

# $LX_{i\tau}$ : un nouvel outil pour la conversion $\text{\LaTeX} \rightarrow \text{XML}$

*Présentation & résultats préliminaires*

Jean-Paul JORDA & Xavier TROCHU  
EDP Sciences

29 juin 2007

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

1. Génèse et présentation d'*L<sup>X</sup>i<sup>r</sup>*
  - ◆ Quel XML ? Quel LaTeX ? Quel stratégie de conversion L<sup>A</sup>T<sub>E</sub>X → XML
  - ◆ *L<sup>X</sup>i<sup>r</sup>*, une brève présentation (objectifs, architecture)
2. Les arcanes, *L<sup>X</sup>i<sup>r</sup>* à l'œuvre.

# Quel XML ?

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Quel XML ?

◆ Quel L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

L<sup>X</sup>i<sup>r</sup> : cahier des  
charges et  
caractéristiques

Balilage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

- ◆ Quel type d'information veut-on récupérer ? La mise en page ? Les informations sémantiques ? La structure ?
  - *un XML plutôt « sémantique »*
  - *pas ou peu d'information de mise en page*
- ◆ Quelle structure pour le document de sortie ? (précisée par une DTD, un Schema ? )
  - *Une structure proche du document d'origine : pas nécessairement de DTD ou de Schema*
- ◆ Que faire des formules ? des tableaux ? des images ?
  - *du MathML*
  - *du XHTML pour les tableaux*
  - *des liens pour les images*

# Quel $\text{\LaTeX}$ ?

Quel XML ? Quel  $\text{\LaTeX}$  ?

◆ Quel XML ?

Quel  $\text{\LaTeX}$  ?

Stratégies pour la conversion  
 $\text{\LaTeX} \rightarrow \text{XML}$

De  $\text{\LaTeX}$  vers  
MathML

$\text{LX}^i r$  : une brève  
présentation

$\text{LX}^i r$ : cahier des  
charges et  
caractéristiques

Balilage  $\text{\LaTeX}$

Du DVI balisé à  
XML

Conclusions &  
Perspectives

- ◆ Quel type de structuration des documents  $\text{\LaTeX}$  ? Quelles classes ? Quels styles ?
  - *Des documents utilisant « correctement »  $\text{\LaTeX}$ , avec un « nettoyage » qui doit être restreint*
- ◆ Quel type de document ?
  - *En premier lieu de la littérature « académique » des articles scientifiques, des livres, des thèses, ..*
- ◆ Seulement  $\text{\LaTeX}$ , ou aussi plain  $\text{\TeX}$ , Context, Musi $\text{\TeX}$ , ... ?
  - *Seulement  $\text{\LaTeX}$*

Quel XML ? Quel  
 $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$  ?

---

Stratégies pour la  
conversion  
 $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X} \rightarrow \text{XML}$

- ◆ La stratégie  
« source  $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$  »
- ◆ La stratégie  
« DVI balisé »
- ◆ Le convertisseur  
Hermes

De  $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$  vers  
MathML

---

$\text{L}^{\text{X}}i^r$  : une brève  
présentation

---

$\text{L}^{\text{X}}i^r$ : cahier des  
charges et  
caractéristiques

---

Balisage  $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$

---

Du DVI balisé à  
XML

---

Conclusions &  
Perspectives

---

# Stratégies pour la conversion $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X} \rightarrow \text{XML}$

# $\text{\LaTeX}$ $\rightarrow$ XML : pourquoi est-ce compliqué ?

Quel XML ? Quel  $\text{\LaTeX}$  ?

---

Stratégies pour la conversion  $\text{\LaTeX}$  $\rightarrow$ XML

---

- ◆ La stratégie « source  $\text{\LaTeX}$  »
- ◆ La stratégie « DVI balisé »
- ◆ Le convertisseur Hermes

De  $\text{\LaTeX}$  vers MathML

---

$\text{\LX}ir$  : une brève présentation

---

$\text{\LX}ir$ : cahier des charges et caractéristiques

---

Balilage  $\text{\LaTeX}$

---

Du DVI balisé à XML

---

Conclusions & Perspectives

---

- ◆  $\text{\TeX}/\text{\LaTeX}$  est bien documenté mais.... compliqué !
  - ◆  $\text{\LaTeX}$  offre une grande liberté aux auteurs : macros complexes, (re)définition de symboles, ...
  - ◆  $\text{\TeX}/\text{\LaTeX}$  est ancien
  - ◆  $\text{\LaTeX}$  est un vaste monde (près de 2000 classes et styles)
  - ◆  $\text{\LaTeX}$  évolue ( $\sim 10\%$  des classes et styles modifiés entre TL2005 et TL2007) ;
  - ◆ Un document  $\text{\LaTeX}$  a de grandes chances de contenir des maths...
- ➔ Avant tout,  $\text{\TeX}/\text{\LaTeX}$  est un système de *composition*.

Quel XML ? Quel L<sup>A</sup>T<sub>E</sub>X ?

---

Stratégies pour la conversion L<sup>A</sup>T<sub>E</sub>X → XML

---

La stratégie « source L<sup>A</sup>T<sub>E</sub>X »

- ◆ La stratégie « DVI balisé »
- ◆ Le convertisseur Hermes

De L<sup>A</sup>T<sub>E</sub>X vers MathML

---

L<sup>A</sup>X<sub>i</sub>r : une brève présentation

---

L<sup>A</sup>X<sub>i</sub>r : cahier des charges et caractéristiques

---

Balisage L<sup>A</sup>T<sub>E</sub>X

---

Du DVI balisé à XML

---

Conclusions & Perspectives

---

Point de départ :

- ◆ L<sup>A</sup>T<sub>E</sub>X est un langage de balises :
- ◆ il a une structure arborescente ;

➔ *On peut construire un analyseur lexico-grammatical pour le transformer en XML*

Quel XML ? Quel  $\text{\LaTeX}$  ?

Stratégies pour la conversion  $\text{\LaTeX} \rightarrow \text{XML}$

La stratégie « source  $\text{\LaTeX}$  »

- ◆ La stratégie « DVI balisé »
- ◆ Le convertisseur Hermes

De  $\text{\LaTeX}$  vers MathML

$\text{\LaTeX}$ ir : une brève présentation

$\text{\LaTeX}$ ir: cahier des charges et caractéristiques

Balilage  $\text{\LaTeX}$

Du DVI balisé à XML

Conclusions & Perspectives

Projet	Auteur principal	Remarques
$\text{\LaTeX}2\text{HTML}$ tth/ttm	Nikos Drakos Ian Hutchinson	Programme en Perl (1993-2002) Programme en C (partiellement commercial)
LaTeXML	Bruce Miller	Développé pour la <i>Digital Library of Mathematical Functions</i> Perl, programmation OO
tralics	José Grimm	Programme en C++ Développé pour les rapports de l'INRIA. Utilisé par la cellule MathDoc.

D'autres outils ont été développés par les éditeurs pour leur besoins internes.



Quel XML ? Quel  $\text{\LaTeX}$  ?

Stratégies pour la conversion  $\text{\LaTeX}$ →XML

La stratégie  
« source  $\text{\LaTeX}$  »

- ◆ La stratégie « DVI balisé »
- ◆ Le convertisseur Hermes

De  $\text{\LaTeX}$  vers MathML

$\text{\LaTeX}$ ir : une brève présentation

$\text{\LaTeX}$ ir: cahier des charges et caractéristiques

Balisage  $\text{\LaTeX}$

Du DVI balisé à XML

Conclusions & Perspectives

- + Ne nécessite pas de distribution  $\text{\TeX}$
- + Relativement simple à mettre en œuvre, pour des documents simples et propres
- Extrêmement compliquée pour des documents moins simples et moins propres !

