

Modélisation et interopérabilité :

Semaine 41, cours 4

Partiel en salle : mardi 10 novembre
12h – 15h

Benoît Valiron <benoit.valiron@monoidal.net>

<http://inf356.monoidal.net/>

1

2

Projets

- Vous avez tous dû recevoir une copie
- Sinon, c'est sur le site du cours
- Instructions :
 - Tout seul ou en binôme
 - À rendre jeudi 3 décembre
 - Deux questions sur Xpath et XSLT seront rajoutés à la fin du mois
 - Mardi prochain je vous demanderai les noms des binômes (ou des uni-ôme), ainsi que la ligne de train choisie.
 - Indépendant des TPs, mais nous sommes là pour répondre à vos questions !
 - COMMENCEZ, ET POSEZ VOS QUESTIONS SI VOUS EN AVEZ !

3

Validation avec les DTDs

Appel général :	Appel général :
<ul style="list-style-type: none">• Pour les fils• Caractère des fils (obligatoire...) dans le format du père	<ul style="list-style-type: none">• Pour l'attribut lui-même• Caractère explicite
<!ELEMENT elt format>	<!ATTLIST elt att format caractère>
Fils de type texte :	Attribut de type texte :
<!ELEMENT elt (#PCDATA)>	<!ATTRIBUT CDATA ...>
CDATA = "character data" ← ici, < et > n'ont pas de sens spécial...	
PCDATA = "parsed character data" ← ici, < et > on un sens spécial	
Section CDATA en XML : <![CDATA[...]]> ← pas de sens spécial pour < et >	

4

Aujourd'hui, fin du tour du propriétaire.

Les formats XML...

- Peuvent définir une arborescence avec une racine unique :
 - XHTML → Racine <html> ... </html>
 - SVG → Racine <svg> ... </svg>
- Peuvent juste définir un ensemble de mots dissociés :
 - Dublin Core, définissant des balises <creator>, <date>, ...
- Mais peuvent aussi juste définir des attributs.

5

6

XLink

- Espace de nom : <http://www.w3.org/1999/xlink>
- But : attacher des liens aux documents XML.
- Méthode :
 - Attacher les attributs href et type du langage Xlink à un élément.
 - href → Lien vers une ressource
 - type → "simple" (mais il y en a d'autres)
 - Il faut définir un préfixe puisque **les attributs sans préfixe sont toujours sans espaces de noms.**

```
<roman xmlns:xlink="http://www.w3.org/1999/xlink"
      xlink:type="simple"
      xlink:href="http://archive.org/pub/etext/etext93/wizoz10.txt">
  <titre>The Wonderful Wizard of Oz</titre>
  <auteur>L. Frank Baum</auteur>
  <année>1900</année>
</roman>
```

7

Le préfixe spécial xml

- Un préfixe spécial ne nécessite pas de définition, le préfixe xml
- Deux attributs spéciaux :
 - xml:lang → Langue (humaine) utilisée
 - xml:space → Les espaces sont-ils significatifs ?
- Se doivent d'être définis dans la DTD, le cas échéant.

8

Formats de documents texte

- XHTML est spécialisé pour l'affichage pour le web.
- D'autres formats existent :
 - Polyvalent : Docbook
 - Orientés impression papier :
 - XSL-FO
 - Suites bureautiques :
 - ODF : OpenOffice
 - OOXML : Microsoft Office

9

Docbook

- Origine : format pour de la documentation technique.
- Sa définition comme standard par OASIS (*Organization for the Advancement of Structured Information Standard*) a généralisé son utilisation.
- Documentation du format sur le site <http://www.docbook.org>

10

Document Docbook

```
<!DOCTYPE xxx PUBLIC "-//OASIS//DTD DocBook XML V4.5//EN"
"http://www.oasis-open.org/docbook/xml/4.5/docbookx.dtd">
```

- Livres, avec la racine <book> ... </book>
 - Découpé en chapitre
 - Page de titre
 - Préface
- Articles, avec la racine <article> .. </article>
 - En un seul tenant, pas de sauts de page
 - Pas de chapitres
 - Exemple : le polycopié du TD 2 écrit en format article de DocBook

