HTML ⟺ LaTeX

Michel Goossens / CERN

Nanterre le 19 janvier 1995

- LaTeX et HTML, les similitudes.
- La conversion de LaTeX vers HTML.
- Étude détaillée de `LaTeX2HTML`.
- La conversion de HTML vers LaTeX.

Similitudes entre LaTeX et HTML

– balisage générique ;

– structure hiérarchique clairement indiquée ;

– jeu de commandes beaucoup plus riche en LaTeX qu'en HTML : trouver des solutions *ad hoc* en générant des images, ou en guidant le logiciel de traduction.

Similitudes entre LaTeX et HTML(suite)

Description	HTML	LaTeX
Commandes de sectionnement		
niveau 1	`<H1>`	`\chapter` ou `\section`
niveau 2	`<H2>`	`\section` ou `\subsection`
niveau 3	`<H3>`	`\subsection` ou `\subsubsection`
niveau 4	`<H4>`	`\subsubsection` ou `\paragraph`
niveau 5	`<H5>`	`\paragraph` ou `\subparagraph`
niveau 6	`<H5>`	`\subparagraph`
nouveau paragraphe	`<P>`	`\par`

Similitudes entre LaTeX et HTML(suite)

Description	HTML	LaTeX
Les listes		
liste numérotée	`<OL>`	`\begin{enumerate}`
liste non-numérotée	`<UL>`	`\begin{itemize}`
entrée dans la liste	`<LI>`	`\item`
liste de description	`<DL>`	`\begin{description}`
terme	`<DT>`	`\item`
définition	`<DD>`	*text*

Similitudes entre LaTeX et HTML(suite)

Description	**HTML**	**LaTeX**
Mise en valeur du texte		
emphatique	`<EM>`texte`</EM>`	`\emph{`texte`}`
italique	`<I>`texte`</I>`	`\textit{`texte`}`
gras	`<B>`texte`</B>`	`\textbf{`texte`}`
chasse fixe	`<TT>`texte`</TT>`	`\texttt{`texte`}`

La conversion de LaTeX vers HTML

`LaTeX2HTML`	de Nikos Drakos (Leeds, Grande Bretagne) ;
`l2x`	de Henning Schulzrinne (Berlin, Allemagne) : traduction de LaTeX en plusieurs autres formats. `l2x -p html.tcl article.tex` proglèmes pour les tableaux, les expressions mathématiques, les tabulations, etc. ;
`tex2RTF`	de Julian Smart (Edinburgh, Grande Bretagne) : transformation de LaTeX en HTML; ne traite pas les tableaux, les expressions mathématiques, les tabulations, etc. ;
`texihtml`	de Lionel Cons (CERN, Genève) : traduction de sources `texinfo` en HTML.

Le convertiseur `LaTeX2HTML` – généralités

– écrit en `perl` ;

– relativement modulaire (extensions possibles, voir ci-dessous)

– traitement « convenable » des commandes LaTeX standard ;

– division du document en une ou plusieurs composantes (sous-documents) en suivant les instructions de l'utilisateur ;

– l'inclusion à chaque page de pictogrammes (icônes) de navigation contenant des hyperliens avec d'autres parties du document ;

– le traitement correct des mathématiques, des tableaux, des figures ou tout environnement arbitraire en faisant appel à LaTeX qui sauve sa sortie sous forme d'images ;

Le convertiseur `LaTeX2HTML` – généralités (suite)

– la gestion d’un facteur de mise à échelle arbitraire à être appliqué aux figures et tableaux qui peuvent être ainsi inclus dans le document HTML sous forme réduite ou comme « icône » (*thumbnail*) ;

– le support pour les programmes de visualisation orientés images (p. ex. `Mosaic`) ou texte simple (p. ex. `Lynx`) ;

– la définition de nouvelles commandes ou environnements, même s’ils sont placés dans un fichier d’extension externe ;

– la génération de notes de bas de page, d’une table des matières, d’une liste des figures ou des tableaux, d’une bibliographie et d’un index ;