Quel XML ? Quel  
 $\text{\LaTeX}$  ?

---

Stratégies pour la  
conversion  
 $\text{\LaTeX} \rightarrow \text{XML}$

---

◆ La stratégie  
« source  $\text{\LaTeX}$  »

La stratégie  
« DVI balisé »

◆ Le convertisseur  
Hermes

De  $\text{\LaTeX}$  vers  
MathML

---

$\text{LX}\text{\LaTeX}$  : une brève  
présentation

---

$\text{LX}\text{\LaTeX}$ : cahier des  
charges et  
caractéristiques

---

Balilage  $\text{\LaTeX}$

---

Du DVI balisé à  
XML

---

Conclusions &  
Perspectives

---

Idées de base (énoncées pas [Sofka, 1998]) :

- ◆ Le moteur  $\text{T}\text{E}\text{X}$  est le meilleur analyseur de document  $\text{\LaTeX}$  qui soit ;
- ◆ On peut « baliser » le DVI produit, en utilisant la commande  $\text{T}\text{E}\text{X} \backslash \text{special}$  ;
- ◆ Le DVI « balisé » peut être analysé pour produire du XML (ou tout autre format)

➔ *C'est la stratégie adoptée par  $\text{LX}\text{\LaTeX}$*

Quel XML ? Quel  
 $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$  ?

---

Stratégies pour la  
conversion  
 $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X} \rightarrow \text{XML}$

---

◆ La stratégie  
« source  $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$  »

La stratégie  
« DVI balisé »

◆ Le convertisseur  
Hermes

De  $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$  vers  
MathML

---

$\text{L}^{\text{X}}i^r$  : une brève  
présentation

---

$\text{L}^{\text{X}}i^r$ : cahier des  
charges et  
caractéristiques

---

Balisage  $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$

---

Du DVI balisé à  
XML

---

Conclusions &  
Perspectives

---

Le DVI (**D**e**V**ice Independent) est le format de sortie de  $\text{T}_{\text{E}}\text{X}$ .  
Un fichier DVI contient une suite d'instructions simples :

- ◆ pour le placement des glyphes ;
- ◆ pour le dessin des traits ;
- ◆ « spéciales », destinées au post-treatment.

Le DVI est un format binaire.

# Critique de la stratégie « DVI balisé »

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

◆ La stratégie  
« source L<sup>A</sup>T<sub>E</sub>X »

◆ La stratégie  
« DVI balisé »

◆ Le convertisseur  
Hermes

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

- Nécessite de reconstruire les caractères et les mots
- Nécessite une ré-écriture de macros parfois compliquée
- + L'analyse du document source est réalisée par T<sub>E</sub>X : pas besoin d'écrire un analyseur complexe

Quel XML ? Quel  $\text{\LaTeX}$  ?

Stratégies pour la conversion  $\text{\LaTeX} \rightarrow \text{XML}$

◆ La stratégie « source  $\text{\LaTeX}$  »

La stratégie « DVI balisé »

◆ Le convertisseur Hermes

De  $\text{\LaTeX}$  vers MathML

$\text{LX}i\text{r}$  : une brève présentation

$\text{LX}i\text{r}$ : cahier des charges et caractéristiques

Balisage  $\text{\LaTeX}$

Du DVI balisé à XML

Conclusions & Perspectives

Projet	Auteur principal	Remarques
tex4ht	E. M. Gurari	Produit divers types de XML Probablement le plus utilisé
<b>Hermes</b>	R. Anghelache	Produit un XML de structure proche de celle du source $\text{\LaTeX}$ Convertisseur DVI $\rightarrow$ XML en C, lex et bison. Développé pour le projet « Living Reviews ».

➔ *Le point de départ d' $\text{LX}i\text{r}$  est **Hermes***

Quel XML ? Quel L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la conversion L<sup>A</sup>T<sub>E</sub>X → XML

◆ La stratégie « source L<sup>A</sup>T<sub>E</sub>X »

◆ La stratégie « DVI balisé »

**Le convertisseur Hermes**

De L<sup>A</sup>T<sub>E</sub>X vers MathML

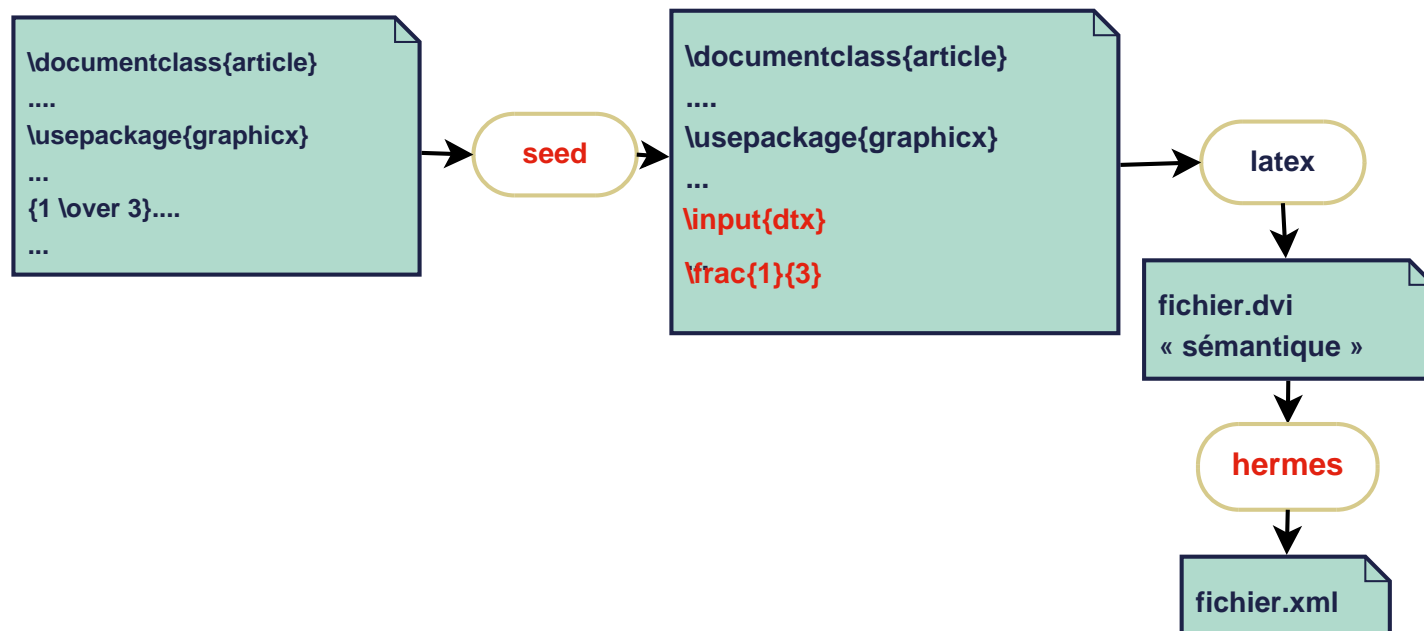
L<sup>X</sup>i<sup>r</sup> : une brève présentation

L<sup>X</sup>i<sup>r</sup>: cahier des charges et caractéristiques

Balilage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à XML

Conclusions & Perspectives



➔ *dtx.tex* : redéfinition de macros pour introduire des balises.

