

NF92

Traitement automatique de
l'information

Pavol BARGER

Publication scientifique

Pavol BARGER

Contenu

- Texte
- Document structuré
- LaTeX

Représentation du texte

- Chaque caractère est représenté comme un numéro -> codage
- Plusieurs codes disponibles
 - morse
 - ASCII (7 ou 8 bits)
 - Unicode (16 bits, usage multilingue)
 - EBCDIC (8 bits)

ASCII

- **Américan Standard Code for Information Interchange**
- Le code ASCII de base représentait les caractères sur 7 bits (c'est-à-dire 128 caractères possibles, de 0 à 127)
 - 0 à 31 ne sont pas des caractères -> *caractères de contrôle* (actions)
 - retour à la ligne (CR)
 - Bip sonore (BEL)
 - 65 à 90 - majuscules
 - 97 à 122 - minuscules
- (Il suffit de modifier le 6ème bit pour passer de majuscules à minuscules, c'est-à-dire ajouter 32 au code ASCII en base décimale.)

ASCII

Dec	Hex	Oct	Char	Dec	Hex	Oct	Char	Dec	Hex	Oct	Char	Dec	Hex	Oct	Char
0	0	000	NUL (null)	32	20	040	#32: Space	64	40	100	#64: @	96	60	140	#96: `
1	1	001	SOH (start of heading)	33	21	041	#33: !	65	41	101	#65: A	97	61	141	#97: a
2	2	002	STX (start of text)	34	22	042	#34: "	66	42	102	#66: B	98	62	142	#98: b
3	3	003	ETX (end of text)	35	23	043	#35: #	67	43	103	#67: C	99	63	143	#99: c
4	4	004	EOF (end of transmission)	36	24	044	#36: \$	68	44	104	#68: D	100	64	144	#100: d
5	5	005	ENQ (enquiry)	37	25	045	#37: %	69	45	105	#69: E	101	65	145	#101: e
6	6	006	ACK (acknowledge)	38	26	046	#38: &	70	46	106	#70: F	102	66	146	#102: f
7	7	007	BEL (bell)	39	27	047	#39: '	71	47	107	#71: G	103	67	147	#103: g
8	8	010	BS (backspace)	40	28	050	#40: (72	48	110	#72: H	104	68	150	#104: h
9	9	011	TAB (horizontal tab)	41	29	051	#41:)	73	49	111	#73: I	105	69	151	#105: i
10	A	012	LF (NL line feed, new line)	42	2A	052	#42: *	74	4A	112	#74: J	106	6A	152	#106: j
11	B	013	VT (vertical tab)	43	2B	053	#43: +	75	4B	113	#75: K	107	6B	153	#107: k
12	C	014	FF (FF form feed, new page)	44	2C	054	#44: ,	76	4C	114	#76: L	108	6C	154	#108: l
13	D	015	CR (carriage return)	45	2D	055	#45: -	77	4D	115	#77: M	109	6D	155	#109: m
14	E	016	SO (shift out)	46	2E	056	#46: .	78	4E	116	#78: N	110	6E	156	#110: n
15	F	017	SI (shift in)	47	2F	057	#47: /	79	4F	117	#79: O	111	6F	157	#111: o
16	10	020	DLE (data link escape)	48	30	060	#48: 0	80	50	120	#80: P	112	70	160	#112: p
17	11	021	DC1 (device control 1)	49	31	061	#49: 1	81	51	121	#81: Q	113	71	161	#113: q
18	12	022	DC2 (device control 2)	50	32	062	#50: 2	82	52	122	#82: R	114	72	162	#114: r
19	13	023	DC3 (device control 3)	51	33	063	#51: 3	83	53	123	#83: S	115	73	163	#115: s
20	14	024	DC4 (device control 4)	52	34	064	#52: 4	84	54	124	#84: T	116	74	164	#116: t
21	15	025	NAK (negative acknowledge)	53	35	065	#53: 5	85	55	125	#85: U	117	75	165	#117: u
22	16	026	STB (synchronous idle)	54	36	066	#54: 6	86	56	126	#86: V	118	76	166	#118: v
23	17	027	ETB (end of trans. block)	55	37	067	#55: 7	87	57	127	#87: W	119	77	167	#119: w
24	18	030	CAU (cancel)	56	38	070	#56: 8	88	58	130	#88: X	120	78	170	#120: x
25	19	031	EM (end of medium)	57	39	071	#57: 9	89	59	131	#89: Y	121	79	171	#121: y
26	1A	032	SUB (substitute)	58	3A	072	#58: :	90	5A	132	#90: Z	122	7A	172	#122: z
27	1B	033	ESC (escape)	59	3B	073	#59: ;	91	5B	133	#91: [123	7B	173	#123: {
28	1C	034	FS (file separator)	60	3C	074	#60: <	92	5C	134	#92: \	124	7C	174	#124:
29	1D	035	GS (group separator)	61	3D	075	#61: =	93	5D	135	#93:]	125	7D	175	#125: }
30	1E	036	RS (record separator)	62	3E	076	#62: >	94	5E	136	#94: ^	126	7E	176	#126: ~
31	1F	037	US (unit separator)	63	3F	077	#63: ?	95	5F	137	#95: _	127	7F	177	#127: DEL