Le convertiseur `LaTeX2HTML` – généralités (suite)

- la traduction des renvois LaTeX en hyperliens qui autorisent non seulement des références aux éléments à l'intérieur d'un même document, mais également à d'autres documents disponibles localement ou sur un site distant ;
- la traduction, si possible, des caractères spéciaux LaTeX en entités ISO-Latin1 ;
- l'extension du langage standard LaTeX avec des commandes pour introduire des hyperliens pointant vers des ressources multimedia résidant sur l'Internet ou des programmes qui doivent être exécutés ;
- la possibilité d'ignorer du texte si une certaine condition est vérifiée ;
- l'utilisation de code HTML « pur » dans le document LaTeX.

Options de la command `latex2html`

```
>latex2html [options] nom_de_fichier.tex
```

version officielle 0.6.4 (août 1994)

```
-split num
-link num
-nolatex
-external_images
-ascii_mode
-t texte-du-titre
-dir répertoire-de-sortie
-address adresse-auteur
```

Options de la command `latex2html` (suite)

`-no_navigation`

`-top_navigation`

`-bottom_navigation`

`-auto_navigation`

`-index_in_navigation`

`-contents_in_navigation`

`-next_page_in_navigation`

`-previous_page_in_navigation`

`-info` *desc*

`-dont_include` *fichier(s)*

`-reuse`

`-init_file` *fichier*

Options de la command `latex2html` (suite)

`-show_section_numbers`

`-h`

version 0.7 α2 (décembre 1994)

`-no_subdir`

`-ps_images`

`-no_reuse`

`-no_images`

`-images_only`

Exemple d'utilisation de LaTeX2HTML

```
\documentclass{article}
\usepackage{makeidx,t1enc}
\usepackage[dvips]{graphics}
\usepackage[french]{babel}
\makeindex
\begin{document}
\begin{center}\Large
  Exemple d'un article en français\\[2mm]\today
\end{center}
\tableofcontents
\listoffigures
\listoftables
\section{Une figure EPS}
\index{section}
Cette section montre comment inclure une figure
PostScript\cite{bib-PS} dans un document \LaTeX.
La figure~\ref{Fpsfig} est insérée dans le texte
à l'aide de la commande
\verb!\includegraphics{colorcir.eps}!.
\index{figure}\index{PostScript}
\begin{figure}
\centering
 \begin{tabular}{c@{\qquad}c}
 \resizebox{3cm}{!}{\includegraphics{colorcir.eps}} &
 \resizebox{3cm}{!}{\includegraphics{tac2dim.eps}}
  \end{tabular}
  \caption{Deux images EPS}\label{Fpsfig}
\end{figure}
```

Exemple d'utilisation de LaTeX2HTML

```
\section{Exemple d'un tableau}
Le tableau~\ref{tab:exa} à la page
\pageref{tab:exa} montre l'utilisation
de l'environnement \texttt{table}.
\begin{table}
\centering
  \begin{tabular}{cccccc}
    \Lcs{primo}   \primo   &
    \Lcs{secundo} \secundo &
    \Lcs{tertio}  \tertio  &
    \Lcs{quatro}  \quatro  &
    2\Lcs{ieme}\ 2\ieme
  \end{tabular}
  \caption{Quelques commandes de l'option
           \texttt{french} de \texttt{babel}}
  \label{tab:exa}\index{tableau}
\end{table}
\begin{thebibliography}{99}
\index{références}
\bibitem{bib-PS}
Adobe Inc. \emph{PostScript, manuel de
                 référence (2ième édition)}
InterÉditions (France), 1992
\end{thebibliography}
\printindex
\index{index}
\end{document}
```

Exemple d'utilisation de LaTeX2HTML

Exemple d'un article en français

7 décembre 1994

Table des matières

Liste des figures

Liste des tableaux

1 Une figure EPS

Cette section montre comment inclure une figure PostScript[1] dans un document LaTeX. La figure 1 est insérée dans le texte à l'aide de la commande `\includegraphics{colorcir.eps}`.