➔ *Un lexique flex de 2000 lignes, une grammaire bison de 11 000 lignes*

Quel XML ? Quel  
L<sub>A</sub>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sub>A</sub>T<sub>E</sub>X → XML

De L<sub>A</sub>T<sub>E</sub>X vers  
MathML

**Le « mode  
math » en T<sub>E</sub>X**

◆ L<sub>A</sub>T<sub>E</sub>X → MathML  
: stratégie  
classique

◆ La conversion  
vers MathML  
selon L<sub>X</sub>i<sub>r</sub>

L<sub>X</sub>i<sub>r</sub> : une brève  
présentation

L<sub>X</sub>i<sub>r</sub>: cahier des  
charges et  
caractéristiques

Balisage L<sub>A</sub>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

Le mode « math » en T<sub>E</sub>X est utilisé pour :

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

**Le « mode  
math » en T<sub>E</sub>X**

◆ L<sup>A</sup>T<sub>E</sub>X → MathML  
: stratégie  
classique

◆ La conversion  
vers MathML  
selon L<sup>X</sup>i<sup>r</sup>

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

Le mode « math » en T<sub>E</sub>X est utilisé pour :

- ◆ les tableaux ;



Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

**Le « mode  
math » en T<sub>E</sub>X**

◆ L<sup>A</sup>T<sub>E</sub>X → MathML  
: stratégie  
classique

◆ La conversion  
vers MathML  
selon L<sup>X</sup>i<sup>r</sup>

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

Le mode « math » en T<sub>E</sub>X est utilisé pour :

- ◆ les tableaux ;
- ◆ les notes dans la marge ;

Quel XML ? Quel  
L<sub>A</sub>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sub>A</sub>T<sub>E</sub>X→XML

De L<sub>A</sub>T<sub>E</sub>X vers  
MathML

**Le « mode  
math » en T<sub>E</sub>X**

◆ L<sub>A</sub>T<sub>E</sub>X→MathML  
: stratégie  
classique

◆ La conversion  
vers MathML  
selon L<sub>X</sub>i<sub>r</sub>

L<sub>X</sub>i<sub>r</sub> : une brève  
présentation

L<sub>X</sub>i<sub>r</sub>: cahier des  
charges et  
caractéristiques

Balisage L<sub>A</sub>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

Le mode « math » en T<sub>E</sub>X est utilisé pour :

- ◆ les tableaux ;
- ◆ les notes dans la marge ;
- ◆ les formules chimiques ;

Quel XML ? Quel L<sub>A</sub>T<sub>E</sub>X ?

Stratégies pour la conversion L<sub>A</sub>T<sub>E</sub>X → XML

De L<sub>A</sub>T<sub>E</sub>X vers MathML

**Le « mode math » en T<sub>E</sub>X**

◆ L<sub>A</sub>T<sub>E</sub>X → MathML : stratégie classique

◆ La conversion vers MathML selon L<sub>X</sub>i<sub>r</sub>

L<sub>X</sub>i<sub>r</sub> : une brève présentation

L<sub>X</sub>i<sub>r</sub>: cahier des charges et caractéristiques

Balisage L<sub>A</sub>T<sub>E</sub>X

Du DVI balisé à XML

Conclusions & Perspectives

Le mode « math » en T<sub>E</sub>X est utilisé pour :

- ◆ les tableaux ;
- ◆ les notes dans la marge ;
- ◆ les formules chimiques ;
- ◆ l’affichage de certains symboles ;

Quel XML ? Quel L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la conversion L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers MathML

**Le « mode math » en T<sub>E</sub>X**

◆ L<sup>A</sup>T<sub>E</sub>X → MathML : stratégie classique

◆ La conversion vers MathML selon L<sup>X</sup>i<sup>r</sup>

L<sup>X</sup>i<sup>r</sup> : une brève présentation

L<sup>X</sup>i<sup>r</sup>: cahier des charges et caractéristiques

Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à XML

Conclusions & Perspectives

Le mode « math » en T<sub>E</sub>X est utilisé pour :

- ◆ les tableaux ;
- ◆ les notes dans la marge ;
- ◆ les formules chimiques ;
- ◆ l’affichage de certains symboles ;
- ◆ Les indices (notes de bas de page) ;

# Le « mode math » en T<sub>E</sub>X

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

---

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

---

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

Le « mode  
math » en T<sub>E</sub>X

◆ L<sup>A</sup>T<sub>E</sub>X → MathML  
: stratégie  
classique

◆ La conversion  
vers MathML  
selon L<sup>X</sup>i<sub>r</sub>

L<sup>X</sup>i<sub>r</sub> : une brève  
présentation

---

L<sup>X</sup>i<sub>r</sub>: cahier des  
charges et  
caractéristiques

---

Balisage L<sup>A</sup>T<sub>E</sub>X

---

Du DVI balisé à  
XML

---

Conclusions &  
Perspectives

---

Le mode « math » en T<sub>E</sub>X est utilisé pour :

- ◆ les tableaux ;
- ◆ les notes dans la marge ;
- ◆ les formules chimiques ;
- ◆ l’affichage de certains symboles ;
- ◆ Les indices (notes de bas de page) ;
- ◆ ...
- ◆ et même pour les formules mathématiques !

➡ *Il ne faut pas tout convertir en MathML*

Quel XML ? Quel  $\text{\LaTeX}$  ?

Stratégies pour la conversion  $\text{\LaTeX}$  → XML

De  $\text{\LaTeX}$  vers MathML

◆ Le « mode math » en  $\text{\TeX}$

$\text{\LaTeX}$  → MathML : stratégie classique

◆ La conversion vers MathML selon  $\text{\Lxir}$

$\text{\Lxir}$  : une brève présentation

$\text{\Lxir}$ : cahier des charges et caractéristiques

Balilage  $\text{\LaTeX}$

Du DVI balisé à XML

Conclusions & Perspectives

- ◆ MathML : empilement de boîtes, identifiants, nombres et opérateurs ;
- ◆ Source  $\text{\TeX}/\text{\LaTeX}$  : suite de commandes plus ou moins imbriquées, opérateurs ;

Stratégie habituelle :

- ◆ balisage des commandes mathématiques ;
- ◆ grammaire pour la construction de MathML ;

Des problèmes liés à certaines structures de  $\text{\TeX}$  :  $\backslash\text{over}$ , etc ...

➔  *$\text{\Lxir}$  utilise une stratégie alternative pour la conversion vers MathML*

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

---

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

---

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

---

- ◆ Le « mode math » en T<sub>E</sub>X
- ◆ L<sup>A</sup>T<sub>E</sub>X → MathML : stratégie classique

La conversion  
vers MathML  
selon *LXir*

*LXir* : une brève  
présentation

---

*LXir*: cahier des  
charges et  
caractéristiques

---

Balisage L<sup>A</sup>T<sub>E</sub>X

---

Du DVI balisé à  
XML

---

Conclusions &  
Perspectives

---

Idées de base :

- ◆ MathML est un empilement de boîtes ;
  - ◆ Le moteur T<sub>E</sub>X construit et positionne des boîtes ;
  - ◆ La procédure de construction des boîtes est accessible dans le fichier de log (avec `\tracingoutput=1`) ;
- ➔ *LXir* analyse les logs pour construire des formules en MathML.

Quel XML ? Quel L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la conversion L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers MathML

◆ Le « mode math » en T<sub>E</sub>X

◆ L<sup>A</sup>T<sub>E</sub>X → MathML : stratégie classique

La conversion vers MathML selon L<sup>X</sup>i<sup>r</sup>

L<sup>X</sup>i<sup>r</sup> : une brève présentation

L<sup>X</sup>i<sup>r</sup>: cahier des charges et caractéristiques

Balilage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à XML

Conclusions & Perspectives

[latexml] Le site de LaTeXML

<http://d1mf.nist.gov/LaTeXML/>

[tralics] Le site de Tralics

<http://www-sop.inria.fr/apics/tralics/>

[Sofka, 1998] Sofka M.D., *T<sub>E</sub>X to HTML Translation via Tagged DVI Files*, TUGboat **19** (1998), p. 214.