Source: www.LaTeXTables.com

Texte

■ Enchaînement de caractères ?

Ex. :

La valorisation à l'UTC Les chercheurs, les enseignants, les personnels, les étudiants de l'UTC donnent un sens à l'innovation ! La valorisation est une des trois missions stratégiques de l'UTC au même titre que la formation et la recherche. L'UTC offre aux entreprises toute une gamme de services : des services liés à la gestion de l'innovation, une pépinière d'entreprises et un service d'aide au montage de projets européens.

-> Besoin d'une mise en forme

Deux types de renseignements

Le contenu est associé à :

- Structure physique
 - Apparence, emplacement, mise en forme
- Structure logique
 - Organisation du document
 - Hiérarchie

Le tout forme un seul texte logique !

Problème majeur

- Mélanger
 - Contenu
 - Mise en forme
 - Structure logique
 - Multimédia



Problème majeur

- Contenu
 - texte
 - images



Problème majeur

- Mise en forme
 - majuscules
 - gras, grand
 - gras
 - italique
 - double colonne
 - majuscules



Problème majeur

- Structure logique
 - titre article
 - auteurs
 - adresse
 - abstract
 - mots clés
 - titre chapitre
 - texte chapitre



Problème majeur

- Solutions
 - WYSIWYG
 - Word
 - Publisher
 - OpenOffice
 - Staroffice
 - Rédaction séparée
 - texte
 - structure logique
 - mise en forme

**Document
structuré**

Document structuré

Plan

- Balisage (markup)
- Latex
- XML
- Parcours de lecture
 - Hypertexte
 - Structure 'site'

Structure d'un document

- Rédaction séparée
 - texte
 - structure logique
 - mise en forme
- 3 fichiers vs 1 seul fichier
- 1 seul fichier
 - texte
 - balises de mise en forme et/ou structure

Balise

Il était un petit homme.

Balise

Il était un `\begin{small}` petit homme `\end{small}`.

Il était un petit homme.

ex. balise commence par \
Problème : comment écrire \ dans le texte ?

Balises

- Des marques placées dans le texte
- Besoin de les distinguer du texte lui-même
- Utilisation de caractères spéciaux
 - Ex: <, %, \$, /, \, _
 - Ces caractères indiquent le début d'une commande

Balises

- Formes de balises
 - `\begin{bal}` ... `\end{bal}`
 - `\blnsrsid9261\charrsid92`
 - `<body>` ... `<\body>`
 - `.bal{...}`
 - `<? ... >`
 - `% ...`, fonction ...
 - `^$?| ... #~&@µ`

LaTeX
RTF
HTML
CSS
PHP
Scilab

Format propriétaire PB

Balises - Règles

- Une balise a une forme fixe, non modifiable (ex. : pas de commentaires dans une balise)
- Deux balises (ex. : `\begin` et `\end`) peuvent former un bloc
- Emboîtement correct, non intercalé
- Avantage : fichier source lisible par l'homme

Structuration logique

- Permet un meilleur repérage dans le document

```
\begin{document}
\title{Histoire de Cali
Mero}
\author{Idefix}
\end{document}
```

Questions :

Tous les documents dont Idefix est l'auteur?

Tous les documents sur Idefix?

