

DocBook-XML et éditeur XMLmind

Bernard Perrot, CNRS - UMR6205

Copyright © <bernard.perrot@univ-brest.fr>

Préambule

Cette présentation
est un crime
de lèse TeX/LaTeX,

désolé...

XML

XML

Qu'est-ce que XML ?

- **Sujet traité aux journées Mathrice au CIRM à l'automne dernier;**
- **D'accord, mais j'étais absent ...**
 - **... oui, mais on n'a pas le temps... mais c'est moins grave, on va parler d'un produit (éditeur XML) qui permet de (presque) se passer de connaissances (pointues) en XML.**

DocBook

DocBook

Historique

- Apparue en 1992 à la suite d'un projet commun de HaL Computer Systems et l'éditeur O'Reilly visant à faciliter l'échange de documentation UNIX.
- **« DocBook est une DTD destinée à la rédaction de livres et d'articles qui «définit un ensemble de balises pour des documents littéraires, et qui fonctionne comme le langage HTML utilisé habituellement sur le Web. [...] Lors de son traitement, un unique fichier SGML DocBook peut produire des fichiers HTML, PDF, PS, TXT ou d'autres formats de publication papier ou électronique. Le traitement est régi par des feuilles de style qui peuvent générer automatiquement une table des matières, la numérotation des pages, la numérotation des chapitres et des sections, et bien d'autres possibilités. »**

(DocBook Install mini-HOWTO)

Aujourd'hui

- **Aujourd'hui, c'est le DocBook Technical Committee de l'OASIS (Organization for the Advancement of Structured Information Standards) qui se charge de sa maintenance et de son évolution;**
- **Initialement développé en SGML, DocBook est désormais disponible en XML;**
 - **ne pas commencer un nouveau projet en SGML !**
- **Près de 400 éléments (balises);**
- **versions simplifiées :**
 - **Simplified DocBook;**
 - **Slides;**
 - **Linux Documentation Project;**
 - **...**

Pour quoi faire ?

- **Surtout destiné et utilisé pour produire de la documentation technique dans les domaines :**
 - **informatique;**
 - **télécommunications;**
 - **technologies de la communication;**
 - **...**

Avantages

- **Séparation du contenu du document et de la présentation;**
- **format libre et très largement documenté :**
 - **manuel « *The Definitive Guide* » en ligne et en téléchargement sur Internet**
- **documents pérennes;**
- **très adaptés aux travaux partagés, collaboratifs, évolutifs.**
 - ***[aparté] un outil pour relancer l'idée de documentation Mathrice collaborative ?***