[hermes] Le site du convertisseur Hermes :

<http://hermes.roua.org/>

[tex4ht] Le site du convertisseur tex4ht :

<http://www.cse.ohio-state.edu/~gurari/TeX4ht/mn.html>



Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

---

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

---

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

---

**L<sup>X</sup>i<sup>r</sup> : une brève  
présentation**

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

---

Balisage L<sup>A</sup>T<sub>E</sub>X

---

Du DVI balisé à  
XML

---

Conclusions &  
Perspectives

---

# *L<sup>X</sup>i<sup>r</sup> : une brève présentation*

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

---

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

---

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

---

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

---

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

◆ Processus de  
conversion avec L<sup>X</sup>i<sup>r</sup>

◆ L<sup>X</sup>i<sup>r</sup> dans une  
chaîne de  
conversion

Balisage L<sup>A</sup>T<sub>E</sub>X

---

Du DVI balisé à  
XML

---

Conclusions &  
Perspectives

---

On s'intéresse à la conversion de documents :

- ◆ avant tout « académiques » (articles, thèses, rapports, ...);
- ◆ éventuellement complexes;
- ◆ qui « compilent » en L<sup>A</sup>T<sub>E</sub>X;
- ◆ utilisant des classes et des styles pris en charge par l'outil.

...

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

◆ Processus de  
conversion avec L<sup>X</sup>i<sup>r</sup>

◆ L<sup>X</sup>i<sup>r</sup> dans une  
chaîne de  
conversion

Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

... vers du XML :

- ◆ contenant les informations « sémantiques » et structurelles du document L<sup>A</sup>T<sub>E</sub>X ;

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

◆ Processus de  
conversion avec L<sup>X</sup>i<sup>r</sup>

◆ L<sup>X</sup>i<sup>r</sup> dans une  
chaîne de  
conversion

Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

... vers du XML :

- ◆ contenant les informations « sémantiques » et structurelles du document L<sup>A</sup>T<sub>E</sub>X ;
- ◆ ne contenant pas (ou peu) d'information de mise en page ;

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

◆ Processus de  
conversion avec L<sup>X</sup>i<sup>r</sup>

◆ L<sup>X</sup>i<sup>r</sup> dans une  
chaîne de  
conversion

Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

... vers du XML :

- ◆ contenant les informations « sémantiques » et structurelles du document L<sup>A</sup>T<sub>E</sub>X ;
- ◆ ne contenant pas (ou peu) d'information de mise en page ;
- ◆ avec des formules de MathML ;

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

◆ Processus de  
conversion avec L<sup>X</sup>i<sup>r</sup>

◆ L<sup>X</sup>i<sup>r</sup> dans une  
chaîne de  
conversion

Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

... vers du XML :

- ◆ contenant les informations « sémantiques » et structurelles du document L<sup>A</sup>T<sub>E</sub>X ;
- ◆ ne contenant pas (ou peu) d'information de mise en page ;
- ◆ avec des formules de MathML ;
- ◆ avec des tableaux de XHTML ;

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

◆ Processus de  
conversion avec L<sup>X</sup>i<sup>r</sup>

◆ L<sup>X</sup>i<sup>r</sup> dans une  
chaîne de  
conversion

Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

... vers du XML :

- ◆ contenant les informations « sémantiques » et structurelles du document L<sup>A</sup>T<sub>E</sub>X ;
- ◆ ne contenant pas (ou peu) d'information de mise en page ;
- ◆ avec des formules de MathML ;
- ◆ avec des tableaux de XHTML ;
- ◆ ...

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

◆ Processus de  
conversion avec L<sup>X</sup>i<sup>r</sup>

◆ L<sup>X</sup>i<sup>r</sup> dans une  
chaîne de  
conversion

Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

... avec un outil qui doit être :

- ◆ utilisable facilement (peu de modifications du document L<sup>A</sup>T<sub>E</sub>X) ;



Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

◆ Processus de  
conversion avec L<sup>X</sup>i<sup>r</sup>

◆ L<sup>X</sup>i<sup>r</sup> dans une  
chaîne de  
conversion

Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

... avec un outil qui doit être :

- ◆ utilisable facilement (peu de modifications du document L<sup>A</sup>T<sub>E</sub>X) ;
- ◆ modulaire ;

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>*ir* : une brève  
présentation

L<sup>X</sup>*ir*: cahier des  
charges et  
caractéristiques

◆ Processus de  
conversion avec L  
*Xir*

◆ L<sup>X</sup>*ir* dans une  
chaîne de  
conversion

Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

... avec un outil qui doit être :

- ◆ utilisable facilement (peu de modifications du document L<sup>A</sup>T<sub>E</sub>X) ;
- ◆ modulaire ;
- ◆ s'appuyant sur les technologies XML (DOM, XSLT, ...);

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>*ir* : une brève  
présentation

L<sup>X</sup>*ir*: cahier des  
charges et  
caractéristiques

◆ Processus de  
conversion avec L<sup>X</sup>*ir*

◆ L<sup>X</sup>*ir* dans une  
chaîne de  
conversion

Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

... avec un outil qui doit être :

- ◆ utilisable facilement (peu de modifications du document L<sup>A</sup>T<sub>E</sub>X) ;
- ◆ modulaire ;
- ◆ s'appuyant sur les technologies XML (DOM, XSLT, ...);
- ◆ configurable aussi facilement que possible ;

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>*ir* : une brève  
présentation

L<sup>X</sup>*ir*: cahier des  
charges et  
caractéristiques

◆ Processus de  
conversion avec L<sup>X</sup>*ir*

◆ L<sup>X</sup>*ir* dans une  
chaîne de  
conversion

Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

... avec un outil qui doit être :

- ◆ utilisable facilement (peu de modifications du document L<sup>A</sup>T<sub>E</sub>X) ;
- ◆ modulaire ;
- ◆ s'appuyant sur les technologies XML (DOM, XSLT, ...) ;
- ◆ configurable aussi facilement que possible ;
- ◆ utilisable sous linux et Windows ;

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>*ir* : une brève  
présentation

L<sup>X</sup>*ir*: cahier des  
charges et  
caractéristiques

◆ Processus de  
conversion avec L  
*Xir*

◆ L<sup>X</sup>*ir* dans une  
chaîne de  
conversion

Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

... avec un outil qui doit être :

- ◆ utilisable facilement (peu de modifications du document L<sup>A</sup>T<sub>E</sub>X) ;
- ◆ modulaire ;
- ◆ s'appuyant sur les technologies XML (DOM, XSLT, ...) ;
- ◆ configurable aussi facilement que possible ;
- ◆ utilisable sous linux et Windows ;
- ◆ facilement intégrable dans une distribution T<sub>E</sub>X ;

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>*ir* : une brève  
présentation

L<sup>X</sup>*ir*: cahier des  
charges et  
caractéristiques

◆ Processus de  
conversion avec L<sup>X</sup>*ir*

◆ L<sup>X</sup>*ir* dans une  
chaîne de  
conversion

Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

... avec un outil qui doit être :

- ◆ utilisable facilement (peu de modifications du document L<sup>A</sup>T<sub>E</sub>X) ;
- ◆ modulaire ;
- ◆ s'appuyant sur les technologies XML (DOM, XSLT, ...) ;
- ◆ configurable aussi facilement que possible ;
- ◆ utilisable sous linux et Windows ;
- ◆ facilement intégrable dans une distribution T<sub>E</sub>X ;
- ◆ documenté ;

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

◆ Processus de  
conversion avec L<sup>X</sup>i<sup>r</sup>

◆ L<sup>X</sup>i<sup>r</sup> dans une  
chaîne de  
conversion

Balilage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

... avec un outil qui doit être :

- ◆ utilisable facilement (peu de modifications du document L<sup>A</sup>T<sub>E</sub>X) ;
- ◆ modulaire ;
- ◆ s'appuyant sur les technologies XML (DOM, XSLT, ...) ;
- ◆ configurable aussi facilement que possible ;
- ◆ utilisable sous linux et Windows ;
- ◆ facilement intégrable dans une distribution T<sub>E</sub>X ;
- ◆ documenté ;
- ◆ libre et ouvert.