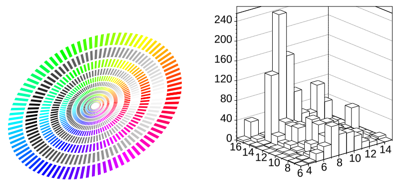


Figure 1: Deux images EPS

2 Exemple d'un tableau

Le tableau 1 à la page 1 montre l'utilisation de l'environnement `table`.

`\primo`	1°	`\secundo`	2°	`\tertio`	3°	`\quatro`	4°	`2\ieme`	2ᵉ

Tableau 1: Quelques commandes de l'option `french` de `babel`

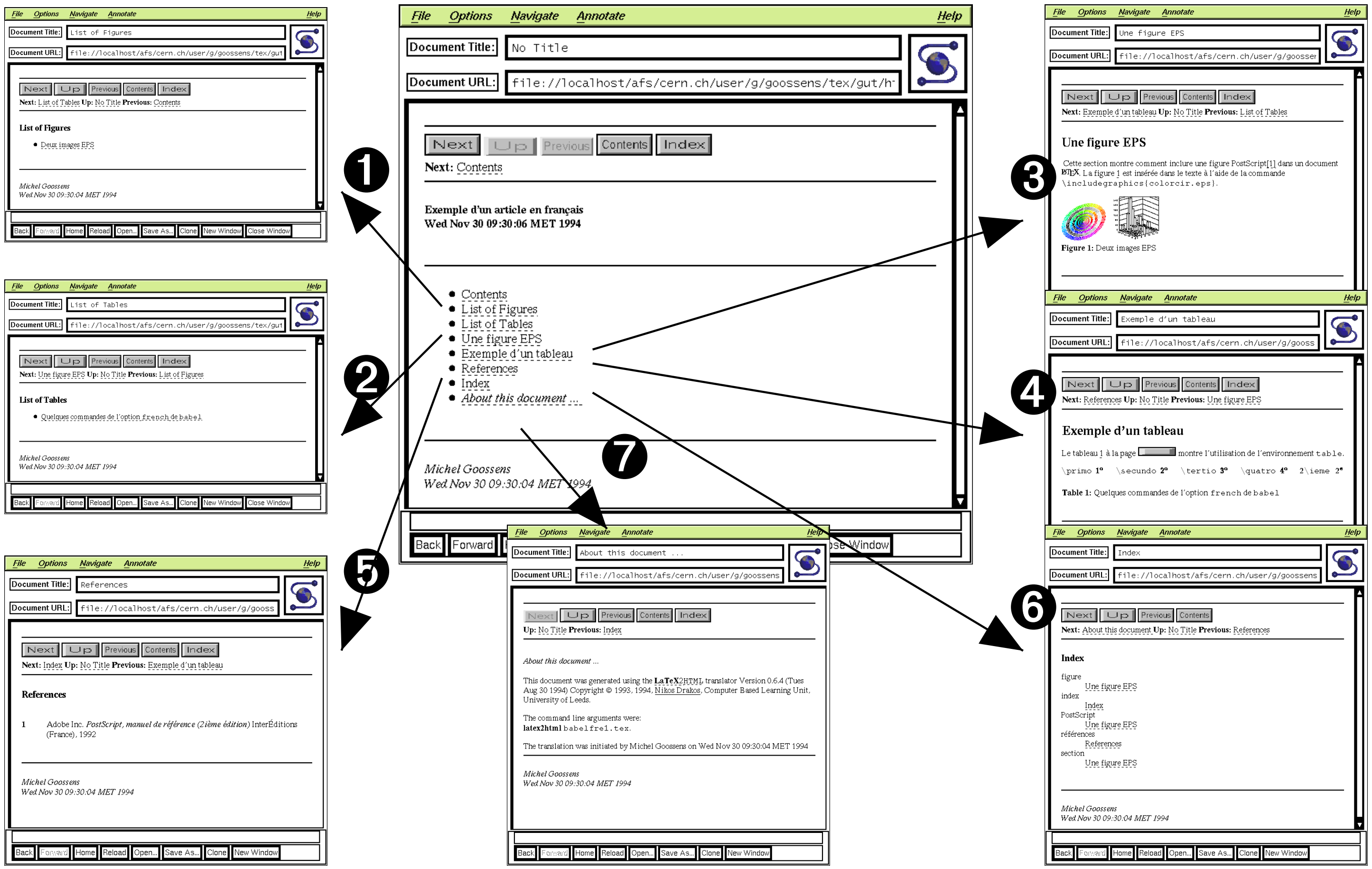
Références

[1] Adobe Inc. *PostScript, manuel de référence (2ième édition)* InterÉditions (France), 1992

Index

1

Exemple d'utilisation de LaTeX2HTML(résultat)



Personnaliser et étendre `LaTeX2HTML`

- les options sur la ligne de commande ;
- le fichier `latextohtml.config` d'initialisation système :
  - définitions communes pour toute une installation ;
  - emplacements des programes et utilitaires (p. ex. LaTeX, `dvips`, `gs`, `pmbplus`) ;
  - initialisation de variables `perl` contrôlant le comportement par défaut de `LaTeX2HTML` ;
  - En fin de fichier il y a la liste des commandes LaTeX à ignorer, et à processées par LaTeX pour inclusion comme fichier image.

Personnaliser et étendre `LaTeX2HTML` (suite)

- `.latex2html-init` permet une initialisation personnelle :
  - se trouve dans le répertoire `$HOME` de l'utilisateur ;
  - peut contenir la même information que `latextohtml.config` ;
  - peut aussi contenir du code `perl` qui spécifie comment certaines commandes LaTeX doivent être traduites en HTML.

