accédés par d'autres moyens
 - **Moteur de recherche** : requête et index
 - (Hyper-) **Liens** : navigation sur la toile

⇒ **Dématérialisation, numérisation** (quelle preuve?)

⇒ Nouvelles structurations (**sites, bases de données**)

Outils pour la création de documents

- ▶ D'un point de vue technico-historique (support)
 - Papyrus, parchemin, papier,
 - Documents texte (ASCII)
 - Documents mis en forme (**HTML**, Word, Latex)
 - Connaissances structurées (**XML**, RDF, SQL)
- ▶ Relation entre **documents** et **connaissances**
 - Utilisation de **thésaurus** / **terminologies**
 - Organisation sous forme de **graphes** (Wikipedia)
- ⇒ Enregistrement linéaire (disque) mais intervention d'un **calcul**
 - **Compilation**
 - **Interprétation**
 - **Présentation**
 - **Rendu graphique**
- ⇒ Une connaissance peut prendre de multiples **formes**

Utilisation de documents

► Domaines

- Bibliothèques / archives
- Législatif
- Sciences (santé)
- Culture
- ...

► Métiers

- Gestion électronique de documents
- Indexation / classification de documents
- Représentation de connaissances
- Ingénierie de systèmes d'information
- Ingénierie multilingue