

INU3011 Documents structurés

Plan de cours Hiver 2022

[Yves MARCOUX](#) - [EBSI](#) - [Université de Montréal](#)

Table des matières

1. [Informations générales](#)
 2. [Objectifs](#)
 3. [Description](#)
 4. [Évaluation](#)
 5. [Calendrier des activités](#)
 6. [Règlements, politiques](#)
 - 6.1 [Évaluation avec seuil](#)
 - 6.2 [Modes de communication](#)
 - 6.3 [Autres règlements et politiques](#)
 7. [Bibliographie](#)
 - 7.1 [Lectures obligatoires \(par ordre de lecture dans le cours\)](#)
 - 7.2 [Lectures recommandées](#)
 - 7.3 [Lectures utiles](#)
 - 7.4 [Diverses sources et ressources utiles](#)
-

1. Informations générales

- **Horaire et locaux :** Lundi de 16h à 18h50 au C-2027 du [Pavillon Lionel-Groulx](#). Certains travaux de laboratoire s'effectuent à l'intérieur de la même plage horaire et au même endroit.
- Cours de trois crédits, ce qui correspond à 135 heures de travail pour l'étudiante¹.
- **[Description officielle du cours](#) :**

Formats de documents et langages de balisage. Historique. Concepts de base. XML et normes périphériques. Modélisation, validation, stylage. Chaînes de traitement et méthodologies d'implantation. Bases de données XML. Préalable : INU1001 ou l'équivalent.
- **Professeur :** [Yves MARCOUX](#) <y marcoux@gmail.com>.
- **[Site Web du cours](#)**
- **[Page StudiUM du cours](#)**

• Modalités :

+/- large ↔

- Les rencontres du lundi 16h ont lieu en classe, en présentiel.
- Au besoin, certaines rencontres pourraient exceptionnellement être en virtuel, sur Zoom. Le cas échéant :
 - Vous serez avisée du changement le plus tôt possible par courriel et via la [page StudiUM du cours](#).
 - L'horaire habituel sera maintenu, soit de 16h à 18h50.
 - Pour la partie laboratoire du cours, il sera possible d'utiliser le [Laboratoire d'informatique virtuel de l'EBSI](#).
 - La partie magistrale du cours (précédent la partie laboratoire) sera enregistrée et pourra être visionnée en différé.
 - Le lien Zoom sera mis sur la [page StudiUM du cours](#) quelques minutes avant le début de la rencontre.
 - Vous pourrez joindre la rencontre virtuelle de l'endroit de votre choix, mais il vous sera possible de le faire à partir du local habituel du cours, soit le C-2027 du [Pavillon Lionel-Groulx](#) (à moins que des directives sanitaires empêchent l'accès physique à cette salle).
Prévoir un casque d'écoute au C-2027.
- Les mini-quiz se font en asynchrone sur [StudiUM](#). Ils sont d'une durée d'entre 30 et 50 minutes et peuvent être effectués de l'endroit et au moment de votre choix (à l'intérieur d'une plage horaire définie), mais *doivent* être faits sur un *ordinateur* et non un téléphone ou une tablette.

¹: Vous remarquerez que, dans ce cours, la question des épécènes est habituellement réglée par l'utilisation du féminin.

2. Objectifs

À la fin du cours, l'étudiante démontrera une compréhension approfondie des enjeux liés à l'utilisation des documents structurés pour la gestion de l'information documentaire. Plus spécifiquement, l'étudiante :

- Démontrera une maîtrise approfondie des concepts et des principes de base des documents structurés.
 - Pourra concevoir des modèles de documents structurés en fonction de besoins donnés.
 - Pourra créer des feuilles de styles en fonction de besoins donnés.
 - Pourra planifier des chaînes de traitement de documents structurés en fonction de besoins donnés.
 - Démontrera une bonne compréhension du processus d'implantation d'un système de documents structurés.
-

3. Description

Tel qu'indiqué ci-dessus, l'objectif principal de ce cours est de faire prendre conscience aux participantes des avantages fondamentaux liés à l'utilisation des documents structurés pour la publication d'information documentaire, ainsi que des défis qu'ils posent aux concepteurs de systèmes d'information. Un objectif secondaire est la familiarisation avec certains types de fonctionnalité des outils de traitement des documents structurés.

Le cours est presque entièrement élaboré autour de XML (*Extensible Markup Language*), le plus important format normalisé de documents structurés existant actuellement. Outre une compréhension approfondie des concepts et principes de base des documents structurés, on vise un niveau de compétence permettant, dans un contexte spécifique bien défini, de concevoir des DTD (*Document Type Definitions*) et/ou des schémas XML (du W3C), de créer des feuilles de styles CSS et XSLT, de sélectionner des outils, de convertir des documents XML vers d'autres formats et de les publier sur le Web, directement et par le truchement d'une base de données XML.