La distribution vient avec un fichier type `dot.latex2html-init` qui peut servir comme modèle. On y trouve également `german.perl`, `french.perl`, `html.perl` et `makeidx.perl`.

Personnaliser et étendre LaTeX2HTML (suite)

```
### Comportement par défaut de latex2html ##################################

$MAX_SPLIT_DEPTH = 8;# arrêter génération de fichiers séparés à ce niveau
$MAX_LINK_DEPTH = 4; # ne montrer plus les noeuds plus bas que ce niveau
$NOLATEX = 0;        # si =1 pas appeler LaTeX pour environnement inconnu
$EXTERNAL_IMAGES = 0;# si =1 laisser images à l'extérieur du document
$ASCII_MODE = 0;     # si =1 pas utiliser d'icônes ou d'images internes
$TITLE = $default_title; # la valeur par défaut est "No Title"
$DESTDIR = '.';      # répertoire cible
# Spécifier format si vous n'aimez pas la valeur par défaut <Nom> <Date>
$ADDRESS = "<I>$address_data[0] <BR>\n$address_data[1]</I>";
$NO_NAVIGATION = 0;  # si =1 pas générer un tableau de navigation
$AUTO_NAVIGATION = 1;# mettre tableau de navigation en haut de chaque page
$WORDS_IN_PAGE = 300;# si nb. lignes > $WORDS_IN_PAGE, mettre également
                     # tableau de navigation en bas de cette page
$INDEX_IN_NAVIGATION = 1;# hyperlien vers index dans tableau de navigation
$CONTENTS_IN_NAVIGATION = 1;# hyperlien vers table des matières
$NEXT_PAGE_IN_NAVIGATION = 1;# hyperlien vers document suivant
$PREVIOUS_PAGE_IN_NAVIGATION = 1;# hyperlien vers document précédent
$INFO = 1;           # si =0 pas produire document "About this document..."
```