Quel XML ? Quel L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la conversion L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers MathML

*LXir* : une brève présentation

*LXir*: cahier des charges et caractéristiques

**Processus de conversion avec *LXir***

◆ *LXir* dans une chaîne de conversion

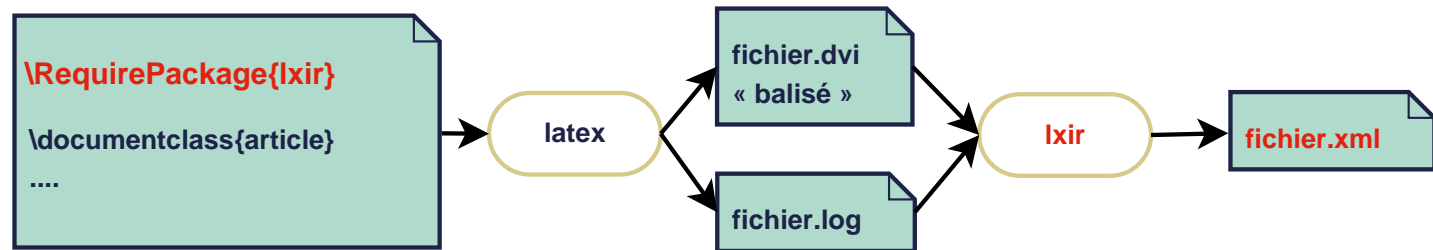
Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à XML

Conclusions & Perspectives

La conversion se fait en deux étapes :

1. Compilation latex, en ajoutant `\RequirePackage` avant le `\documentclass` ;  
 ➔ *On obtient un DVI « balisé »*
2. Lancement de la commande `lxir fichier.dvi`, qui génère le XML.





Quel XML ? Quel L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la conversion L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers MathML

LXir : une brève présentation

LXir: cahier des charges et caractéristiques

◆ Processus de conversion avec LXir

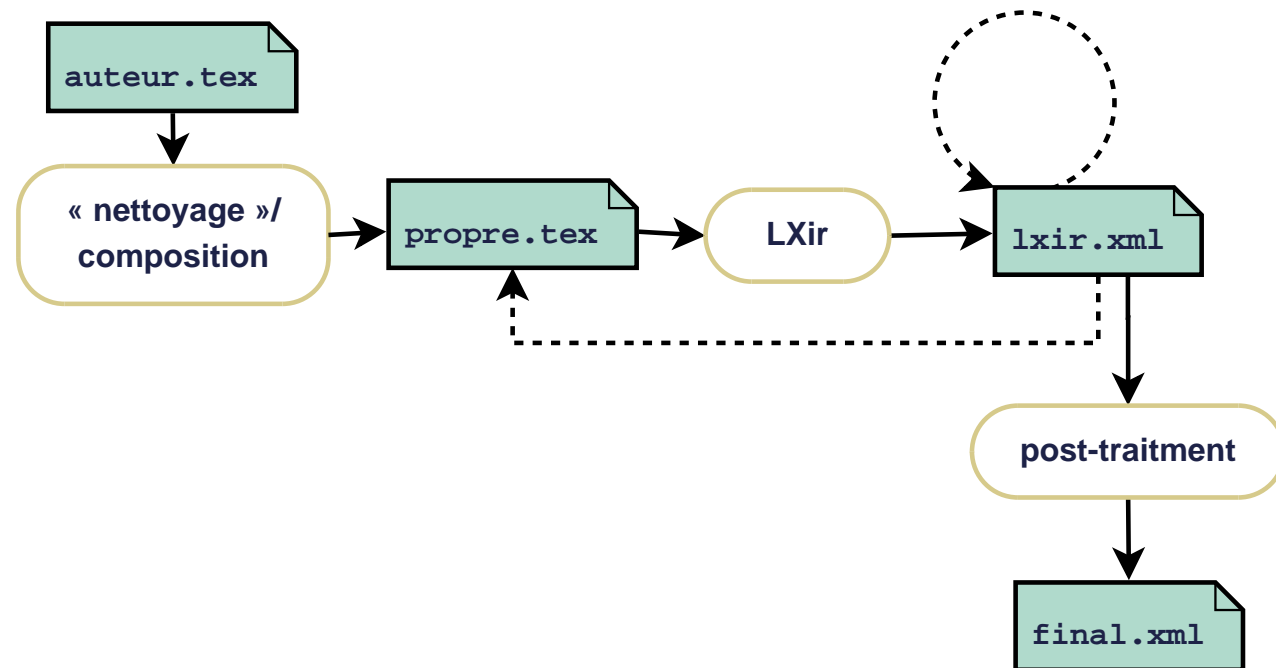
**LXir dans une chaîne de conversion**

Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à XML

Conclusions & Perspectives

LXir devra en général s'insérer dans une chaîne de conversion.



Quel XML ? Quel  
 $L_{A}T_{E}X$  ?

---

Stratégies pour la  
conversion  
 $L_{A}T_{E}X \rightarrow XML$

---

De  $L_{A}T_{E}X$  vers  
MathML

---

$L_{X}i_{r}$  : une brève  
présentation

---

$L_{X}i_{r}$ : cahier des  
charges et  
caractéristiques

---

Balilage  $L_{A}T_{E}X$

◆ Classes et styles  
balisés

Du DVI balisé à  
XML

---

Conclusions &  
Perspectives

---

[Sofka, 1998]

- ◆ Balisage de la séparation des mots : séparation par 1sp ;
- ◆ `\let\@@italiccorr=\relax;`
- ◆ On empêche les césures ;
- ◆ `\pagestyle{empty};`
- ◆ On empêche autant que possible les sauts de page

Quel XML ? Quel  $\LaTeX$  ?

Stratégies pour la conversion  $\LaTeX \rightarrow$  XML

De  $\LaTeX$  vers MathML

$LXir$  : une brève présentation

$LXir$ : cahier des charges et caractéristiques

Balisage  $\LaTeX$

◆ Classes et styles balisés

Du DVI balisé à XML

Conclusions & Perspectives

Définies dans *lxirtag.sty* :

- ◆  $\xBegin\{mytag\} \dots \xEnd\{myTag\}$   
 $\rightarrow \langle \text{span class}="mytag" / \rangle \dots \langle / \text{span} \rangle$
  - ◆  $\xEmpty\{mytag\}$   
 $\rightarrow \langle \text{span class}="mytag" / \rangle$
  - ◆  $\xEmptyA\{mytag\}\{attr=val\}$   
 $\rightarrow \langle \text{span class}="mytag" attr="val" / \rangle$
  - ◆  $\ift@g0n$  : Activateur de balisage
- *Le balisage est intégrable directement dans un nouveau style.*

Quel XML ? Quel  $\text{\LaTeX}$  ?