Le cours *INUI001 Introduction à l'information numérique* est préalable. Ce préalable est précisé plus avant dans la lecture [Introduction au cours](#), à faire la première semaine de cours.

Une des principales activités est un projet consistant à modéliser des documents, à en saisir quelques-uns en XML et à les publier de diverses façons :

- Sur le Web, en XML, en HTML et dans une base de données XML.
- En format LaTeX, pour en tirer du prêt-à-papier (PDF).

La nature des documents sera discutée en classe.

Une conférencière est habituellement invitée à parler d'une utilisation concrète des documents structurés autour du Cours 12.

4. Évaluation

Évaluation

Activité	Pondération
✂ TP 1 : Rallye de bienvenue*	5%
🔗 Trois « mini-quiz » sur StudiUM (10% + 10% + 10%)*	30%
✂ TP 2 : Encodage d'un texte*	10%
✂ TP 3 : Création d'un modèle XML*	10%
✂ TP 4 : Modélisation préliminaire	15%
✂ TP 5 : Modélisation et stylage	20%
✂ TP 6 : Publication Web	10%

*Les évaluations marquées d'un astérisque sont individuelles. Les TP 4, 5 et 6 constituent le projet mentionné précédemment. Ils sont réalisés en équipes de deux personnes, mais celles qui le souhaitent peuvent aussi le faire individuellement (SVP, en aviser le professeur le plus tôt possible).

La note de passage à un cours s'exprime en cote littérale. Elle est de « D » pour les étudiants inscrits à un programme de premier cycle et de « C » pour les étudiants inscrits à un programme des cycles supérieurs (même dans un cours de premier cycle).

L'évaluation avec seuil (voir [section à cet effet](#) ci-après) est appliquée dans ce cours.

Voir [ici la méthode de calcul de la note finale du cours](#).

+/- large ↔

Pour vous aider à suivre l'évolution des résultats de vos évaluations au cours du trimestre, et notamment connaître les notes requises dans les évaluations à venir pour atteindre une cote donnée (en tenant compte ou non du seuil), un [Évaluateur de réussite](#) est disponible. N'hésitez pas à l'utiliser.

Le respect des directives données dans les protocoles de travaux est un critère d'évaluation.

5. Calendrier des activités

Note : Les contenus et leur répartition entre les cours seront ajustés en cours de session.

Calendrier des activités

Date – Cours	Contenu
2022-01-10 – C1	Logistique et fonctionnement du cours, tour guidé des ressources. XML, c'est quoi ? En quoi ce cours peut vous être utile dans votre carrière ?
2022-01-17 – C2	Historique de XML. Structuration de l'information en XML.
2022-01-18 mardi 23:55	✂ 5% Remise : TP 1 Rallye de bienvenue (travail individuel)
2022-01-24 – C3	Entités prédéfinies et entités caractère. oXygen et caractères spéciaux, émojis.
2022-01-25 semaine	🔗 10% Mini-quiz 1
2022-01-31 – C4	Validité en XML : mécanismes, bénéfices, expérience-auteur, modèles normalisés, écosystème autour d'un modèle. Présentation du TP 2.
2022-02-07 – C5	Langage des DTD.
2022-02-08 mardi 23:55	✂ 10% Remise : TP 2 Encodage d'un texte (travail individuel)
2022-02-14 – C6	Modélisation : nature et démarche. Travailler avec un modèleXML3011. Principes de modélisation (début). Présentation du TP 3.
2022-02-15 semaine	🔗 10% Mini-quiz 2
2022-02-21 – C7	Principes de modélisation (fin). Points spécifiques de modélisation. Présentation du TP 4, aperçu des TP 5 et 6.
2022-02-22 mardi 23:55	✂ 10% Remise : TP 3 Création d'un modèle XML (travail individuel)
2022-02-28	<i>Semaine d'activités libres — Pas de cours</i>
2022-03-07 – C8	Traitement de documents XML : modèle de données XPath, XPath.
2022-03-08 mardi 23:55	✂ 15% Remise : TP 4 Modélisation préliminaire