Un fichier .latex2html-init

```
$ADDRESS = "<I>Michel Goossens<BR>Division CN<BR>Tél. 3363</I>".
           "<BR>\n$address_data[1]</I>";
$MAX_SPLIT_DEPTH = 2; # arrêter séparation au niveau sous-section
$MAX_LINK_DEPTH = 1;  # montrer liens descendants juste pour les sections
$EXTERNAL_IMAGES = 1; # laisser images à l'extérieur du document

$DONT_INCLUDE = "memo:psfig:times:preprint:revtex:eqsecnum:" .
"aps:doublespace:float:harvard:tabls";

$MATH_SCALE_FACTOR = 1.6;# échelle pour les images en ligne
$FIGURE_SCALE_FACTOR = 0;# échelle pour les images mises en évidence
                         # = 0, garder les dimensions originales
```

Un fichier `.latex2html-init` (suite)

```
### Commandes à ignorer ##################################
&ignore_commands( <<_IGNORED_CMDS_);
documentclass # [] # {}
usepackage # [] # {}
linebreak# []
mbox
makebox# [] # []
_IGNORED_CMDS_

### Commandes à passer directement à LaTeX ##############
&process_commands_in_tex (<<_RAW_ARG_CMDS_);
fbox # {}
framebox # [] # [] # {}
_RAW_ARG_CMDS_

1; # Ceci doit être la dernière ligne
```

Étendre la fonctionnalité de `LaTeX2HTML`

```
sub do_cmd_Ucom {
    local($_) = @_;
    s/$next_pair_pr_rx//o;
    join('',qq+<KBD>$&</KBD>+,$_);
}

sub do_env_EnumZW {
    local($_) = @_;
    local($count) = 0;
    s|\\item|do {++$count; qq!<DT><IMG ALIGN=TOP ALT=""
    SRC="http://somewhere/icons/circled$count.xbm"><DD>!}|eog;
  "<DL COMPACT>$_</DL>";
    }
```

Extensions hypertextes de `LaTeX2HTML` – le fichier `html`

Introduction d'hyperliens

\htmladdimg{⟨*URL*⟩}
\htmladdnormallink{*texte*}{⟨*URL*⟩}
\htmladdnormallinkfoot{*texte*}{⟨*URL*⟩}

Renvois par hyperliens à d'autres documents

\externallabels{⟨*URL répertoire externe*⟩}{⟨*fichier des étiquettes*⟩}
\externalref{⟨*étiquette dans fichier externe*⟩}

Notons que ⟨*fichier des étiquettes*⟩ doit exister avant le traitement par `LaTeX2HTML` du fichier qui contient la commande `\externallabels` lui faisant référence. Ceci veut dire que les fichiers doivent être traités dans le bon ordre.

Extensions hypertextes de `LaTeX2HTML`—le fichier `html`

Inclure du code HTML *pur*

```
\begin{rawhtml}
  ...
\end{rawhtml}
```

Le contenu de l’environnement est ignoré par LaTeX.

Code conditionnel

```
\begin{latexonly}
  version papier
\end{latexonly}
```

```
\begin{htmlonly}
  version électronique
\end{htmlonly}
```

Extensions hypertextes de `LaTeX2HTML`—le fichier `html`

Contrôler l'apparence des renvois

```
\hyperref{texte_h}{texte_d1}{texte_d2}{étiquette}
\htmlref{texte}{étiquette}
```

étiquette doit être défini avec une commande `\label`.

Pour `\htmlref` *texte* sera présent dans les deux versions mais dans la version électronique HTML le texte sera activé en devenant un hyperlien renvoyant à *étiquette*.

Extensions hypertextes de LaTeX2HTML—le fichier html

Contrôler la dimension des images

Contrôle globale avec les variables `perl` `MATH_SCALE_FACTOR` et `FIGURE_SCALE_FACTOR`.