Stratégies pour la conversion  $\text{\LaTeX} \rightarrow \text{XML}$

De  $\text{\LaTeX}$  vers MathML

$\text{LX}^i\text{r}$  : une brève présentation

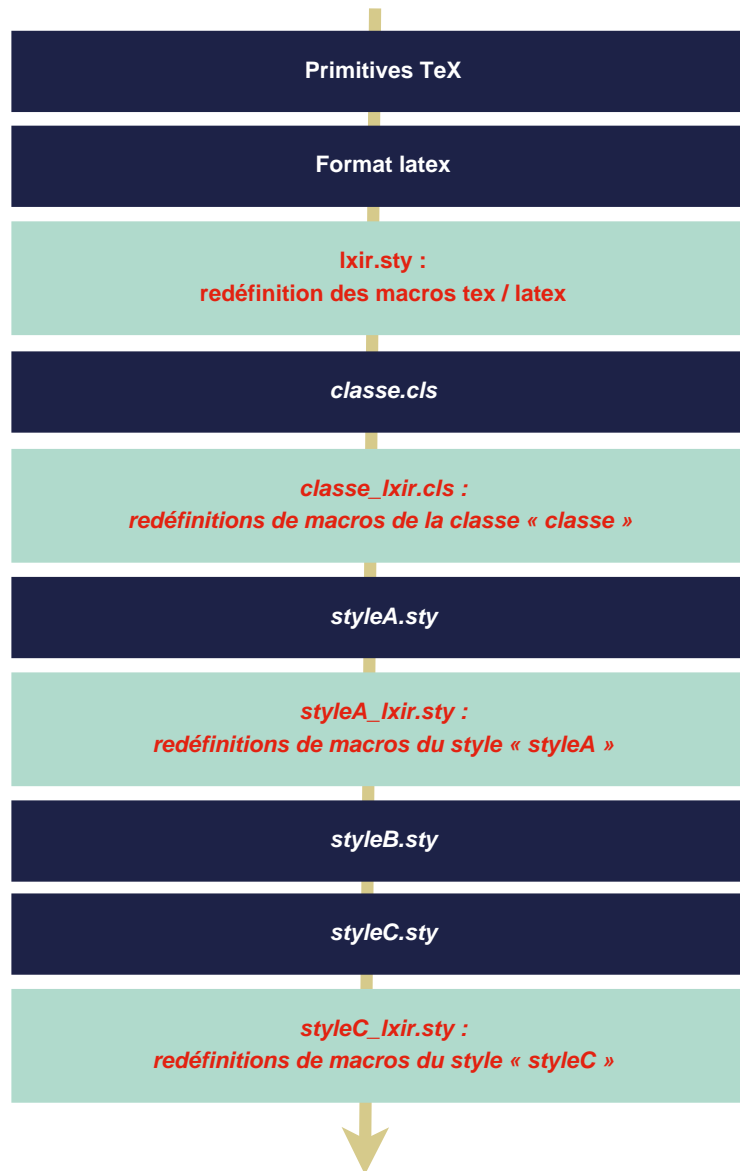
$\text{LX}^i\text{r}$ : cahier des charges et caractéristiques

Balisage  $\text{\LaTeX}$

◆ Classes et styles balisés

Du DVI balisé à XML

Conclusions & Perspectives



Modification du source  $\text{\LaTeX}$  :

```

\RequirePackage{lxir}
\documentclass{classe}
\usepackage{styleA}
\usepackage{styleB}
\usepackage{styleC}

```

Quel XML ? Quel L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la conversion L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers MathML

L<sup>X</sup>i<sup>r</sup> : une brève présentation

L<sup>X</sup>i<sup>r</sup>: cahier des charges et caractéristiques

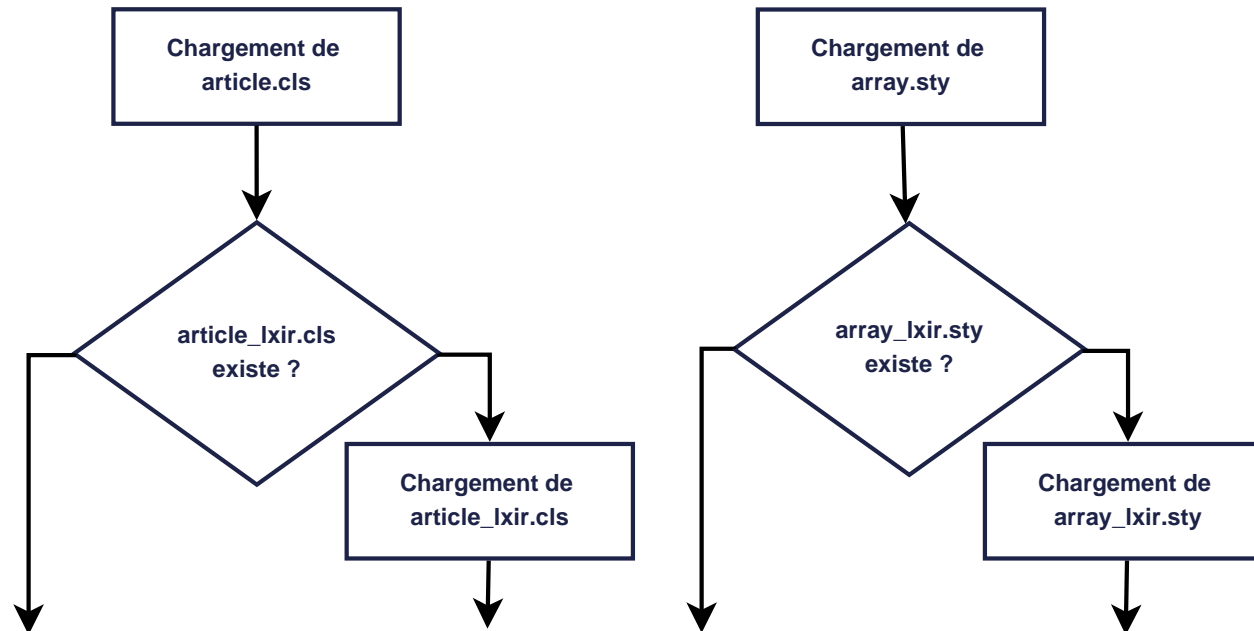
Balilage L<sup>A</sup>T<sub>E</sub>X

◆ Classes et styles balisés

Du DVI balisé à XML

Conclusions & Perspectives

Redéfinition dans *lxir.sty* du mécanisme de chargement des classes et des styles :



Quel XML ? Quel L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la conversion L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers MathML

L<sup>X</sup>i<sup>r</sup> : une brève présentation

L<sup>X</sup>i<sup>r</sup>: cahier des charges et caractéristiques

Balisage L<sup>A</sup>T<sub>E</sub>X

◆ Classes et styles balisés

Du DVI balisé à XML

Conclusions & Perspectives

Nous avons utilisé plusieurs stratégies pour baliser les macros :

- ◆ Encadrement :

```
\let\t@gfbox=\fbox
\def\fbox#1{%
  \xBegin{fbox}\t@gfbox{#1}\xEnd{fbox}}
```

- ◆ Simplification :

```
\def\@xfloat #1[#2]{%
  \xBegin{float}\xEmptyA{params}{floatType=#1}
  \def \@captive {#1}}
  \def\end@float{\xEnd{float}}
```

- ◆ À façon (par ex. pour les tableaux)

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X→XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>ir : une brève  
présentation

L<sup>X</sup>ir: cahier des  
charges et  
caractéristiques

Balisage L<sup>A</sup>T<sub>E</sub>X

◆ Classes et styles  
balisés

Du DVI balisé à  
XML

Conclusions &  
Perspectives

Le balisage des maths est relativement simple :

- ◆ Activation de logs enrichis ;
- ◆ Balisage des environnements mathématiques ;
- ◆ Balisage du début et de la fin du mode mathématique ;
- ◆ Balisage des opérateurs mathématiques.