2022-03-14 – C9	XSLT. Présentation du TP 5.
2022-03-15 <i>semaine</i>	Rencontres séparées par équipe : retour sur TP 4.
2022-03-21 – C10	Notions et directives complémentaires. Travail en laboratoire.
2022-03-28 – C11	LaTeX; production de LaTeX et de PDF à partir de XML.
2022-04-04 – C12	Chaînes de traitement XML. Méthodes de création de documents XML : exportation, conversion, « upconversion ». Conférencière invitée (événement et date à confirmer).
2022-04-05 <i>semaine</i>	🔗 10% Mini-quiz 3
2022-04-11 – C13	Publication Web de documents XML : statique, dynamique. Bases de données XML. Présentation du TP 6.
2022-04-18	<i>Lundi de Pâques — Pas de cours</i>
2022-04-25 – C14	Les espaces de noms XML (<i>XML Namespaces</i>). Aperçu de la TEI.
2022-04-29 <i>vendredi 23:55</i> <i>Dernier jour de la session</i>	✘ 20% Remise : TP 5 Modélisation et stylage ✘ 10% Remise : TP 6 Publication Web

+/- large ↔

6. Règlements, politiques

6.1 Évaluation avec seuil

L'*évaluation avec seuil* est appliquée. Pour réussir le cours, il faut donc obtenir une moyenne pondérée supérieure ou égale à la [note de passage](#) non seulement sur l'ensemble des activités évaluées du cours, mais également sur le *sous-ensemble* des activités évaluées du cours qui sont *réalisées individuellement (et non en équipe)*. Si la moyenne pondérée obtenue sur le sous-ensemble des activités évaluées du cours réalisées individuellement est *inférieure à la note de passage*, alors cette moyenne pondérée est considérée comme étant la note globale obtenue pour le cours; autrement dit, les activités évaluées réalisées en équipe ne sont alors pas comptées et le cours est échoué.

6.2 Modes de communication

Le mode de communication privilégié du professeur vers vous est la section *Messages du professeur* de la [page StudiUM du cours](#).

Le professeur peut également vous joindre par courriel. L'adresse utilisée sera celle enregistrée dans [Mon accès UdeM](#); assurez-vous qu'elle soit valide et fonctionnelle en tout temps. SVP, lire votre courriel tous les jours.

Le mode de communication privilégié pour joindre professeur est le courriel. L'adresse courriel du professeur est <y marcoux@gmail.com>. **SVP, inscrire la mention [INU3011] (incluant les crochets) au début de la ligne sujet de votre message.**

Vous êtes aussi encouragée à poser des questions dans le forum *Vos questions-réponses* sur la [page StudiUM du cours](#).

6.3 Autres règlements et politiques

+/- large ↔

Tous les règlements, politiques et directives énoncés dans le [Guide étudiant du 1^{er} cycle](#) s'appliquent, incluant le *Code d'honneur de l'EBSI*. Une attention particulière est à porter aux éléments suivants :

Travaux en équipe

Pour les travaux réalisés en équipe, le professeur se réserve le droit d'évaluer séparément chaque membre d'une équipe.

Règlement disciplinaire sur le plagiat ou la fraude concernant les étudiants

Toute infraction au règlement sur le plagiat ou la fraude sera traitée suivant la procédure indiquée dans le règlement.

Retard dans la remise des travaux

Tout retard non justifié dans la remise d'un travail sera sanctionné : 5% de la note est retranché par jour de calendrier de retard jusqu'à un maximum de 35%; à la 8^{ième} journée de calendrier, la note F ou zéro (0) est attribuée.

Qualité de la langue

Un maximum de 10% de la note globale d'un travail pourra être retranché pour mauvaise qualité de la langue dans les travaux (ne s'applique pas aux quiz).

7. Bibliographie

[Réserve du cours à la BLSH](#)

7.1 Lectures obligatoires (par ordre de lecture dans le cours)

- MARCOUX, Yves. « Introduction au cours ». En ligne <<https://cours.ebsi.umontreal.ca/INU3011/horizon/intro-au-cours.html>>.
- W3C. *XML Essentials*. En ligne <<https://www.w3.org/standards/xml/core>>.
- MARCOUX, Yves. [Premier tour d'horizon de XML](#). Quatre textes d'introduction à XML.
- MARCOUX, Yves. [Travailler avec un modèle XML 3011](#).
- MARCOUX, Yves. « Attributs et modélisation en XML ». En ligne <<https://cours.ebsi.umontreal.ca/INU3011/matthem/attrib-modelisation.html>>.
- MARCOUX, Yves. « Stylage des éléments vides en XSLT ». En ligne <<https://cours.ebsi.umontreal.ca/INU3011/matthem/XSLT-stylage-elements-vides.html>>.
- LaTeX Project. "An introduction to LaTeX." *LaTeX project site*, février 2008. En ligne <<https://latex-project.org/intro.html>>.
- UNWALLA, Mike. "LaTeX: an introduction." *Communicator*, printemps 2006. En ligne <<https://www.techscribe.co.uk/ta/latex-introduction.pdf>>.
- TEI Consortium. [TEI: History](#).