Éléments de *DocBook*

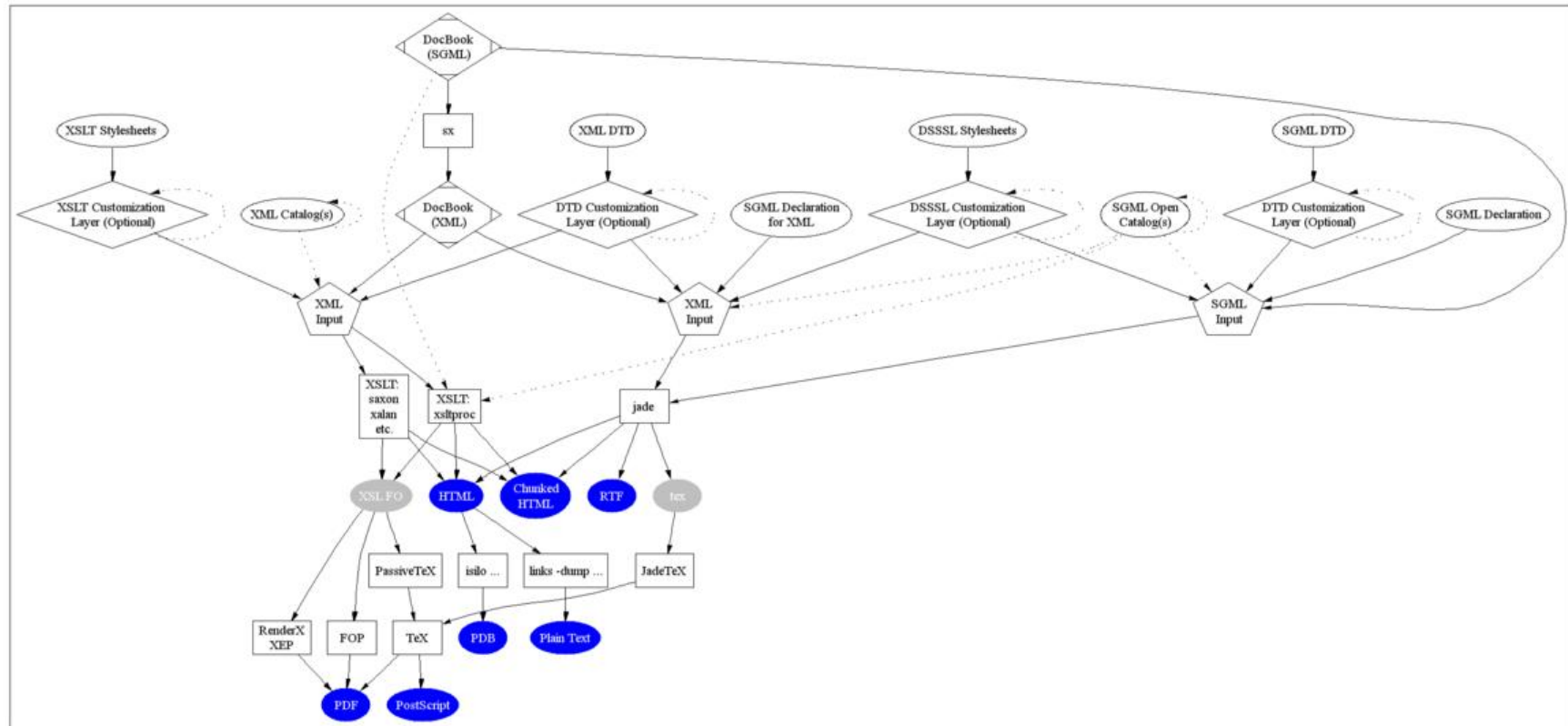
- Près de 400 éléments répartis en 2 catégories :
 - hiérarchie : éléments structurels
 - information : éléments qui contiennent les données elles-mêmes (le contenu)
- Les éléments de départ sont :
 - *Set* : élément de plus haut niveau (collection, recueil de livres, ...);
 - *Book* : élément de départ le plus commun;
 - *Article* : plus simple que le *Book* pour des publications courtes;
 - *Article* est également le seul élément de *Simplified DocBook*
 - *Slides* : pour les présentations (telle que celle-ci)

Comment faire ?

- **Deux étapes bien distinctes pour produire un document *DocBook* :**
 - **l'écriture du texte proprement dit (en XML), en respectant la DTD de *DocBook* :**
 - **avec vi, Emacs, Nedit, Notepad, votre éditeur de texte favori...**
 - **avec un éditeur orienté XML, tel que XMLmind qui va être introduit ci-après.**
 - **la transformation de ce texte balisé en document lisible :**
 - **PDF, PostScript, HTML, ...**
 - **ce qui donnent le schéma de publication suivant :**

