M1 MSc DataScience - Methods and tools for technical and scientific writing

Structured documents, Markdown and Latex

Semestre 1, 2021-2021

Lecturer: Marco Winckler (UCA, I3S)
winckler@unice.fr

UNIVERSITÉ CÔTE D'AZUR

Structured documents/markup languages

The origin of structured documents/markup languages: SGML

- SGML : Standard Generalized Markup Language
  - ISO 8879 – 1989 (started with GML in the 60's)
  - Based on tags to represent commands on documents
  - Readable (yet hardly) by humans
  - Documents should comply to a DTD: document Type Definition
  - Include the notion of document validity
- SGML was hard to use, ex. programming printer machines… replaced by other markup languages

Markup languages

- Goals:
  - Separation of concerns about the document: structure, content, presentation, links, etc.
  - Process documents without taking into account the presentation and the contents
  - Check the validity of documents (ex. contain all required elements)
  - Create documents cross-platforms
- Examples of markup languages:
  - HTML, XML, RTF, Mark down, Latex…

Example of Markup language HTML

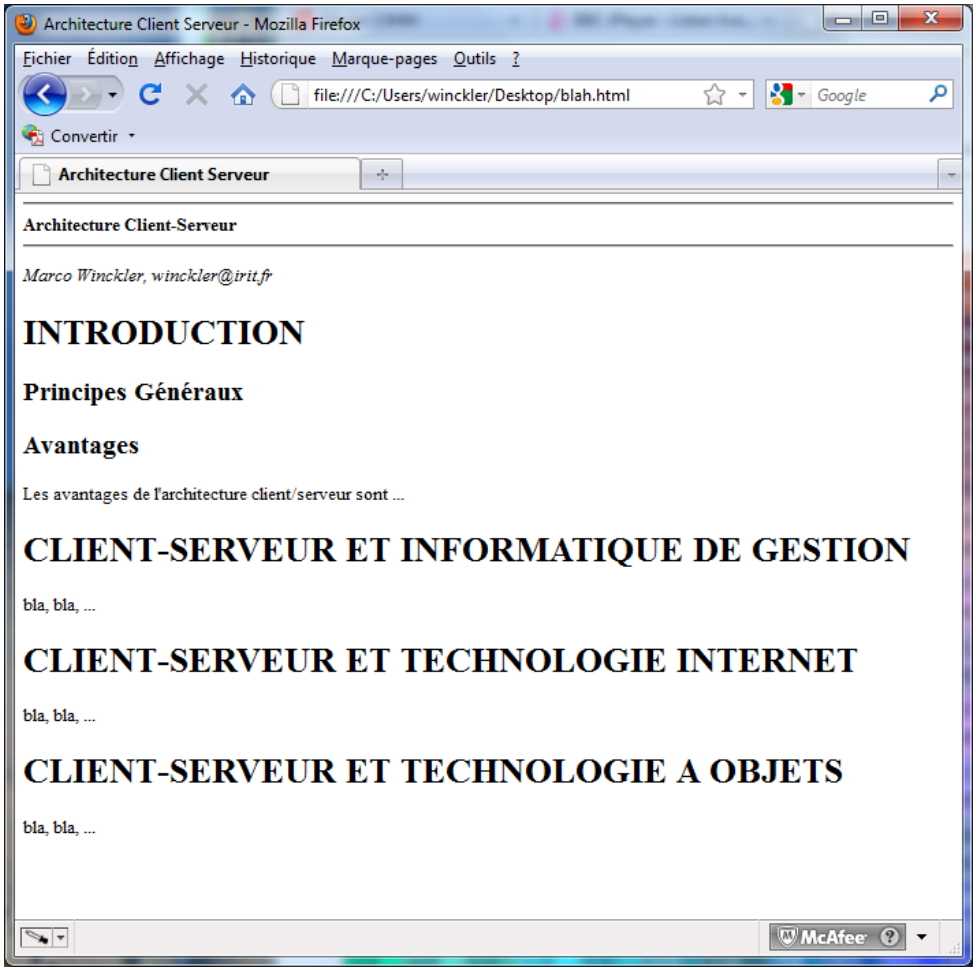
Structure of a HTML document

- Tags: imply a command
  - contents : creates a paragraph
  - **Bold/Grass** : deprecated tag for style bold
  - **paragraph in bold** : notice the combined tags
  - </br> : auto closing tag used for break a line in a paragraph
- Attributes : add value to moderate a command
  - <tag attribute="style:bold"> Content </tag>
- Entities: commands to replace special characters (used in other commands or not standard in all platforms)
  - "<" represents the < sign.
  - ">" represents the > sign.
  - "" represents the " mark.
  - å (in decimal) represents the letter "a" with a small circle above it.
  - И (in decimal) represents the Cyrillic capital letter "I".
  - 水 (in hexadecimal) represents to the Chinese character for water.