11

Exemple

```
<?xml version="1.0" encoding="utf-8"?>
<article>
  <articleinfo>
    <title>TD-2 : Espaces de noms</title>
    <author>
      <firstname>Benoit</firstname>
      <surname>Valiron</surname>
    </author>
  </articleinfo>
  <sect1>
    <title>Déclaration d'un espace de noms</title>
    <sect2>
      <title>Un titre de sous-section</title>
      <para>
        Texte de <emphasis>l'exercice 1</emphasis>.
      </para>
    </sect2>
  </sect1>
  <sect1>
    <title>Appartenance à un espace de noms</title>
    <para>
      <orderedlist>
        <listitem><para>Question 2.1</para></listitem>
        <listitem><para>Question 2.2</para></listitem>
        <listitem><para>Question 2.3</para></listitem>
      </orderedlist>
    </para>
  </sect1>
</article>
```

12

Principaux éléments pour un article

- Cinq niveaux hiérarchique <sect1> à <sect5>
- Paragraphes encadrés par <para> ... </para>
- Listes
 - Ordonnées : orderedlist
 - Non-ordonnées : itemizedlist
 - Items des listes : listitem
- Citations : blockquote
- Morceau de code (DTD, XML, programme...) : programlisting
- Dans le texte :
 - Emphase : emphasis
 - Liens hypertextes : ulink
 - Code : code

13

Exemple : programlisting

```
<?xml version="1.0" encoding="utf-8"?>
<article>
  <articleinfo>
    <title>TD-2 : Espaces de noms</title>
  </articleinfo>
  <sect1>
    <title>Une DTD pour RSS</title>
    <para>
      Un flot de données peut être encodé en format RSS.
      Une DTD simplifiée pour ce format est la suivante.
      <programlisting><![CDATA[<!ELEMENT rss (channel+)>
<!ELEMENT channel (title, description, item*)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT description (#PCDATA)>
<!ELEMENT item (title|pubDate|description)*>
<!ELEMENT pubDate (#PCDATA)>]]></programlisting>
      Chaque <code>channel</code> correspond à une
      liste de données, avec un titre, une
      description et les données, une
      par <code>item</code>.
    </para>
  </sect1>
</article>
```

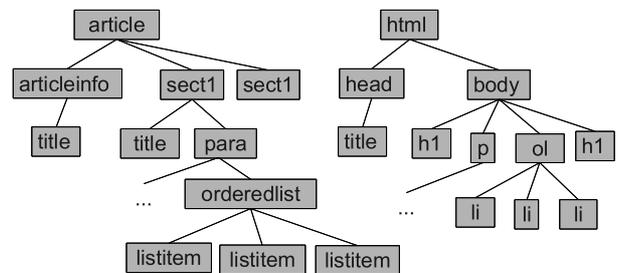
14

Comparaison avec XHTML

<pre><?xml version="1.0" encoding="utf-8"?> <article> <articleinfo> <title>TD-2 : Espaces de noms</title> <author> <firstname>Benoit</firstname> <surname>Valiron</surname> </author> </articleinfo> <sect1> <title>Un titre</title> <para> Du texte <emphasis>important</emphasis> </para> <orderedlist> <listitem> <para>Q 2.1</para> </listitem> <listitem> <para>Q 2.2</para> </listitem> <listitem> <para>Q 2.3</para> </listitem> </orderedlist> </sect1> <sect1> Autre section... </sect1> </article></pre>	<pre><?xml version="1.0" encoding="utf-8"?> <html> <header> <title>TD-2 : Espaces de noms</title> </header> <body> <h1>Déclaration d'un espace de noms</h1> <p> Du texte important. </p> Question 2.1 Question 2.2 Question 2.3 <h1>Autre titre</h1> </body> </html></pre>
--	---

15

Comparaison avec XHTML



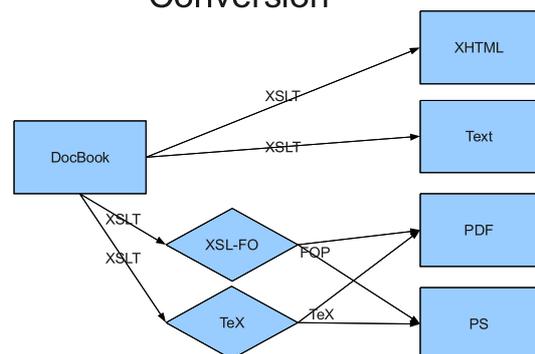
16

Extensions

- Docbook vient avec de nombreuses possibilités:
 - Table des matières
 - Index
 - Format pour de la bibliographie
 - ...
- Le format est modulaire : nul besoin de tout connaître pour produire un document.

17

Conversion



18

Résumé

- Docbook est orienté utilisateur humain
- Il est possible sans difficulté de concevoir un document docbook en utilisant un éditeur de texte de type Notepad+, emacs ou gedit (le TD 2 a été créé comme ça).
- Docbook est modulaire : de nombreuses fonctionnalités peuvent être ajoutées mais n'entrave pas l'écriture de document simple.
- Un fichier docbook peut être transformé en de nombreux formats.