schéma du modèle de publication

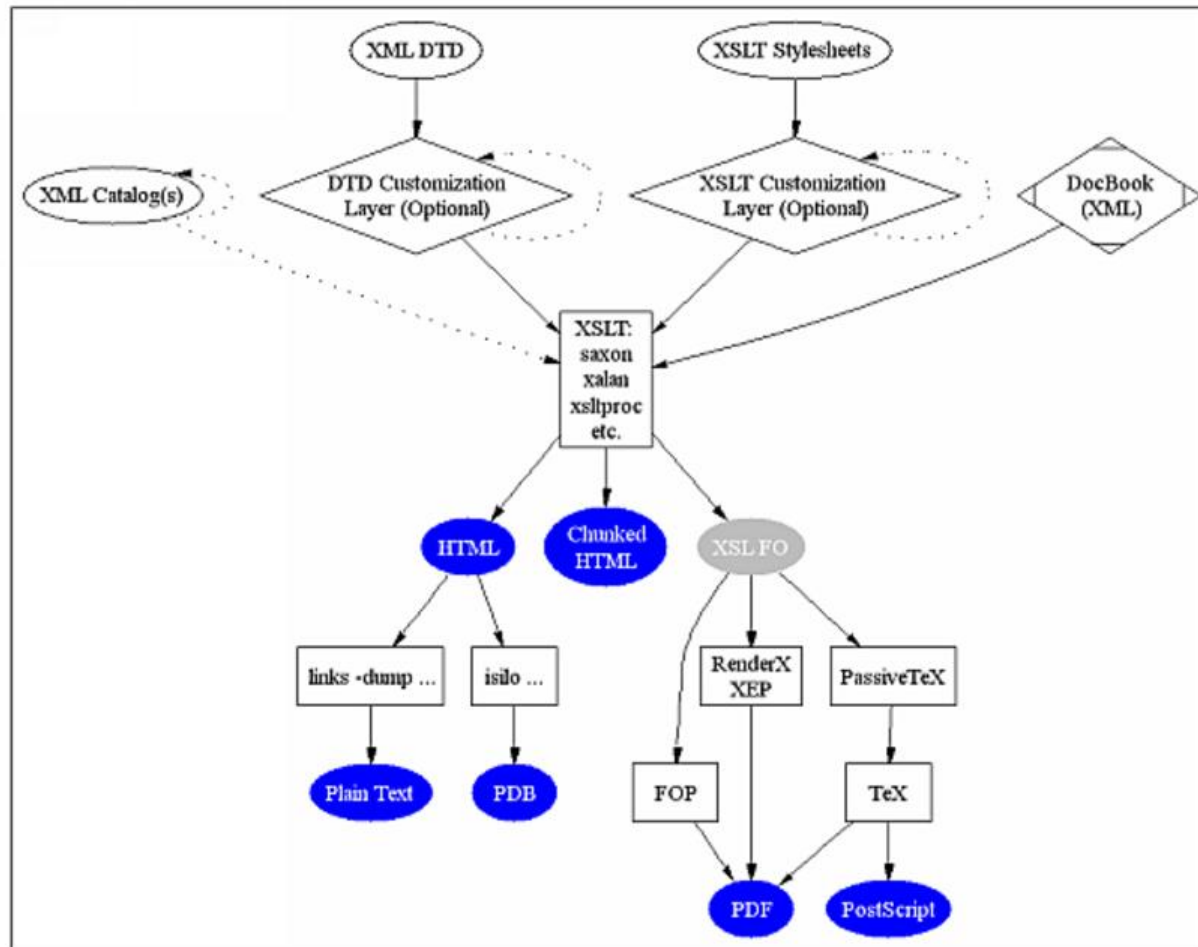
- schéma complet (SGML, XML)



Full DocBook SGML/XML Publishing Model

schéma du modèle de publication

- schéma XML/XSLT seul



DocBook XML Publishing (XSLT)

Difficultés

- **Au niveau de l'écriture, la principale difficulté sera le respect de la syntaxe XML et de la DTD de *DocBook*, c'est à dire faire que le document soit « bien formé ».**
 - **cette contrainte est forte avec l'usage d'un éditeur de texte habituel (générique);**
 - **extrêmement simplifiée avec un éditeur orienté XML comme XMLmind (car contraint absolument au respect de XML et de la DTD).**
- **Au niveau de la production du document final, les difficultés sont liées au côté parfois (souvent ?) obscur de l'usage des outils dédiés à cette tâche (Jade, xsltproc, ...);**

Difficultés (Suite)

- **cette difficulté est également très simplifiée par l'usage d'outils prenant en charge les transformations (simple clic sur la fonction de transformation désirée)**
- **Il reste assez technique de modifier les feuilles de styles (XSL, CSS) si on souhaite adapter/modifier/changer la présentation standard des outils utilisés.**

Structure d'un document DocBook-XML

Structure d'un document DocBook-XML

Les balises XML

- **Élément** `elem` contenant le texte « Abc de F » :

```
<element> Abc de F </element>
```

- **Élément vide (deux formes) :**

```
<element></element>  
<element/>
```

- **Élément avec attributs :**

```
<element attribut1="valeur1"  
attribut2="valeur2"> texte </element>
```

- **Commentaire :**

```
<!-- ceci est un commentaire -->
```

Les balises XML (Suite)

- **Instruction XML (déclaration de type, définition d'une entité, ...):**

```
<!instruction param1 param2...>
```

- **Instruction de traitement (destinée à l'application) :**

```
<?instruction param1="value1" param2="value2"?>
```

L'entête

- Si usage d'un éditeur spécialisé, elle sera normalement générée automatiquement.