LATEX

Présentation

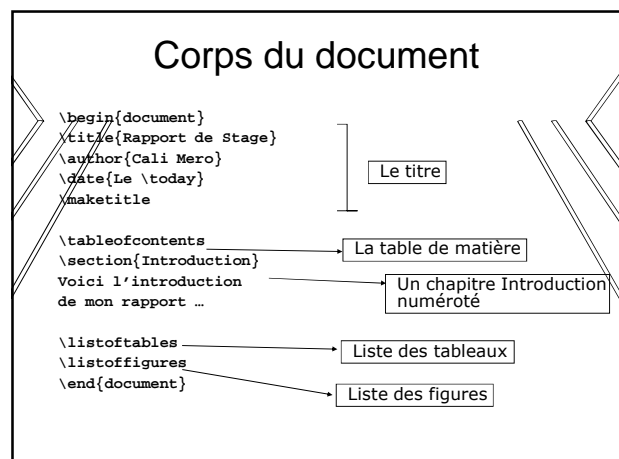
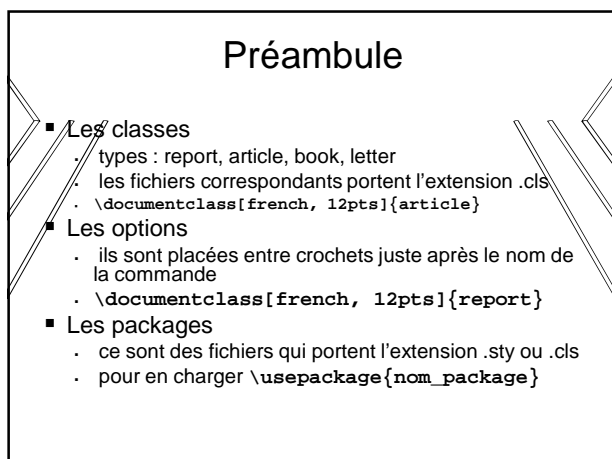
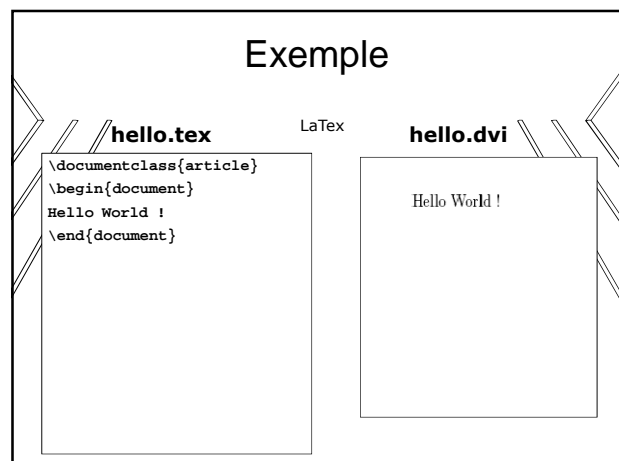
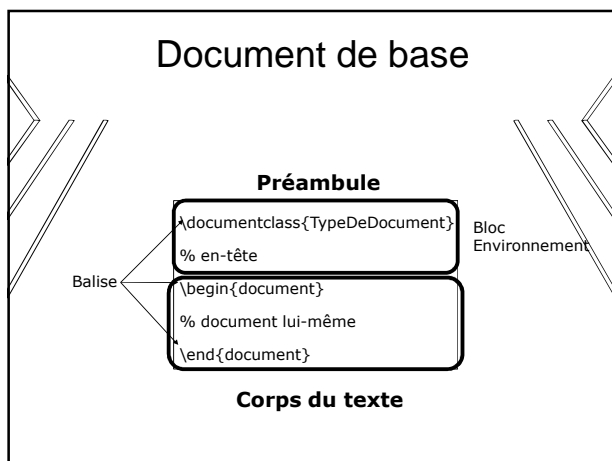
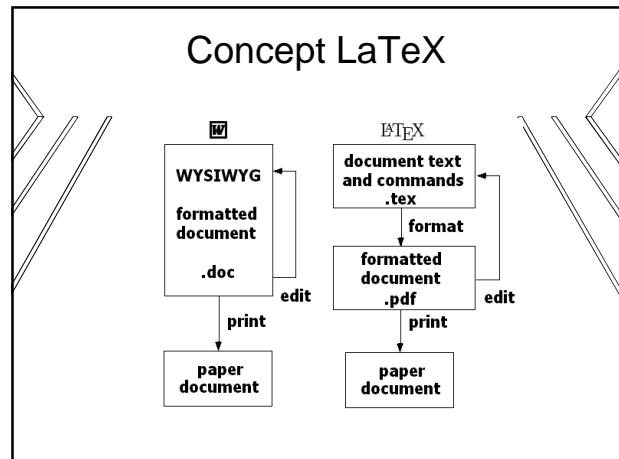
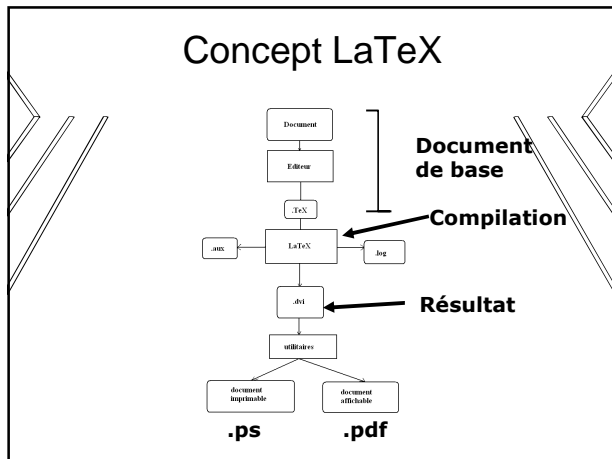
- TeX en 1978 (Donald E. Knuth)
- LaTeX – sur-ensemble de TeX (Leslie Lamport)
 - Rendre TeX plus simple
- Séparation de la forme et du contenu
- Système de préparation de documents
 - Non Wysiwyg
 - Besoin de compilation
 - Non temps réel
 - Gratuit