19

XSL-FO

- Recommandation du W3C depuis 2001.
- Format XML pour décrire un document imprimé : Taille du papier, marges, ...
- Espace de noms : <http://www.w3.org/1999/XSL/Format>

20

XSL-FO : Format

```
<root>
<layout-master-set>
  <simple-page-master master-name="couverture" ...>
    <!-- page unique : description de la couverture -->
  </simple-page-master>
  <page-master-sequence master-name="contenu" ...>
    <!-- description physique des pages internes -->
  </page-master-sequence>
</layout-master-set>
<page-sequence master-reference="couverture">
  <!-- Texte de la couverture -->
</page-sequence>
<page-sequence master-reference="contenu">
  <!-- Texte du livre -->
</page-sequence>
</root>
```

21

XSL-FO : Mise en forme

- Simple-page-master

```
<simple-page-master master-name="contenu"
  page-height="11cm" page-width="8cm"
  margin-top="1cm" margin-bottom="1cm"
  margin-left="1cm" margin-right="1cm">
  <region-body region-name="contenu-region" />
</simple-page-master>
```

- Page-sequence

```
<page-sequence master-reference="contenu">
  <flow flow-name="contenu-region">
    <block font-family="Helvetica" font-size="20pt"
      font-weight="bold" text-align="center">
      Un titre
    </block>
    <block font-family="NewRoman" font-size="12pt"
      font-weight="normal" text-align="left">
      Du texte
    </block>
  </flow>
</page-sequence>
```

22

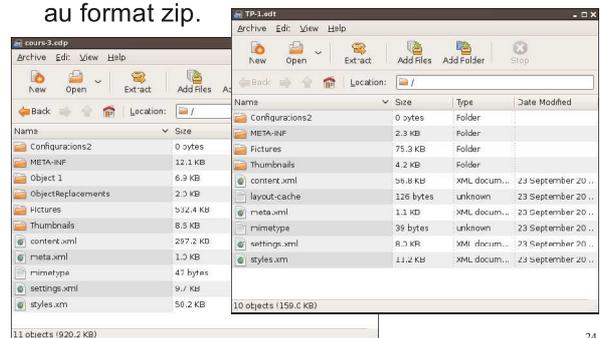
Commentaire

- XSL-FO est particulièrement verbeux. Ceci dit, la structure du document est suffisamment simple pour pouvoir écrire un canevas et s'en servir pour formater du texte.
- Il est spécialement utile dans le cas de transformations automatiques, par exemple en partant de DocBook.
- Son usage principal est la production de PDF au niveau des serveurs de contenu.

23

ODF et OpenOffice

- Un fichier OpenOffice est un fichier compressé au format zip.



24

Dedans...

- Du XML !
- En particulier :
 - content.xml → le contenu du document
 - styles.xml → la description des styles utilisés
 - meta.xml → méta-données : titre, auteur, ...
 - settings.xml → les réglages de la dernière sauvegarde (niveau de zoom, position du curseur...)

25

content.xml : Structure

```
<office:document-content
xmlns:office="urn:oasis:names:tc:opendocument:xmlns:office:1.0"
xmlns:style="urn:oasis:names:tc:opendocument:xmlns:style:1.0"
xmlns:text="urn:oasis:names:tc:opendocument:xmlns:text:1.0"
xmlns:table="urn:oasis:names:tc:opendocument:xmlns:table:1.0"
xmlns:draw="urn:oasis:names:tc:opendocument:xmlns:draw:1.0"
xmlns:fo="urn:oasis:names:tc:opendocument:xmlns:xsl-fo-compatible:1.0"
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:dc="http://purl.org/dc/elements/1.1/"
...>
<office:scripts>
<!-- Scripts définis -->
</office:scripts>
<office:font-face-decls>
<style:font-face ...>
<!-- Polices utilisées -->
</office:font-face-decls>
<office:automatic-styles>
<style:style ...>
<!-- Styles internes -->
</office:automatic-styles>
<office:body>
<!-- Corps du document -->
</office:body>
```