Par exemple :

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE article PUBLIC
"-//OASIS//DTD DocBook XML V4.5//EN"
"http://www.oasis-open.org/docbook/xml/4.5/docbookx.dtd">
```

- première ligne : prologue, obligatoire, indique la version XML et l'encodage.
- deuxième ligne : déclaration de la DTD *DocBook*. Définie ici de manière publique.

Pour une déclaration interne (SYSTEM) on aurait pu avoir (pour une autre version de la DTD dans cet exemple) :

L'entête (Suite)

```
<!DOCTYPE article SYSTEM  
"/usr/share/sgml/docbook/xml-dtd-4.3-1.0-25/docbookx.dtd">
```

- **Ce prologue est suivi du contenu du document, qui doit commencer par l'élément racine précisé dans le prologue :**

```
<article>  
  <title>Le titre de l'article</title>  
  <section>  
    ....  
</article>
```

Hiérarchie des éléments

Par ordre hiérarchique, voici la liste simplifiée (non exhaustive) des éléments les plus importants :

1. `<set>` : ensemble de un ou plusieurs livres;
2. `<book>` : livre, type de document le plus complet;
3. `<part>` : partie d'un livre;
4. **composants** : `<chapter>` (chapitre d'un livre), `<article>` (utilisé habituellement comme élément racine) ou `<reference>` (utilisé pour des pages de référence, dans le genre manpage);
5. `<section>` : élément pouvant s'imbriquer plusieurs fois, définissant ainsi des sous-sections;
6. **blocs** : éléments apparaissant sous forme de blocs séparés les uns des autres, et prenant toute la largeur de la page: `<para>`, `<xxxxlist>`, `<table>`, `<mediaobject>`, ...

Hiérarchie des éléments (Suite)

7. *inline* : éléments incorporés dans un bloc de texte, comme `<emphasis>` utilisé au début de cette ligne.

On peut citer aussi des éléments pouvant faire partie d'un `<book>` tels que `<toc>`, `<preface>`, `<index>`, ..., non obligatoires.

De même, tous les éléments cités ne sont pas toujours obligatoires, un `<article>` peut avoir un `<para>` sans `<section>`.

Le manuel *DocBook : the Definitive Guide* (existe en français) est la référence.

Meta-informations

<code><title></code>	titre du document
<code><author></code>	nom de l'auteur
<code><honorific></code>	(M., Mme, Dr)
<code><firstname></code>	
<code><surname></code>	
<code><lineage></code>	(Junior, Senior...),
<code><othername></code>	
<code><affiliation></code>	l'organisation auquel appartient un individu
<code><jobtitle></code>	la profession au sein de l'organisation
<code><orgname></code>	le nom de l'organisation
<code><orgdiv></code>	le nom d'une division
<code><address></code>	l'adresse d'un individu
<code><street></code>	
<code><pub></code>	(boîte postale)
<code><postcode></code>	
<code><citystate></code>	

Meta-informations (Suite)

<code><country></code>	
<code><phone></code>	
<code><fax></code>	
<code><email></code>	
<code><otheraddr></code>	
<code><copyright></code>	copyright, avec année(s) et détenteur
<code><year></code>	
<code><holder></code>	
<code><date></code>	date de publication ou de révision
<code><releaseinfo></code>	informations sur la distribution actuelle
<code><revhistory></code>	historique des révisions
<code><revision></code>	entrée décrivant une révision
<code><revnumber></code>	numéro de révision
<code><revremark></code>	commentaire associé
<code><legalnotice></code>	informations légales
<code><keywordset></code>	ensemble de mots-clefs
	correspondants au document
<code><keyword></code>	entrée correspondant à un mot-clef

Éléments *inline* courants

<code><abbrev></code>	abréviation
<code><acronym></code>	acronyme (abréviation se prononçant comme un mot normal)
<code><citetitle></code>	titre d'une oeuvre
<code><emphasis></code>	italique
<code><foreignphrase></code>	expression étrangère
<code><personname></code>	nom d'une personne
<code><replaceable></code>	partie à remplacer par l'utilisateur (comme <code><sectn></code> à la place de <code><sect1></code> , <code><sect2></code> ...)
<code><quote></code>	citation
<code><subscript></code>	indice (exemple)
<code><superscript></code>	exposant (exemple)
<code><markup></code>	expression à représenter tel quel
<code><productname></code>	nom d'un produit
<code><trademark></code>	marque déposée

Liens hypertextes

<code><anchor></code>	définie une cible pour un lien
<code><xref></code>	lien interne avec texte généré par la cible (attribut <code>xrefname</code>)
<code><link></code>	lien interne
<code><ulink></code>	lien externe grâce à une URL
<code><olink></code>	lien généré par <code><modespec></code>
<code><modespec></code>	définit un lien à générer pour <code><olink></code>

Notes de bas de pages

<code><footnote></code>	note de bas de page
<code><footnoteref></code>	référence vers une note

Callout: Les *callouts* permettent d'associer une annotation à une marque dans un bloc particulier (image, listing...). Le principe est le même que pour une note de bas de page dans du texte ordinaire.

<code><area></code>	<code><areaset></code>
<code><areaspec></code>	<code><callout></code>
<code><calloutlist></code>	<code><co></code>
<code><coref></code>	<code><screenco></code>
<code><programlistingco></code>	<code><mediaobjectco></code>
<code><imageobjectco></code>	

Listes

<code><simplelist></code>	liste simple, peut être définie horizontale ou verticale grâce à l'attribut <code>type</code> ,
<code><member></code>	entrée, contient directement le texte,
<code><itemizedlist></code>	liste avec puce,
<code><listitem></code>	entrée, doit contenir un élément bloc comme <code><para,></code>
<code><orderedlist></code>	liste ordonnée,
<code><listitem></code>	entrée (cf. au dessus),
<code><variablelist></code>	liste dont chaque entrée associe un terme à une description,