Contrôle au niveau de chaque image :

`\htmlimage{`*options*`}`

`scale=`*facteur_d'échelle*

`external`

`thumbnail=`*facteur_de_réduction*

```
\begin{figure}
  \htmlimage{thumbnail=0.24}
  \includegraphics{mondessin.eps}
  \caption{Description de mon dessin}\label{fig-mondessin}
\end{figure}
```

Fichier principal (`ex2.tex`)

```
\documentclass{article}
\usepackage{html,t1enc}
\usepackage[dvips]{graphics}
\usepackage[french]{babel}
\begin{document}
\begin{center}
\Large Exemple d'un document composé
\end{center}

\htmladdnormallink{Les Images}%
           {../ex21/ex21.html}

\externallabels{../ex21}%
     {../ex21/labels.pl}
Référence à une figure
externe~\externalref{Fpsfig}.

\htmladdnormallink{Les tableaux}%
             {../ex22/ex22.html}

\externallabels{../ex22}%
               {../ex22/labels.pl}
Référence à un tableau
externe~\externalref{tab-exa}.

\htmladdnormallink{La bibliographie}%
             {../ex2bib/ex2bib.html}
\end{document}
```

Fichier avec images (ex21.tex)

```
\documentclass{article}
\usepackage{html,t1enc}
\usepackage[dvips]{graphics}
\usepackage[french]{babel}
\makeindex
\begin{document}
\section{Une figure EPS}\label{sec-figure}
Cette section montre comment inclure une
\externallabels{../ex2bib}%
     {../ex2bib/labels.pl}%
figure PostScript\externalref{bibPS}
dans un document \LaTeX. La
\hyperref{figure}{figure }{}{Fpsfig}
est insérée dans le texte à l'aide
de la commande
\verb!\includegraphics{colorcir.eps}!.
\begin{figure}
\htmlimage{thumbnail=0.2}
\centering
 \begin{tabular}{c@{\qquad}c}
  \resizebox{6cm}{!}{%
    \includegraphics{colorcir.eps}} &
  \resizebox{6cm}{!}{%
    \includegraphics{tac2dim.eps}}
 \end{tabular}
 \caption{Deux images EPS}\label{Fpsfig}
\end{figure}
\end{document}
```

Fichier avec tableau (ex22.tex)

```
\documentclass{article}
\usepackage{html,t1enc}
\usepackage[dvips]{graphics}
\usepackage[french]{babel}
\newcommand{\Lcs}[1]{%
    \texttt{\symbol{'134}#1}\enspace}
\begin{document}
\section{Exemple d'un tableau}
\label{sec-tableau}
Le \hyperref{tableau}{tableau }{}{tab-exa}
montre l'utilisation de l'environnement
\texttt{table}.
\begin{table}
\centering
  \begin{tabular}{cccccc}
    \Lcs{primo} \primo     &
    \Lcs{secundo} \secundo &
    \Lcs{tertio} \tertio   &
    \Lcs{quatro} \quatro   &
    2\Lcs{ieme}\ 2\ieme
  \end{tabular}
  \caption{Quelques commandes de l'option
           \texttt{french} de
           \texttt{babel}}\label{tab-exa}
\end{table}
\end{document}
```

Fichier de bibliographie (`ex2bib.tex`)

```
\documentclass{article}
\usepackage{html,t1enc}
\usepackage[dvips]{graphics}
\usepackage[french]{babel}
\makeindex
\begin{document}
\begin{thebibliography}{99}
\bibitem{bib-PS}\label{bibPS}
Adobe Inc. \emph{PostScript, manuel de
référence (2ième édition)}
InterÉditions (France), 1992
\end{thebibliography}
\end{document}
```

Exemple d'un document composé

```
répertoire principal      569  ex20.tex
====================      721  ex21.tex
                          627  ex22.tex
                          322  ex2bib.tex

    sous-répertoire ex20                    sous-répertoire ex2bib
    --------------------                    ----------------------
    1187  ex20.html                          844  ex2bib.html
     109  images.pl                          109  images.pl
      93  labels.pl                          141  labels.pl

    sous-répertoire ex21                    sous-répertoire ex22
    --------------------                    --------------------
    1755  T_18854_figure15.gif               624  _15561_table12.gif
   12118  _18854_figure15.gif               1047  ex22.html
     122  _18854_tex2html_wrap57.gif         512  images.pl
    1345  ex21.html                          687  images.tex
     539  images.pl                          191  labels.pl
     589  images.tex
     190  labels.pl
```