27

content.xml : exemple

```
<office:font-face-decls>
<style:font-face style-name="OpenSymbol"
svt:font-family="OpenSymbol" />
<style:font-face style-name="StarSymbol"
svt:font-family="StarSymbol" />
<style:font-face style-name="Liberation Mono"
svt:font-family="Liberation Mono"
style:font-family-generic="modern" style:font-pitch="fixed" />
<style:font-face style-name="Liberation Serif"
svt:font-family="Liberation Serif"
style:font-family-generic="roman"
style:font-pitch="variable" />
<style:font-face style-name="Liberation Sans"
svt:font-family="Liberation Sans"
style:font-family-generic="swiss"
style:font-pitch="variable" />
<style:font-face style-name="Dejavu LGC Sans"
svt:font-family="Dejavu LGC Sans"
style:font-pitch="variable" />
</office:font-face-decls>
<office:automatic-styles>
<style:style style-name="P1" style="family="paragraph"
style:parent-style-name="Standard"
<style:paragraph-properties text-align="center"
style:justify-single-words="false" />
</style:style>
```

29

content.xml

- Contenu différent en fonction du type de document : texte, tableur, présentation...
- Élément racine toujours le même : office:document-content
- Espaces de noms:
 - urn:oasis:names:tc:opendocument:xmlns:text:1.0
 - urn:oasis:names:tc:opendocument:xmlns:presentation:1.0

26

content.xml : exemple

```
<office:document-content
xmlns:office="urn:oasis:names:tc:opendocument:xmlns:office:1.0"
xmlns:style="urn:oasis:names:tc:opendocument:xmlns:style:1.0"
xmlns:text="urn:oasis:names:tc:opendocument:xmlns:text:1.0"
xmlns:table="urn:oasis:names:tc:opendocument:xmlns:table:1.0"
xmlns:draw="urn:oasis:names:tc:opendocument:xmlns:draw:1.0"
xmlns:fo="urn:oasis:names:tc:opendocument:xmlns:xsl-fo-compatible:1.0"
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:meta="urn:oasis:names:tc:opendocument:xmlns:meta:1.0"
xmlns:svg="urn:oasis:names:tc:opendocument:xmlns:svg-compatible:1.0"
xmlns:chart="urn:oasis:names:tc:opendocument:xmlns:chart:1.0"
xmlns:dr3d="urn:oasis:names:tc:opendocument:xmlns:dr3d:1.0"
xmlns:math="http://www.w3.org/1998/Math/MathML"
xmlns:form="urn:oasis:names:tc:opendocument:xmlns:form:1.0"
xmlns:script="urn:oasis:names:tc:opendocument:xmlns:script:1.0"
xmlns:core="http://openoffice.org/2004/core"
xmlns:oooc="http://openoffice.org/2004/oooc"
xmlns:dom="http://www.w3.org/2001/xml-events"
xmlns:foaf="http://www.w3.org/2000/foaf#"
xmlns:rss="http://www.w3.org/2001/XMLSchema"
xmlns:rst="http://www.w3.org/2001/XMLSchema-instance"
xmlns:rpt="http://openoffice.org/2009/report"
xmlns:of="urn:oasis:names:tc:opendocument:xmlns:of:1.2"
xmlns:rid="http://docs.oasis-open.org/.opendocument/meta/rid#fak"
xmlns:fld="urn:oasis:names:experimental:ooxm-interop:xmlns:field:1.0"
xmlns:orma="urn:oasis:names:experimental:ooxm-odf-interop:xmlns:form:1.0"
office:version="1.2">
```

28

content.xml : exemple

```
<style:font-family-generic="system"
style:font-pitch="variable" />
</office:font-face-decls>
<office:automatic-styles>
<style:style style-name="P1" style="family="paragraph"
style:parent-style-name="Standard"
<style:paragraph-properties text-align="center"
style:justify-single-words="false" />
</style:style>
<style:style style-name="P2" style="family="paragraph"
style:parent-style-name="Standard"
<style:paragraph-properties text-align="center"
style:justify-single-words="false" />
</style:style>
<style:style style-name="P3" style="family="paragraph"
style:parent-style-name="Standard"
<style:paragraph-properties text-align="center"
style:justify-single-words="false" />
</style:style>
<style:tab-stop style:position="0.1002in" />
<style:tab-stop style:position="0.3997in" />
<style:tab-stop style:position="0.7002in" />
<style:tab-stop style:position="0.9998in" />
</style:tab-stops>
<style:paragraph-properties
style:font-size="20pt"
<style:text-properties font-size="20pt">
```

30

OOXML : Microsoft Office

- Vrai nom : ECMA-376 : Office Open XML file format.
- Maintenant, après de nombreuses péripéties, c'est un standard ISO.
- Descendant du format binaire : .xls, .doc, .ppt
- Néanmoins, le corps du document est en XML.