<code><varlistentry></code>	entrée,
<code><term></code>	terme à décrire,
<code><listitem></code>	description (cf. au dessus),

Listes (Suite)

<code><segmentedlist></code>	liste comprenant plusieurs titres, et dont chaque entrée possède une sous-entrée pour chaque titre,
<code><segtitle></code>	titre, contient directement le texte,
<code><seglistitem></code>	entrée, (doit contenir autant de <code><seg></code> qu'il y a de <code><segtitle></code> en début de liste),
<code><seg></code>	sous-entrée, contient le texte directement.

Tableaux

<code><informaltable></code>	tableau sans titre
<code><table></code>	tableau avec titre
<code><tgroup></code>	contient le tableau et spécifie le nombre de colonnes dans l'attribut <code>cols</code>
<code><spansec></code>	associe un nom à un étalement sur plusieurs colonnes (optionnel)
<code><colspec></code>	formate une colonne (optionnel)
<code><thead></code>	première(s) rangée(s) du tableau, généralement l'en-tête (optionnel)
<code><tfoot></code>	dernière(s) rangée(s) du tableau (optionnel)
<code><tbody></code>	corps du tableau, contient les rangées sans formatage particulier
<code><row></code>	contient une rangée
<code><entry></code>	cellule du tableau (pouvant s'étaler sur plusieurs colonnes ou rangées)
<code><entrytbl></code>	cellule contenant un sous tableau

Images

<code><mediaobject></code>	insère un objet en mode bloc
<code><inlinemediaobject></code>	insère un objet en mode inline (permet de l'inclure dans un <code><ulink></code>)
<code><caption></code>	titre de l'objet
<code><imageobject></code>	image correspondant à l'objet
<code><imagedata></code>	fichier ou entité contenant l'image
<code><textobject></code>	texte correspondant à l'objet, contient <code><textdata></code> ou un autre élément bloc (comme <code><para></code>)
<code><textdata></code>	fichier ou entité contenant le texte

Termes et données informatiques

Éléments blocs :

<code><screen></code>	texte sur un écran, reproduit le contenu tel quel (existe aussi avec callouts)
<code><programlisting></code>	listing, reproduit le contenu tel quel (existe aussi avec callouts)
<code><synopsis></code>	décrit la syntaxe d'une commande ou fonction (beaucoup d'éléments associés : se référer à la doc!)
<code><lineannotation></code>	ligne de commentaire pour un des 3 éléments in extenso ci dessus
<code><screenshot></code>	copie d'écran, contient un <code><mediaobject></code>

Termes et données informatiques (Suite)

Éléments *inline* :

<code><computeroutput></code>	texte généré par l'ordinateur,
<code><userinput></code>	texte tapé par l'utilisateur,
<code><application></code>	nom d'un logiciel,
<code><command></code>	nom d'un exécutable,
<code><filename></code>	nom d'un fichier
<code><systemitem></code>	terme relatif à un système (comme le nom d'un OS ou d'une ressource),
<code><hardware></code>	composant hardware
<code><sgmltag></code>	balise SGML ou XML,
<code><constant></code>	nom d'une constante,
<code><errortext></code>	message d'erreur,
<code><envar></code>	variable d'environnement,
<code><variable></code>	variable
<code><prompt></code>	invite au début d'une ligne de commande,
<code><property></code>	propriété,

Termes et données informatiques (Suite)

<code><guibutton></code>	nom d'un bouton dans une GUI (interface graphique utilisateur),
<code><guimenu></code>	nom d'un menu dans une GUI,
<code><guimenuitem></code>	nom d'une entrée dans un menu,
<code><accel></code>	lettre raccourci clavier dans un menu (comme F dans Fichier,)
<code><shortcut></code>	raccourci clavier
<code><keycombo></code>	combinaison de touches
<code><keycap></code>	nom d'une touche

Autres éléments blocs :

<code><literallayout></code>	texte reproduit in extenso (tel quel/verbatim)
<code><lineannotation></code>	commentaire
<code><example></code>	exemple avec un titre
<code><informalexample></code>	exemple sans titre
<code><figure></code>	figure avec un titre pouvant être flottante (attribut <code><float></code> à 1)
<code><informalfigure></code>	figure sans titre pouvant être flottante
<code><blockquote></code>	citation
<code><attribution></code>	auteur d'une citation
<code><abstract></code>	résumé
<code><highlights></code>	point principaux
<code><caution></code> , <code><important></code>	admonitions
<code><warning></code> , <code><note></code> , <code><tip></code>	
<code><remark></code>	
<code><qandaset></code>	ensemble de questions/réponses
<code><qandadiv></code>	titre d'une division de l'ensemble

Autres éléments blocs : (Suite)

<code><qandaentry></code>	entrée comprenant une question et (optionnel) une ou plusieurs réponses
<code><question></code>	question (le "Q :" par défaut peut être changé par <code><label></code>)
<code><answer></code>	réponse (le "A :" par défaut peut être changé par <code><label></code>)
<code><label></code>	change le label d'une question ou réponse
<code><sidebar></code>	bloc séparé du reste du texte normal du document
<code><procedure></code>	description d'une procédure, d'un algorithme
<code><step></code>	étape dans la procédure
<code><substeps></code>	sous-étape

Et encore...

- écriture de bibliographies : `<bibliography>`
- glossaires : `<glossary>`
- tables des matières : `<toc>`
(souvent générées automatiquement)
- index : `<index>`
- pages de référence : `<reference>`
- modules EBNF, HTML Form, SVG et MathML.

Entités internes

- Il s'agit de mot-clés permettant par exemple de remplacer une expression souvent employée.
- Elles sont définies ainsi en début de document (à la fin du tag `DOCTYPE`, entre crochets) :
 - `<!ENTITY entité "expression">`
- Puis elles sont instanciées par `&entité;`