LaTeX

- Système de rédaction de texte à l'aide des balises
- Accent sur la structure et l'organisation du document
 - Cohérence de la présentation graphique
- Gestion automatique de mise en page
- Adapté à la rédaction scientifique

Avantages

- La typographie est un art...
- Meilleure structuration du document en se concentrant sur le fond et non sur la forme
- Nombreuses extensions
- Portable et gratuit
- Pas de problème de compatibilité et de portabilité de fichiers



Corps du document

```
\begin{document}
\title{Rapport de Stage}
\author{Cali Mero}
\date{Le \today}
\maketitle
```

<code>\part</code>
<code>\chapter</code>
<code>\section</code>
<code>\subsection</code>
<code>\subsubsection</code>
<code>\paragraph</code>
<code>\subparagraph</code>

Commandes de sectionnement

```
\tableofcontents
\section{Introduction}
Voici l'introduction
de mon rapport ...
```

```
\listoftables
\listoffigures
\end{document}
```

Corps du document

```
\begin{document}
\title{Rapport de Stage}
\author{Cali Mero}
\date{Le \today}
\maketitle
```

Rapport de Stage
Cali Mero
Le September 18, 2006

```
\tableofcontents
\section{Introduction}
Voici l'introduction
de mon rapport ...
```

```
\listoftables
\listoffigures
\end{document}
```

Contents

- 1 Introduction
- Voici l'introduction de mon rapport
- List of Tables
- List of Figures

Avec \documentclass{article}

Sectionnement

Chaque commande de sectionnement fait :

- Numérotation hiérarchique de titres
- Insertion de titres dans la table des matières
- Formatage de titres

Forme étoilée (`\section*{...}`) supprime la numérotation

Mise en forme

Taille de caractères

- `\begin{small} petit \end{small}`
- `\begin{large} grand \end{large}`
- `\begin{LARGE} GRAND \end{LARGE}`

- autres: tiny, scriptsize, footnotesize, small, normalsize, large, Large, LARGE, huge

petit
grand
GRAND

Mise en forme

Styles

[illegible]

Listes

```

// Itemized List
\begin{itemize}
  \item First item
  \item Second item
  \item Third item
\end{itemize}

```

```
Enumerated List —————
\begin{enumerate}
  \item First item
  \item Second item
  \item Third item
\end{enumerate}
```

```
Descriptive List (items in bold)
\begin{description}
  \item[First item] Description.
  \item[Second item] Description.
  \item[Third item] Description.
\end{description}
```

Itemized List

- First item
- Second item
- Third item

- Enumerated List

- (1) First item
- (2) Second item
- (3) Third item

Descriptive List (items in bold)

First item: Description.
Second item: Description.
Third item: Description.