37

Un fichier .docx

- Décompressé :
 - [Content_Types].xml → décrit le contenu de l'archive
 - _rels → dossier contenant des fichiers XML décrivant les relations entre les composantes de l'archive.
 - DocProps → dossier avec les propriétés
 - De l'application : app.xml
 - Du document : core.xml
 - Définies par l'utilisateur : custom.xml
 - Un répertoire pour le type de document (ex : word) contenant document.xml

38

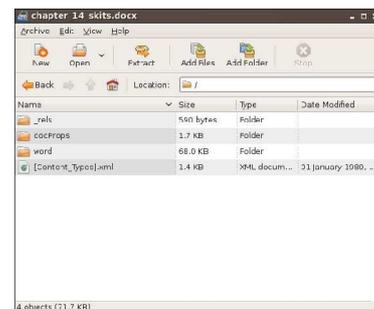
Exemple de document.xml

Un paragraphe contenant les mots "gras" et "italique" en gras et en italique :

```
<w:p>
  <w:r>
    <w:rPr><w:b/></w:rPr>
    <w:t>gras</w:t>
  </w:r>
  <w:r>
    <w:rPr><w:i/></w:rPr>
    <w:t>italique</w:t>
  </w:r>
</w:p>
```

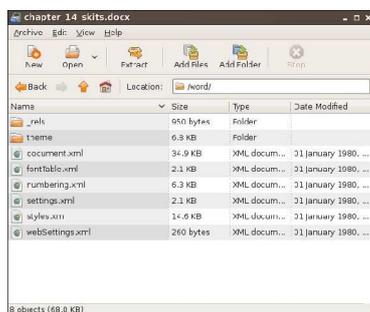
39

Un exemple tiré du ouaibe (je n'ai pas Office 2007...)



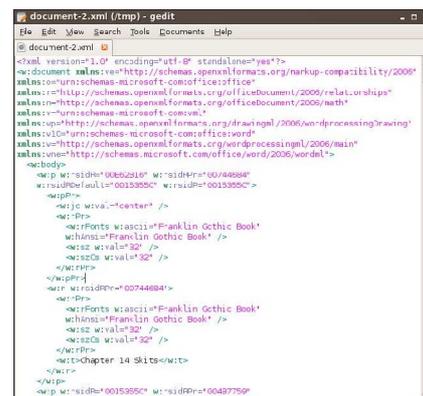
40

Un exemple tiré du ouaibe (je n'ai pas Office 2007...)



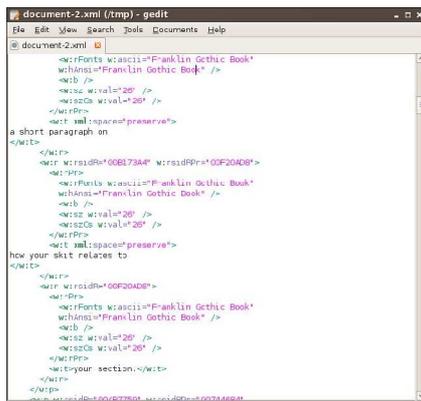
41

document.xml



42

document.xml



```
document-2.xml (/tmp) - gedit
File Edit View Search Tools Documents Help
document-2.xml
<w:font w:ascii="Franklin Gothic Book"
w:hings="Franklin Gothic Book" />
<w: />
<w:sz w:val="20" />
<w:color w:val="20" />
</w:font>
<w:xml:space="preserve">
a short paragraph on
</w: />
<w: />
<w:rsid="00B17344" w:rsidPr="00F20408">
<w:Pr>
<w:font w:ascii="Franklin Gothic Book"
w:hings="Franklin Gothic Book" />
<w: />
<w:sz w:val="20" />
<w:color w:val="20" />
</w:Pr>
<w:xml:space="preserve">
how your skit relates to
</w: />
<w: />
<w:rsid="00F20408">
<w:Pr>
<w:font w:ascii="Franklin Gothic Book"
w:hings="Franklin Gothic Book" />
<w: />
<w:sz w:val="20" />
<w:color w:val="20" />
</w:Pr>
<w:four section.</w: />
</w: />
</w:Pr>
</w: />
```

43

Commentaires

- Comme pour ODF, on a un format orienté machine.
- À la différence d'ODF, les styles sont inclus dans le contenu.
- Il est particulièrement peu lisible, avec des balises et des espaces de noms peu explicites, ce qui ne l'empêche pas d'être verbeux.
- Il porte le lourd héritage du format binaire

44

Conclusion

- Il y a mille façons d'utiliser le format XML pour encoder du texte :
 - Format humain
 - Format machine
 - Orienté ouaibe ou orienté impression papier

45