7.2 Lectures recommandées

- BITOUZÉ, Denis; CHARPENTIER, Jean-Côme. *LaTeX, l'essentiel : pour une prise en main rapide et efficace*. Paris : Pearson education c2010. (Réserve BLSH Z 253.4 L38 B59 2010 DVDROM)
- GLUSHKO, Robert J.; McGRATH, Tim. *Document engineering: analyzing and designing documents for business informatics and Web services*. The MIT Press, 2005, 703 p. (Réserve BLSH QA 76.9 T48 G58 2005)

- KAY, Michael. *XSLT Programmer's Reference, Second Edition*. Wrox Press, 2001, 939 p. (Réserve BLSH QA 76.73 X58 K39 2001)
- KAY, Michael. *XSLT 2.0 Programmer's Reference, Third Edition*. Wiley, 2004.
- KAY, Michael. *XSLT 2.0 and XPath 2.0 Programmer's Reference, 4th Edition*. Wiley, 2008. [En ligne via Sofia – Université de Montréal](#),

Note : Plusieurs autres ressources sur XSL/XSLT sont répertoriées dans [Sofia – Université de Montréal](#).

- MARCOUX, Yves. « Les jeux de caractères ». En ligne <<https://marcoux.ebsi.umontreal.ca/ensejgn/6052/JeuxCar/JeuxCar.html>>.
- MILLER, Dick R.; CLARKE, Kevin S. *Putting XML to Work in the Library: Tools for Improving Access and Management*. Chicago, ALA, 2004. (Réserve BLSH Z 678.93 X54 M55 2004)
- NELLHAUS, Tobin. "XML, TEI, and Digital Libraries in the Humanities." *Portal: Libraries and the Academy* 1.3 (2001): 257-277. En ligne (accès VPN ou Proxy) <<https://muse.jhu.edu/article/27158>>.
- ROMARY, Laurent. "Questions & Answers for TEI Newcomers." *Jahrbuch für Computerphilologie* 10. Mentis Verlag, 2009. En ligne <<https://arxiv.org/abs/0812.3563>>.
- Text Encoding Initiative Consortium. *A Gentle Introduction to XML*. En ligne <<http://xml.coverpages.org/TEI-GentleIntroXML.pdf>>.
- TENNANT, Roy. *XML in Libraries*. Neal-Schuman, 2002. En réserve BLSH [Z 678.93 X54 X65 2002](#).
- TENNISON, Jeni. *Beginning XSLT*. Apress, 2004. En réserve BLSH [QA 76.73 X58 T47 2004](#).
- VAN DONGEN, M.R.C. *LaTeX and friends*. Heidelberg ; New York : Springer c2012, accessible en ligne via Atrium.
- W3C. *XML en 10 points*. En ligne <<https://www.w3.org/XML/1999/XML-in-10-points.fr.html>>.

7.3 Lectures utiles

- CAGLE, Kurt. "Why XSLT and XQuery Are Coming Back." [En ligne sur LinkedIn](#), 2020-06-19.
- DVORAK, John C. "Killing the Web," *PC Magazine*, octobre 2000. En ligne <<https://www.pcmag.com/archive/killing-the-web-4353>>.
- KORPELA, Jukka. *A tutorial on character code issues*. En ligne <<https://jkorpela.fi/chars.html>>. Mises à jour régulières.
- MALER, Eve; EL ANDALOUSSI, Jeanne. *Developing SGML DTDs: From Text to Model to Markup*. Prentice Hall PTR, 1996. (Réserve BLSH QA 76.76 H94 M26 1996)
- NELSON, Theodor Holm. "Embedded Markup Considered Harmful." En ligne: <<https://www.xml.com/pub/a/w3j/s3.nelson.html>>.
- PILLOU, Jean-François. *Introduction à XML*. En ligne <<https://www.commentcamarche.net/contents/1332-xml-introduction-a-xml>>.
- ROUMIEUX, Olivier. « XML sans frontières ». *Archimag*, no 159, novembre 2002.
- RYMER, John R. "Why 90 percent of XML standards will fail." En ligne <<https://www.zdnet.com/article/why-90-percent-of-xml-standards-will-fail-5000114625/>>.
- SAMPSON, Geoffrey. *e.biz: The Anatomy of Electronic Business*. Elsevier, 2005. (Réserve BLSH HF 5548.32 S26 2004)
- W3Schools. "XML Tutorial." En ligne <<https://www.w3schools.com/xml/>>.
- W3Schools. "XPath Tutorial." En ligne <https://www.w3schools.com/xml/xpath_intro.asp>.
- W3Schools. "XSLT Tutorial." En ligne <https://www.w3schools.com/xml/xsl_intro.asp>.