Tables

- Environnement : tabular
- Les colonnes sont séparées entre elles par &
- On écrit ligne par ligne
- 3 types de colonnes :
 - l : colonne alignée à gauche (left)
 - r : colonne alignée à droite (right)
 - c : colonne justifiée (center)

```
\begin{tabular}{lrc}
Premier mot & Deuxième mot & Troisième mot \\
Première phrase & Deuxième phrase & Troisième phrase \\
\end{tabular}
```

Premier mot	Deuxième mot	Troisième mot
Première phrase	Deuxième phrase	Troisième phrase

Tables

- Une barre | entre deux descripteurs de colonne insérera un filet entre deux colonnes
- Un \hline entre deux lignes les séparera par un filet

```
\begin{tabular}{|l|l|l|}
\hline
Premier mot & Deuxième mot & Troisième mot \\
\hline
Première phrase & Deuxième phrase & Troisième phrase \\
\hline
\end{tabular}
```

Premier mot	Deuxième mot	Troisième mot
Première phrase	Deuxième phrase	Troisième phrase

Formules mathématiques

- Environnements spécifiques
 - \$... \$ remarque -> dans le texte
 - \[... \] remarque -> mode formule en dessous du texte
 - \begin{math} ... \end{math}
 - \begin{equation} ... \end{equation}
 - \begin{displaymath} ... \end{displaymath}
- Exemple lettre grecques
 - \$\lambda, \xi, \pi, \mu, \Phi, \Omega\$

$$\lambda, \xi, \pi, \mu, \Phi, \Omega$$

Exemples

- \sqrt{x} $\sqrt{x^2 + \sqrt{y}}$ $\sqrt[3]{2}$
- $\overline{m+n}$ or $\underbrace{a+b+\dots+z}_{26}$

$$\overline{m+n} \text{ or } \underbrace{a+b+\dots+z}_{26}$$

Indices et exposants

- a_{-1} x^2 $e^{-\alpha t}$ a^{3}_{ij}

$$a_1 \quad x^2 \quad e^{-\alpha t} \quad a_{ij}^3$$

Exercice

$$\text{corr}(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{[\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2]^{1/2}}$$

```
\begin{equation}
\text{corr}(X,Y)=\frac{\sum_{i=1}^n (x_i-\bar{x})(y_i-\bar{y})}{[\sum_{i=1}^n (x_i-\bar{x})^2 \sum_{i=1}^n (y_i-\bar{y})^2]^{1/2}}
\end{equation}
```

Matrices

```
\left[ \left( \begin{array}{cccccc}
1 & 2 & 3 & \cdots & n \\
2 & 3 & 4 & \cdots & n+1 \\
3 & 4 & 5 & \cdots & n+2 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
n & n+1 & n+2 & \cdots & 2n-1
\end{array} \right) \right]
```

$$\left(\begin{array}{cccccc} 1 & 2 & 3 & \cdots & n \\ 2 & 3 & 4 & \cdots & n+1 \\ 3 & 4 & 5 & \cdots & n+2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n & n+1 & n+2 & \cdots & 2n-1 \end{array} \right)$$

Figures

Environnement figure
 Besoin d'utiliser un package (`\usepackage{graphicx}`)
 Concept d'un flottant
 Ex :
`\begin{figure}[options]`
`\includegraphics[options]{figure_name.eps}`
`\caption{Figure Caption would go here}`
`\label{fig_label}`
`\end{figure}`

Figure Options h – here, t – top, b – bottom, p – separate page

Environnement verbatim

```
\begin{verbatim}
a b
c
d e eee e
f
\end{verbatim}
```

Copie exactement le texte et garde son formatage

```
a b
c
d e eee e
f
```

verbatiminput

```
\usepackage{verbatim}
.
.
.
\verbatiminput{fichier.txt}
```

Permet de inclure le listing du fichier dans un document.

```
a b
c
d e eee e
f
```

Commandes diverses

- `\newpage`
- `\tableofcontents`
- `\listoffigures`
- `\listoftables`
- `\makeindex`
- `\LaTeX`
- `\c[c], \e`
- `\layout, \textwidth`
- `\newfont`
- `\hspace`
- `\documentclass[titlepage, a4paper, twocolumn]{article}`

Hypertexte

```
\documentclass[12pt]{amsart}
\begin{document}
\tableofcontents
\section{Introduction}
\label{intro}
Introductory text would be here.
\section{Body}
blah, blah, blah.
Remember what I said in
Section \ref{intro}
There is more in \ref{extra}
\appendix
\section{Extra Stuff}
\label{extra}
This is the appendix.
\end{document}
```

CONTENTS	
1. Introduction	1
2. Body	1
Appendix A. Extra Stuff	1
1. INTRODUCTION	
Introductory text would be here.	
2. BODY	
blah-blah-blah	
Remember what I said in Section 1?	
There is more in A	
APPENDIX A. EXTRA STUFF	
This is the appendix.	

Latex libre

- Pour Linux
 - Tex/Latex
 - Emacs
 - Lyx
- Pour Windows :
 - MikTeX
 - Texnic Center
 - Texmaker
 - Lyx