Dans `texmf.cnf`, il faut autoriser les longues lignes dans les fichiers de log :

```
max_print_line = 1000 % For LXir
```

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

Balilage L<sup>A</sup>T<sub>E</sub>X

**Classes et styles  
balisés**

Du DVI balisé à  
XML

Conclusions &  
Perspectives

Quelques classes et styles ont déjà été balisés :

Classe ou style	État
aa.cls	OK
article.cls	OK
book.cls	OK
report.cls	OK
array.sty	OK
natbib.sty	À tester
tabularx.sty	À finir
enumerate.sty	OK
babel.sty	À faire
graphicx.sty	OK
color.sty	OK



# Du DVI à XML : schéma de principe

Quel XML ? Quel L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la conversion L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers MathML

L<sup>X</sup>i<sup>r</sup> : une brève présentation

L<sup>X</sup>i<sup>r</sup>: cahier des charges et caractéristiques

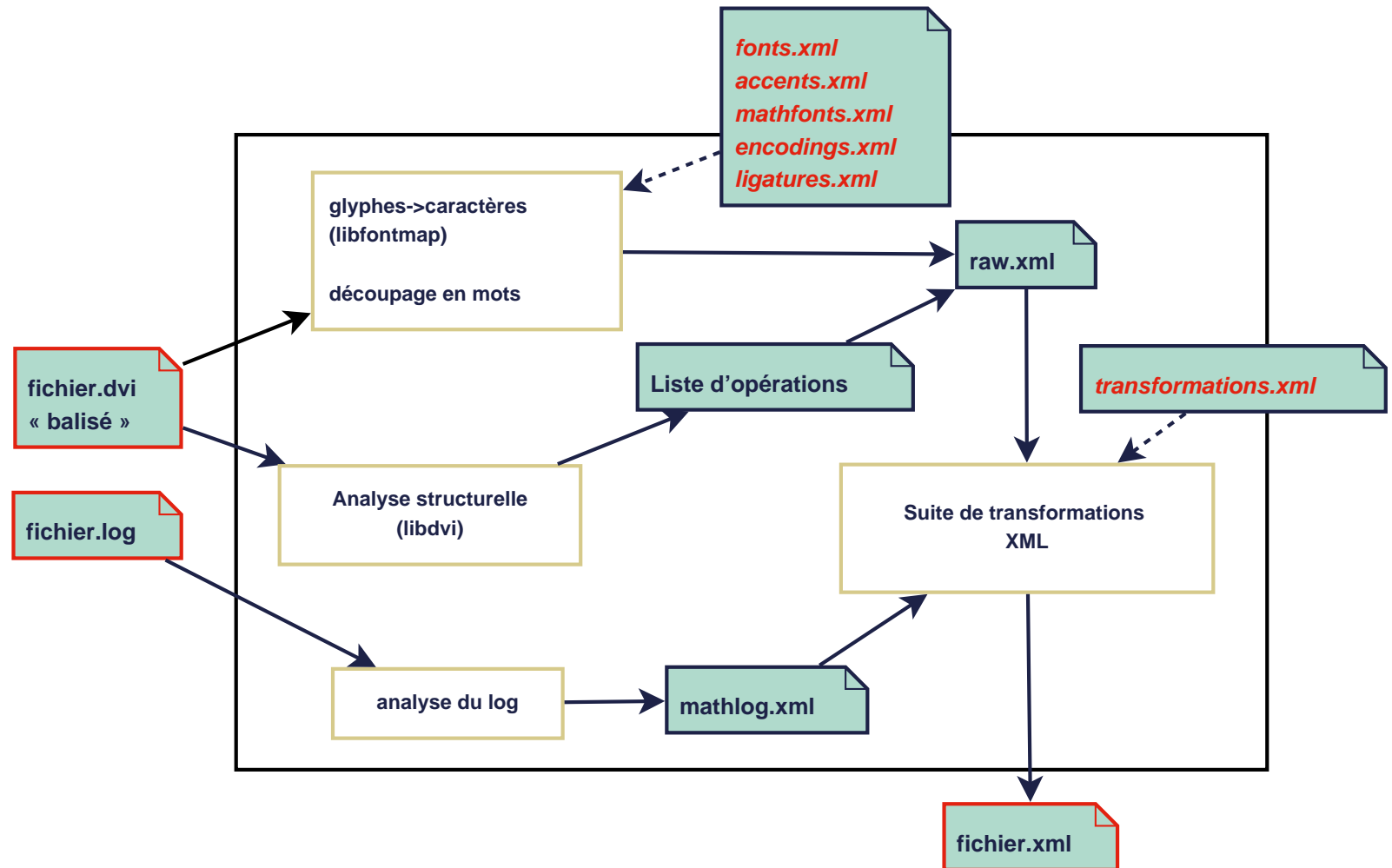
Balilage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à XML

◆ Du DVI à XML : détails

◆ Chaîne de transformation XML

Conclusions & Perspectives



Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

---

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

---

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

---

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

---

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

---

Balilage L<sup>A</sup>T<sub>E</sub>X

---

Du DVI balisé à  
XML

---

Du DVI à XML :  
détails

◆ Chaîne de  
transformation  
XML

Conclusions &  
Perspectives

---

- ◆ **l<sup>x</sup>i<sup>r</sup>** : Programme en C, utilise *libxml2* ;
- ◆ Deux bibliothèques en C : *libfontmap*, *libdvi* ;
- ◆ *fonts.xml*, *mathfonts.xml*, *ligatures.xml* et *encodings.xml* : bases de données pour la conversion glyphes → caractères ;
- ◆ *transformations.xml* : Description de la chaîne de transformation XML.
- ◆ Format XML de sortie : XHTML « enrichi » : visualisable avec firefox (feuille de style CSS).

Quel XML ? Quel L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la conversion L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers MathML