7.4 Diverses sources et ressources utiles

- ANDRIES, Patrick; et al. Une traduction en français de la norme XML 1.0. En ligne <<http://xml.coverpages.org/xml10-french.html>>.
- BRAY, Tim. La norme XML annotée par un des éditeurs, Tim Bray. En ligne <<https://www.xml.com/axml/axml.html>>.
- COVER, Robin. "The Cover Pages: The Online Resource for Markup Language Technologies." En ligne <<http://xml.coverpages.org/>>.
- GRDS. « Groupe départemental de Recherche sur les Documents Structurés ». En ligne <<https://grds.ebsi.umontreal.ca/>>.
- GRDS. [XML en route au Gouvernement du Québec](#) (rapport du <GRDS> sur l'utilisation de XML au Gouvernement du Québec; format PDF)
- GRDS. [Cadre de référence gouvernemental en gestion intégrée des documents](#) (rapport du <GRDS> sur la gestion intégrée des documents au Gouvernement du Québec; format PDF)
- ISO. *Freely Available Standards*. En ligne <<https://standards.iso.org/ittf/PubliclyAvailableStandards/>>.

- KARP, Alan H. “Viewpoint: Making money selling content that others are giving away.” *Communications of the ACM*, jan. 2003. En ligne <<https://dl.acm.org/doi/10.1145/602421.602438>>.
- MARCOUX, Yves. « Les formats normalisés de documents électroniques ». *ICO Québec*, vol. 6, nos 1-2, printemps 1994, pp. 56-65. En ligne <<https://marcoux.ebsi.umontreal.ca/publi/ico1994/ico94.htm>>.
- MARCOUX, Yves. « Place de SGML parmi les nouvelles architectures documentaires ». *Actes de la conférence « Technologie SGML 1996 » organisée par le Centre de recherche en droit public de l'Université de Montréal*, CRDP, 1996, pp. 1-13. En ligne <<https://marcoux.ebsi.umontreal.ca/publi/ottawa/marcoux.html>>.
- MARCOUX, Yves; SÉVIGNY, Martin. “Why SGML? Why now?” *Journal of the American Society for Information Science*, vol. 48, no 7, July 1997, Special Topic Issue: Structured Information/Standards for Document Architectures, pp. 584-592.
- MICHARD, Alain. *XML: Langage et applications*. Nouvelle édition, Eyrolles, 2001, 499 p. (Réserve BLSH QA 76.76 H94 M53 2001)
- OASIS. “Advancing open standards for the information society.” En ligne <<https://www.oasis-open.org/>>.
- [Registre- référentiel XML développé au <GRDS> pour le Gouvernement du Québec.](#)
- [Répertoire d'outils XML gratuits de Lars Marius Garshol \(Norvège\).](#)
- SOLARI, J.J. Une traduction en français de la norme XML 1.1. En ligne <<http://www.yoyodesign.org/doc/w3c/xml11/>>.
- Unicode Consortium. “Unicode Home Page.” En ligne <<https://unicode.org/>>.
- W3C. Spécification de XML 1.0 (5^e édition, datée du 26 novembre 2008). En ligne <<https://www.w3.org/TR/REC-xml>>.
- W3C. Spécification de XML 1.1 (2^e édition, datée du 29 septembre 2006). En ligne <<https://www.w3.org/TR/xml11>>.
- W3C. Spécification des schémas XML du W3C (2^e édition, datée du 28 octobre 2004). En ligne <<https://www.w3.org/TR/xmlschema-0/>>.
- W3C. “W3C Data Activity: Building the Web of Data.” En ligne <<https://www.w3.org/2013/data/>>.
- W3Schools. “W3Schools Online Tutorials.” Nombreux tutoriels gratuits sur plusieurs normes, du W3C et autres. En ligne <<https://www.w3schools.com/>>.
- XML.org . “XML.org advances the use of open standards by providing educational information, discussion areas, and collaborative resources.” En ligne <<http://www.xml.org/>>.
- Yergeau, F.; et al. *Terminologie bilingue XML*. En ligne <<https://yergeau.com/w3c/xml10/termino.html>>.