```
<?xml version="1.0" encoding="..."?>
<!DOCTYPE book PUBLIC "... " "..." [
<!ENTITY bz "bizarre">
]>
<book>
...
<para>
- Oui, vous regardez votre couteau et vous
  dîtes &bz;, &bz;. Alors je croyais que ...
</para><para>
```

Entités internes (Suite)

```
- Moi, j'ai dit &szlig;, &szlig;, comme c'est étrange !  
  Pourquoi aurais je dit &szlig;, &szlig; ?  
</para><para>  
- Je vous assure mon cher cousin,  
  que vous avez dit &szlig;, &szlig;.  
</para><para>  
- Moi, j'ai dit &szlig;,  comme c'est &szlig; !"   
</para>  
...
```

Entités externes

Elles permettent d'inclure un fichier externe dans le document.

```
<?xml version="1.0" encoding="..."?>
<!DOCTYPE book PUBLIC "... "..." [
<!ENTITY chap1 SYSTEM "/home/bernard/doc/pommes.xml">
<!ENTITY chap2 SYSTEM "/home/bernard/doc/poires.xml">
<!ENTITY chap3 SYSTEM "/home/bernard/doc/tatin.xml">
]>
<book>
  <title>Les tartes aux fruits</title>
  &chap1;
  &chap2;
  &chap2;
</book>
```


Alternative : XInclude

Permet la référence externe et l'inclusion de tout ou partie (c'est le « partie » qui est intéressant) d'un autre document XML.

Pour cela, dans le document référencé, l'attribut « id » de l'élément référencé doit être renseigné :

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE article PUBLIC
  "-//OASIS//DTD DocBook XML V4.5//EN"
  "http://www.oasis-open.org/docbook/xml/4.5/docbookx.dtd">
<article>
  <title>Le titre de l'article</title>
  <section id="section-beautemps"><title>Beau temps</title
  ...
</section>
...
```

Alternative : XInclude (Suite)

Dans le document appelant :

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE article PUBLIC
  "-//OASIS//DTD DocBook XML V4.5//EN"
  "http://www.oasis-open.org/docbook/xml/4.5/docbookx.dtd">