L<sup>X</sup>i<sup>r</sup> : une brève présentation

L<sup>X</sup>i<sup>r</sup>: cahier des charges et caractéristiques

Balilage L<sup>A</sup>T<sub>E</sub>X

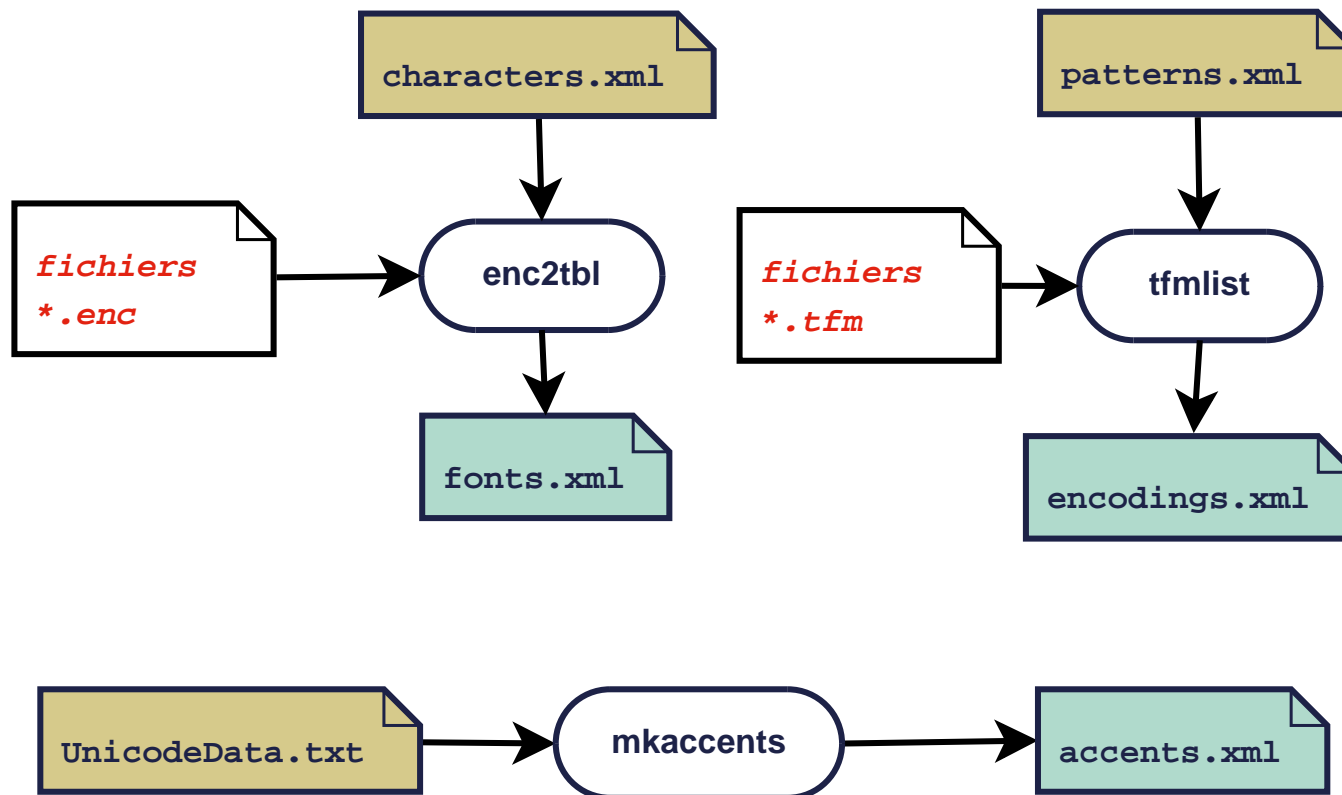
Du DVI balisé à XML

**Du DVI à XML : détails**

◆ Chaîne de transformation XML

Conclusions & Perspectives

Des programmes permettent de générer les fichiers de données pour les fontes :



Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

Balilage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

◆ Du DVI à XML :  
détails

Chaîne de  
transformation  
XML

Conclusions &  
Perspectives

Une cinquantaine d'étapes élémentaires du « raw.xml » au XML final.

Selon les transformations à réaliser :

- ◆ réorganisation de l'arbre XML, avec des fonctions C en DOM/libxml2
- ◆ transformation de l'arbre XML, avec des transformations XSLT

Le fichier `transformations.xml` décrit :

- ◆ le type de transformation (fonction C, transformation XSLT, script extérieur) ;
- ◆ L'ordre des transformations ;
- ◆ si la transformation génère un fichier XML.

pour le texte et pour les mathématiques.

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

---

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

---

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

---

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

---

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

---

Balilage L<sup>A</sup>T<sub>E</sub>X

---

Du DVI balisé à  
XML

---

**Conclusions &  
Perspectives**

- ◆ Le travail des prochains mois
- ◆ Le futur site Web
- ◆ Conclusions

# *Conclusions & Perspectives*

Quel XML ? Quel  
 $\text{\LaTeX}$  ?

Stratégies pour la  
conversion  
 $\text{\LaTeX} \rightarrow \text{XML}$

De  $\text{\LaTeX}$  vers  
MathML

$\text{LX}^i\text{r}$  : une brève  
présentation

$\text{LX}^i\text{r}$ : cahier des  
charges et  
caractéristiques

Balilage  $\text{\LaTeX}$

Du DVI balisé à  
XML

Conclusions &  
Perspectives

**Le travail des  
prochains mois**

◆ Le futur site  
Web

◆ Conclusions

- ◆ Choix de la license libre (GPL, LGPL ?)
  - ◆ Intégration dans cyberthèse
  - ◆ Tests intensifs, débogages
  - ◆ Développements : balisage de nouvelles classes et de nouveaux styles, sortie des formules en  $\text{\LaTeX}$ , optimisations, ... ;
  - ◆ Documentation
  - ◆ Diffusion : création du site web, présentations, articles, ...
- ➔ *Objectif : une version diffusable largement et un site Web cet automne.*

Quel XML ? Quel L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la conversion L<sup>A</sup>T<sub>E</sub>X→XML

De L<sup>A</sup>T<sub>E</sub>X vers MathML

L<sup>X</sup>i<sup>r</sup> : une brève présentation

L<sup>X</sup>i<sup>r</sup>: cahier des charges et caractéristiques

Balilage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à XML

Conclusions & Perspectives

◆ Le travail des prochains mois

**Le futur site Web**

◆ Conclusions

<http://www.latex-lxir.org> permettra d'accéder :

- ◆ aux sources (serveur SVN) ;
- ◆ aux binaires pour Windows ;
- ◆ à une liste de discussion et/ou de diffusion ;
- ◆ à la documentation ;
- ◆ ...

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

Balilage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

◆ Le travail des  
prochains mois

◆ Le futur site  
Web

Conclusions

◆ L<sup>X</sup>i<sup>r</sup> devrait atteindre les objectifs fixés



Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

Balilage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

- ◆ Le travail des prochains mois
- ◆ Le futur site Web

Conclusions

- ◆ L<sup>X</sup>i<sup>r</sup> devrait atteindre les objectifs fixés
- ◆ Le projet est amené à évoluer

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

◆ Le travail des  
prochains mois

◆ Le futur site  
Web

Conclusions

- ◆ L<sup>X</sup>i<sup>r</sup> devrait atteindre les objectifs fixés
- ◆ Le projet est amené à évoluer
- ◆ Il sera ouvert aux contributions

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

◆ Le travail des  
prochains mois

◆ Le futur site  
Web

Conclusions

- ◆ L<sup>X</sup>i<sup>r</sup> devrait atteindre les objectifs fixés
- ◆ Le projet est amené à évoluer
- ◆ Il sera ouvert aux contributions
- ◆ Il sera utilisé :

Quel XML ? Quel L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la conversion L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers MathML

L<sup>X</sup>i<sup>r</sup> : une brève présentation

L<sup>X</sup>i<sup>r</sup>: cahier des charges et caractéristiques

Balisage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à XML

Conclusions & Perspectives

◆ Le travail des prochains mois

◆ Le futur site Web

Conclusions

- ◆ L<sup>X</sup>i<sup>r</sup> devrait atteindre les objectifs fixés
- ◆ Le projet est amené à évoluer
- ◆ Il sera ouvert aux contributions
- ◆ Il sera utilisé :
  - ❖ À EDP Sciences : conversion vers la DTD Journal Publishing (~ 2000 articles/an);

Quel XML ? Quel  
L<sup>A</sup>T<sub>E</sub>X ?

Stratégies pour la  
conversion  
L<sup>A</sup>T<sub>E</sub>X → XML

De L<sup>A</sup>T<sub>E</sub>X vers  
MathML

L<sup>X</sup>i<sup>r</sup> : une brève  
présentation

L<sup>X</sup>i<sup>r</sup>: cahier des  
charges et  
caractéristiques

Balilage L<sup>A</sup>T<sub>E</sub>X

Du DVI balisé à  
XML

Conclusions &  
Perspectives

◆ Le travail des  
prochains mois

◆ Le futur site  
Web

Conclusions

- ◆ L<sup>X</sup>i<sup>r</sup> devrait atteindre les objectifs fixés
- ◆ Le projet est amené à évoluer
- ◆ Il sera ouvert aux contributions
- ◆ Il sera utilisé :
  - ❖ À EDP Sciences : conversion vers la DTD Journal Publishing (~ 2000 articles/an) ;
  - ❖ Avec cyberthèse : conversion de thèses vers TEI Lite