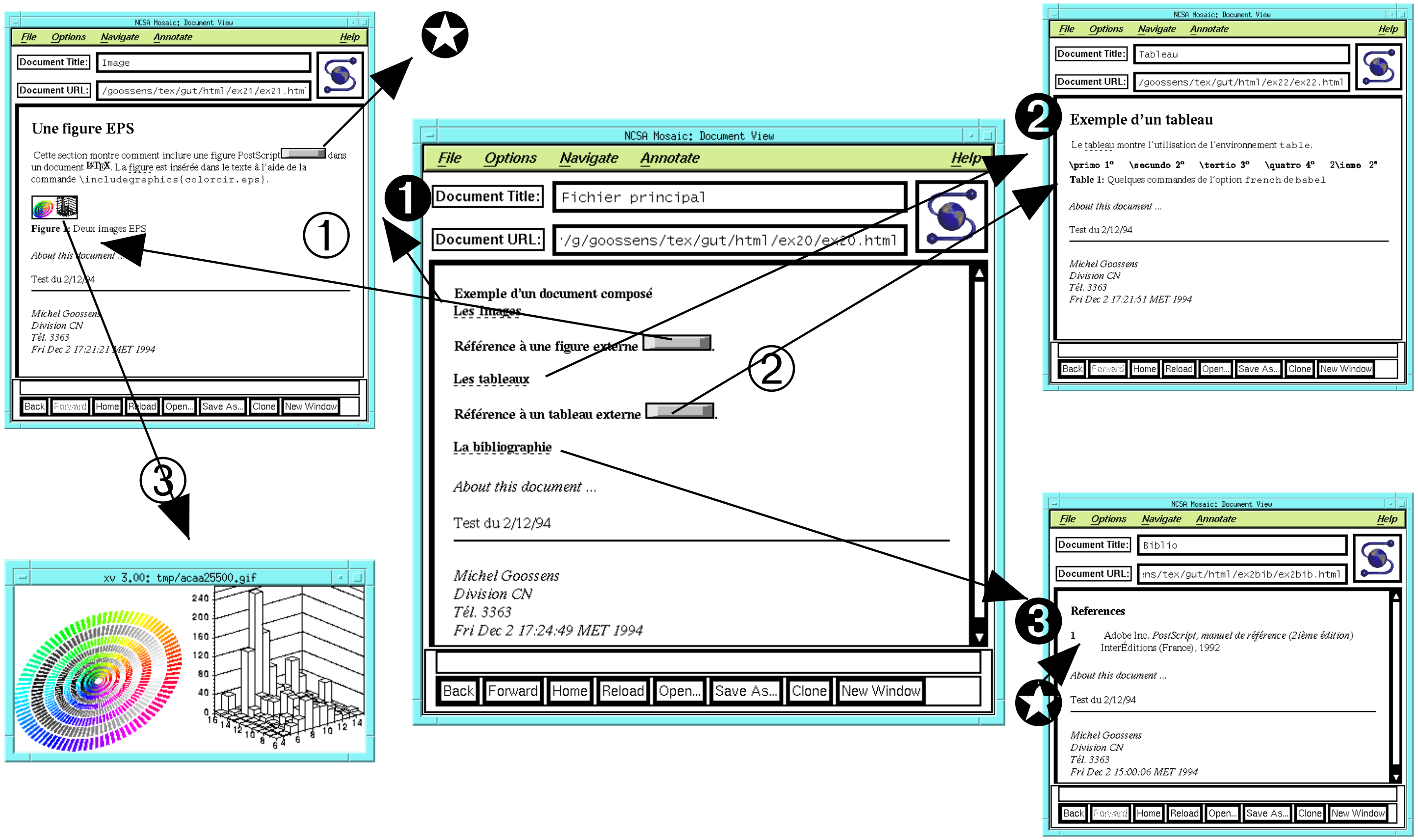
Exemple d'un document composé (suite)

```
# Fichier moninit.pl
# Personnalisation de latex2html
$ADDRESS = "<I>Michel Goossens<BR>Division CN<BR>Tél. 3363</I>".
           "<BR>\n$address_data[1]</I>";
$MAX_SPLIT_DEPTH = 0; # pas séparer en sous-documents
$MAX_LINK_DEPTH = 0;  # pas montrer liens descendants
$NO_NAVIGATION = 1;   # pas générer un tableau de navigation

1; # Ceci doit être la dernière ligne

> latex2html -init_file moninit.pl -t "Image" \
             -info "Test du 2/12/94" ex21.tex
```