<article>
  <title>Le titre d'un article</title>

  <section><title>La pluie</title>...</section>

  <xi:include href="./beau.xml"
    xmlns:xi="http://www.w3.org/2001/XInclude"
    xpointer="element(section-beau-temps)" />
  ...
```

Caractères spéciaux

Il s'agit en fait aussi d'entités (entités SGML, comme en HTML, air connu), instanciées également par `&entité;` :

<code>half</code>	½	<code>num</code>	#
<code>copy</code>	©	<code>laquo</code>	«
<code>lt</code>	<	<code>amp</code>	&
<code>reg</code>	®	<code>raquo</code>	»
<code>gt</code>	>	<code>ast</code>	*
<code>trade</code>	™	<code>para</code>	¶
<code>dollar</code>	\$	<code>commat</code>	@
<code>micro</code>	μ	<code>hellip</code>	...

Transformations

- **Généralement, on va vouloir transformer le source DocBook-XML en document Web (HTML, XHTML), ou imprimable (PS, PDF).**
- **Pour produire un document (X)HTML, on pourra utiliser `xsltproc`.**
- **Pour produire un document PS ou PDF, on pourra utiliser `Openjade`, ou un outils tirés des *docbook-utils* (qui sont en fait des wrappers au dessus de `Openjade`). `Openjade` est une application Java, qui transite par un fichier *TeX*.**
- **Plus conforme au schéma de production (XSL-FO) pour les sorties imprimables, on trouvera des produits comme Apache-FOP (open-source, gratuit), ou RenderX-XEP (commercial, gratuit ou payant).**

Transformations (Suite)

- **L'usage de ces outils en ligne de commande est souvent assez technique (nombreux arguments parfois ésotériques) et rébarbative, une solution intégrée éditeur-transformateur avec interface graphique est présentée ci-après.**

XMLmind

XMLmind

Qu'est-ce que XXE ?

- ***XXE (XMLmind XML Editor)* est un éditeur XML.**
- **Ce n'est pas un outil orienté développeur, mais plutôt destiné aux rédacteurs de documentation techniques.**
- **Deux versions (actuelle = v3.5.2):**
 - **une gratuite, dite aussi « standard ».**
 - **une payante, dite aussi « professionnelle » (licence individuelle = 220 €TTC)**
 - **la version standard a quelques restrictions, la plus importante pour l'usage en éditeur de documentation est l'impossibilité de convertir (en HTML, PDF, ...) directement à partir du produit, il faudra passer par la ligne de commande externe (la version 3.5 standard était moins restrictive)**

Qu'est-ce que XXE ? (Suite)

- **multi-plateforme : Unix, Windows, MacOS. En fait c'est une application 100% Java, il suffit d'avoir une JVM fonctionnelle (les distributions incluent ou non une JVM pour les environnements Windows ou MacOS).**
- **Conversions (addons) en : HTML, PDF, PS, RTF, Word2003, OpenOffice2**
- **<http://www.xmlmind.com/xmleditor/>**

Du même éditeur

- ***XMLmind FO Converter* : conversion XML vers traitement de texte usuels : Word2003, RTF, OpenOffice2 :**
 - **version standard gratuite entièrement fonctionnelle (ligne de commande)**
 - **version professionnelle payante avec API (inutile sauf à des développeurs/intégrateurs)**
 - ***Aptconvert* : conversion depuis le format APT (Almost Plain Text) vers :**
 - **HTML, XHTML, PDF, PostScript, (MS Word loadable) RTF, DocBook SGML and DocBook XML.**
 - **règles de balisage hyper-simplifiée (indentation, quelques caractères spéciaux).**
- Exemple de texte en APT :**

Du même éditeur (Suite)

```
This is the title of a section
```

```
This is a paragraph which contains several sentences.  
Sentence one. Sentence two. Sentence three.  
This paragraph is followed by a bulleted list.
```

```
* List item 1.
```

```
* List item 2. This <word> will be rendered  
in italic. This other <<word>> will be  
rendered in bold.
```

