

INU3011 Documents structurés

Cours 1

Bienvenue - Introduction

Plan

- Bienvenue !
- Organisation du cours
- Visite guidée des ressources
- Introduction à XML

Bienvenue !

- Qui je suis ?
- Qui êtes-vous ?
- Préalable: INU1001 ou l'équivalent

Horaire et structure

- Lundi 16h-18h50
- Local C-2043 Pavillon Lionel-Groulx
- Structure :
 - Questions, discussion sur les lectures
 - Présentation "magistrale"
 - Volet pratique "laboratoire"
 - Exercices / protocoles de la semaine
 - Discussions, coaching individuel ou en équipe

Les mini-quiz

- Les mini-quiz sont sur StudiUM
- **Individuels**
- Durée entre 35 et 50 minutes
- À faire au **moment de votre choix** le **mardi, mercredi ou jeudi** de la semaine où ils sont programmés

StudiUM

- Outil principal de communication
- Section "Nouvelles"
- Forum de discussion "Vos questions/réponses"
- Organisation par semaine
 - Semaines ajoutées au fur et à mesure

Visite guidée des ressources (1/3)

- Le [site Web du cours](#)
 - Plan de cours
 - Évaluateur de réussite
 - Nombreux exemples
 - Aussi accessibles *via* des liens dans les textes ou les présentations PDF
 - Lectures, protocoles et exercices sont sur le site du cours, mais accédés par StudiUM

Visite guidée des ressources (2/3)

- Laboratoire d'informatique virtuel de l'EBSI
 - Postes en Windows 11
 - Préconfigurés adéquatement pour le cours
 - Préinstallés :
 - oXygen + autres requis en fin de trimestre
 - Nécessite de travailler en permanence avec une connexion VPN active

Visite guidée des ressources (3/3)

- Logiciel oXygen, excellent éditeur XML
 - Utilisation obligatoire dans le cours
 - Installation sur votre ordi personnel
 - **Fortement recommandée**
 - Permise par la licence campus UdeM
 - Procédure d'installation sur page StudiUM du cours *et non* via la ~~Logithèque UdeM~~

"Introduction au cours"

- Lecture donnant des informations pratiques sur le cours, les prérequis et les choix technologiques
- Inclut une capsule vidéo :
 - [Introduction à oXygen](#) (8 min.)

Premier tour d'horizon de XML

- Ensemble de quatre textes avec index
- "Préface" *Comment lire ce Tour d'horizon?*
 - Capsule vidéo d'accompagnement (12 min.)
 - Définit "Lecture active"
- Notions syntaxiques XML de base
- Sous-ensemble "minimum viable" de XML
- Lecture active fortement préconisée

XML, c'est quoi?

XML, c'est un...

Format de documents numériques (1/2)

- Ensemble de règles ou conventions...
- Permettant de représenter de l'information sous forme de fichiers informatiques
 - Donc, représentée en binaire (0 et 1)

Format de documents numériques (2/2)

- Deux niveaux de règles :
 - **Syntaxique** : comment l'information binaire doit être *structurée* dans les fichiers
 - **Sémantique** : comment l'information binaire doit être *interprétée* pour être perceptible par l'humain

Format de documents *structurés*

- Extensible Markup Language
- Utilise des règles de balisage à l'intérieur de fichiers texte Unicode-UTF-8*
 - Donc : tous les documents XML **sont aussi des fichiers texte**
 - Des *balises* (à la `<html>`) identifient début et fin des éléments constitutifs du document

**D'autres jeux de caractères sont aussi possibles*

Format *normalisé*

- Recommandations du W3C
 - XML 1.0
 - 1^e éd. février 1998
 - 5^e éd. novembre 2008
 - XML 1.1
 - 1^e éd. février 2004
 - 2^e éd. septembre 2006
 - Ajouts très mineurs par rapport à XML 1.0
 - Encore peu utilisé

XML définit deux niveaux de conformité syntaxique (1/3)

1. Le bien-formé (*well-formedness*)
2. La validité
 - Les règles du bien-formé sont *obligatoires* pour tout document XML
 - Certains documents se conformant à des règles additionnelles de forme sont *valides*

XML définit deux niveaux de conformité syntaxique (2/3)

- Les règles additionnelles de validité varient d'un type de document à l'autre
 - Elles déterminent notamment les *noms d'élément* permis et leur signification
- Un ensemble de règles additionnelles pour un certain type de document constitue ce qu'on appelle un *modèle XML*
 - ou une DTD, pour *Document Type Definition*

XML définit deux niveaux de conformité syntaxique (3/3)

- Un modèle XML définit donc un format de document numérique, *plus précis et spécifique* que les règles du bien-formé
- Plusieurs modèles XML sont *eux-mêmes normalisés* au sein de différentes instances ou communautés d'intérêt

Où XML est-il utilisé?

- Trois exemples où le XML est visible
 - [Archives UdeM](#)
 - [Œuvres de Carl Maria von Weber](#)
 - [Lois canadiennes](#)
 - [California Bar Examination Pass List](#)
- Un exemple où le XML n'est pas visible par les utilisateurs
 - [Portail européen des archives](#)

Pourquoi rendre le XML disponible publiquement ?

- Pour une organisation ou un groupe d'intérêt, rendre disponible publiquement son information en XML, c'est faire la preuve d'une volonté réelle de partage et de collaboration
- Permet l'exploitation de l'information en aval sans entente préalable : une façon de maximiser la diffusion

XML est partout !

- Juste pour voir, recherchez les fichiers nommés *.xml sur votre ordi... *surprise !*
 - Par exemple, dans `C:\ProgramData`
- Montre d'entraînement, téléphones, etc.
- [XML Essentials](#) (une des lectures de cette semaine) :
 - "It is the age of XML: it is everywhere."

En quoi ce cours peut vous être utile dans votre carrière?

Raison 1 – Omniprésence

- La quantité d'information disponible en XML est énorme et croissante
- Les chances que vous soyez exposés à de l'information en XML sont très grandes
- Plus vous en connaissez sur XML, mieux vous serez en mesure de traiter adéquatement cette information

Raison 2 – Modélisation

- Le design de tout système d'information est fondé sur la *modélisation de l'information*, qui détermine les traitements possibles
- XML = meilleur cadre possible pour comprendre l'impact des décisions de modélisation sur les traitements possibles
- Comprendre ce lien fera de vous un meilleur interlocuteur avec les équipes de développement, quelles que soient les technologies utilisées

Raison 3 – Autonomie

- Pour peu que vous œuvriez au sein d'une organisation technologiquement avant-gardiste, vous pourriez avoir la possibilité d'intervenir vous-même sur des systèmes XML :
 - modélisation, traitement avec XSLT, publication sous différentes formes, bases de données XML