Exemple d'un document composé (suite)

```
This is LaTeX2HTML Version 0.6.4 (Tues Aug 30 1994)...
Computer Based Learning Unit, University of Leeds.

OPENING /afs/cern.ch/user/g/goossens/ex21.tex

Loading /usr/local/lib/latex2html/styles/html.perl...

Reading ...
Reading ex21.aux ..........................
Translating ...0/2..........1/2........2/2......
Generating images using latex ...
This is TeX, Version 3.1415 (C version 6.1)
(18854_images.tex
LaTeX2e <1994/06/01> patch level 3
Hyphenation patterns for english, loaded.

Generating postscript images using dvips ...
This is dvipsk 5.58c Copyright 1986, 1994 Radical...
' TeX output 1994.12.02:1830' -> 18854_image
(-> 18854_image001) <tex.pro><special.pro>[1]
(-> 18854_image002) <tex.pro>
<special.pro>[2<colorcir.eps><tac2dim.eps>]
GS>GS>Writing 18854_image001.ppm
GS>Writing _18854_tex2html_wrap57.gif
GS>GS>Writing 18854_image002.ppm
GS>Writing _18854_figure15.gif
GS>GS>Writing 18854_image002.ppm
GS>Writing T_18854_figure15.gif

Doing section links ....
Done.
```

Exemple d'un document composé (résultat)



La conversion de HTML vers LaTeX

- les utilitaire de visualisation offrent souvent des sorties PostScript ou Ascii (p. ex. Mosaic) ;
- parfois un veut un texte plus structuré pour l'imprimer « correctement » ;
- `HTML2LaTeX` traduit (une grande partie de) HTML vers LaTeX ;
- `SGML2TeX` est plus général et transforme un source SGML arbitraire en LaTeX.

HTML2LaTeX, un convertiseur HTML vers LaTeX

- programme écrit en C par Nathan Torkington (Nouvelle Zélande) ;
- partie traduction basée sur analyseur (parser) HTML de Mosaic du NCSA ;
- chaque balise HTML est traduite dans la balise correspondante LaTeX ;
- traduction pas idéale (p. ex. `<PRE>` pas traité correctement).

```
> html2latex -n - < file.html | more
```

L'option `-n` signifie que les sections seront numérotées.

```
> html2latex -t 'Mes premiers pas en HTML' \
             -a 'Prénom Nom' -p            \
             -c -o'[12pt,twoside]{article}' mon-article
```

`SGML2TeX`, un convertiseur SGML en LaTeX

- écrit par Peter Flynn (Cork, Irlande) ;
- convertit des balises SGML en instructions TeX ;
- actuellement est écrit en PCL pour MS-DOS sur PC ;
- sera (probablemt) réécrit dans un langage plus portable ;
- pas de vérification du source SGML ;
- l'utilisateur définit l'équivalent LaTeX des éléments, attributs ou entités ;
- un fichier de configuration (prédéfinitions) ;
- par défaut les éléments sont mis dans une forme acceptable par LaTeX :
  - les balises deviennent `\start...` et `\finish...` ;
  - les entités `&ent;` deviennent `\ent{}` ;
  - les attributs deviennent des argument LaTeX.