```
* List item 3.
```

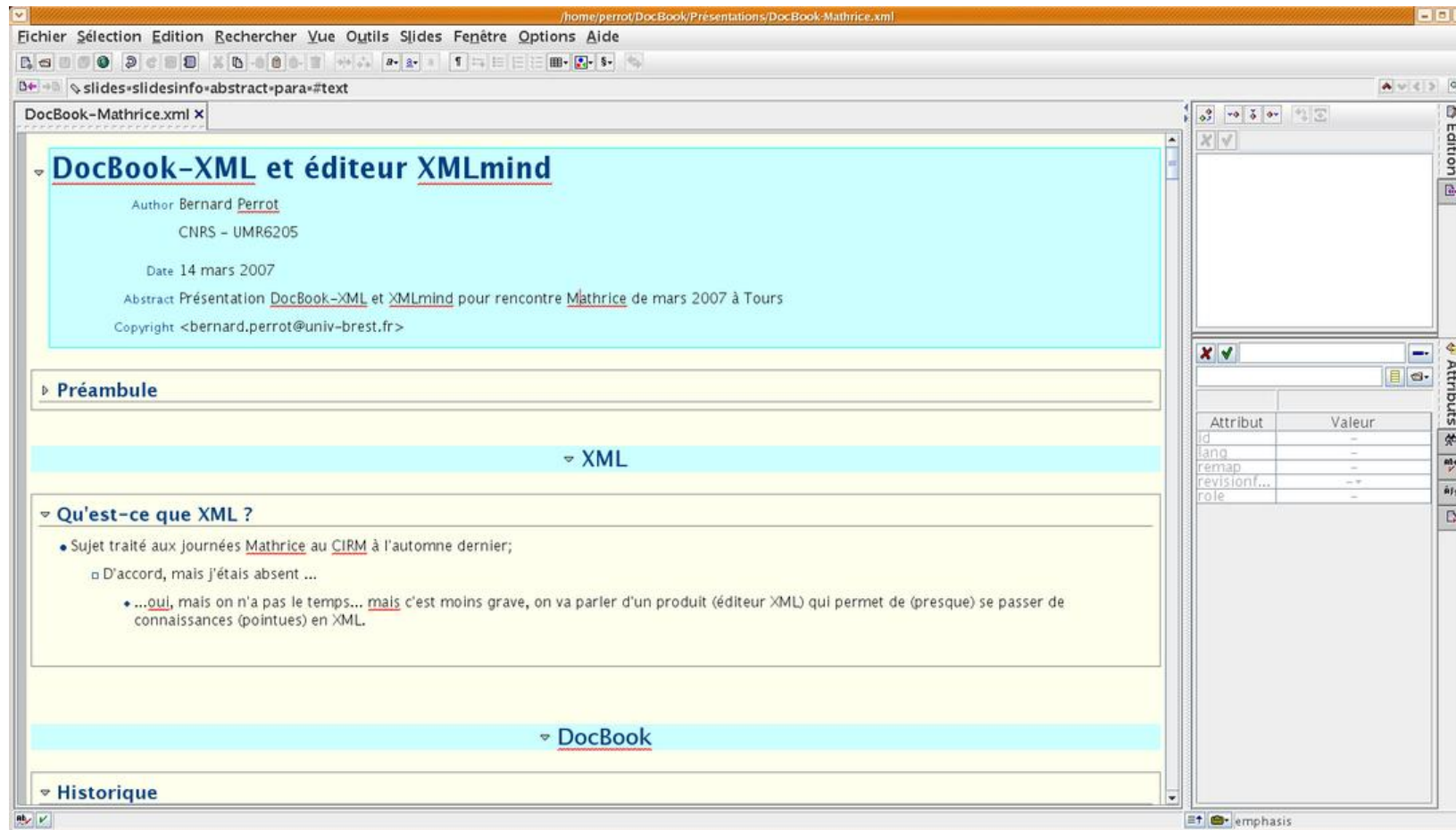
- **c'est gratuit et permet de récupérer en très peu de temps, et convertir (par exemple) en DocBook des anciennes documentations écrites en ASCII au kilomètre.**

RenderX-XEP

- **C'est un convertisseur XSL-FO, donc permettant de produire des documents PDF ou PostScript à partir de sources XML.**
- **Plus performant de Apache-FOP (mais la version 0.93 de Apache-FOP commence à être plus utilisable, tableaux par exemple).**
- **Versions gratuite et payante (licence desktop = 299\$) : la version gratuite (personnelle) est entièrement fonctionnelle, mais ajoute un petit texte en très bas de page indiquant que ce document est produit avec RenderX-XEP;**
- **s'intègre (gratuit ou payant) à XMLmind pour la production de PS et PDF.**
- **<http://www.renderx.com/tools/xep.html>**

Une démonstration...

... valant mieux qu'un long discours...



The screenshot shows the XMLmind editor interface. The main window displays a presentation slide with the following content:

- DocBook-XML et éditeur XMLmind**
- Author [Bernard Perrot](#)
- CNRS - UMR6205
- Date 14 mars 2007
- Abstract Présentation [DocBook-XML](#) et [XMLmind](#) pour rencontre [Mathrice](#) de mars 2007 à Tours
- Copyright <bernard.perrot@univ-brest.fr>

Below the slide, the editor shows a table of attributes for the selected element:

Attribut	Valeur
id	-
lang	-
remap	-
revisionf...	--
role	-

(Enchaînement sur une démonstration en ligne du produit)

QUESTIONS ?

DES QUESTIONS ?