Example

```
<html>
<HEAD>
        <TITLE>
                Architecture Client Serveur
        </TITLE>
</HEAD>
<BODY>
        <HR>
        <STRONG>
        Architecture Client-Serveur
        </STRONG>
        <HR>
        <P>
        <ADDRESS>
        Marco Winckler, winckler@irit.fr
        </ADDRESS>
        <H1>INTRODUCTION</H1>
                <H2>Principes Généraux</H2>
                <H2>Avantages</H2>
                Les avantages de l'architecture client/serveur sont ...<P>
        <H1>CLIENT-SERVEUR ET INFORMATIQUE DE GESTION</H1>
        bla, bla, ...
        <H1>CLIENT-SERVEUR ET TECHNOLOGIE INTERNET</H1>
        bla, bla, ...
        <H1>CLIENT-SERVEUR ET TECHNOLOGIE A OBJETS</H1>
        bla, bla, ...
</BODY>
</html>
```

Display on a Web browser



Hierarchy of tags in HTML documents

```
<html>

    <head>
        <title>
            En-Tête du document dans la barre de titre…
        </title>
        <meta name="Author"  content="Winckler">
        <base href="/usr/home/winckler"> </base>
    </head>

    <body bgcolor="#C0C0C0" text="#000000">
        ... Le contenu du document...
                <h1> Titre de niveau 1 </h1>
                <h2> Titre de niveau 2 </h2>
                <p> Paragraphe est indiquée par </p>
    </body>

</html>
```

Markdown

- Simplified way to write content for the web
- Simplified version of markup language
- Ex.

  italics
- Used to share structure of documents with others

Exercise 1

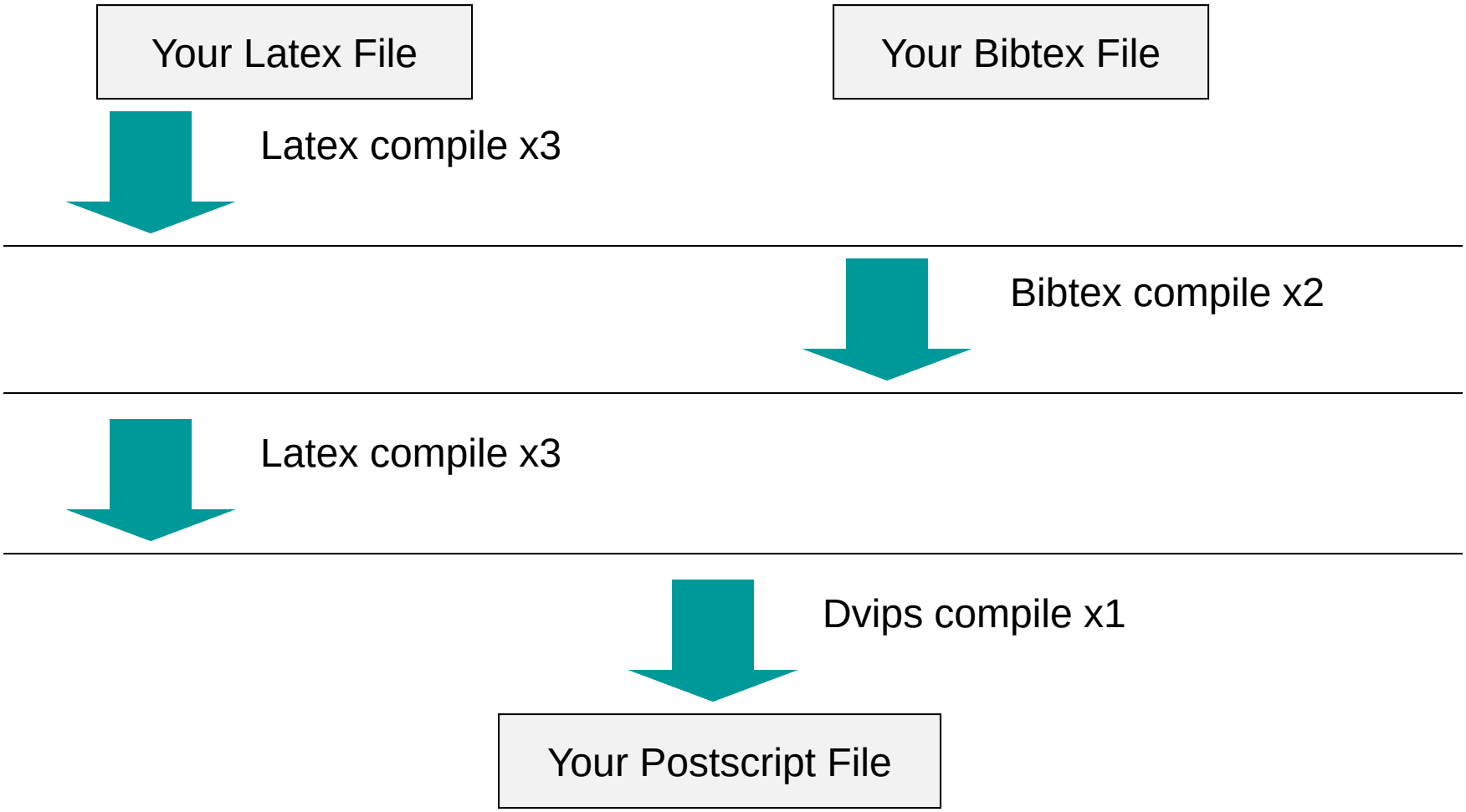
- Follow the tutorial about Markdown at:

  https://www.markdowntutorial.com/

Introduction to Latex

- Essentially a Markup Language (like HTML, XML and RTF)
- An extension to TeX system
- TeX written by Donald Knuth in 70´s

Creating Latex Files



Latex File Structure

- Document Class

  Predefined Formats (article, report, book,..).

- Packages used

  Added Functionality (graphics, reference style,...).

- Main Body

  Text and Bibliography References.

The Basics

- Document Class

  `\documentclass[`*`options`*`]{class}`

  options = a4paper, 11pt, 12pt, 10pt, twocolumn, landscape,...

  class = article, report, book,...

- Packages

  `\usepackage{package name}`

  `epsfig = insert PS pictures into the document`

  `fancyhdr = easy definition of footer and header`

Body of Text

- Start with `\begin{document}`
- End with `\end{document}`
- Typesetting Text
  - `\\` or `\newline` and `\newpage`
  - Quotations
  - Bold `\textbf{……}` or `\bf`
  - Italics `\emph{……}` or `\textit{……}` or `\it`
  - Underline `\underline{……}` or `\ul`

Body of Text cont…

- Including Multiple Files
  - `\input{filename.tex}`

Format

- Sections
  - `\section{…}` = 1. Latex is Great
  - `\subsection{…}` = 1.1 Why Latex is Great
  - `\subsubsection{…}` = 1.1.1 Reason One
  - `\appendix` - changes numbering scheme
  - `\chapter{…}` - To be used with book and report document classes
- Titles, Authors and others
  - `\title{…}` `\author{…}`
  - `\footnote{…}`

Format Contd.

- `\maketitle` - Display Title and Author
- `\tableofcontents` - generates TOC
- `\listoftables` - generates LOT
- `\listoffigures` - generates LOF
- Labels
  - `\label{`*`marker`*`}` - Marker in document.
  - `\pageref{`*`marker`*`}` - Displays page no. of marker.
  - `\ref{`*`marker`*`}` - Displays section location of marker.
- Itemise
  - Use either *enumerate*, *itemize* or *description.*
  - *see handout for example.*

Lists

- Source
  - `\begin{itemize}`
  - `\item Apple`
  - `\item Orange`
  - `\end{itemize}`
- Result
  - `Apple`
  - `Orange`

Lists

- `Enumerate` instead of `itemize` gives a numbered list
- Lists can be recursive

Environment

- Something between
  - `\begin{name}`
  - `\end{name}`
- Many command, for example `\bf` affect the text until the end of environment
- Can be recursive
- Examples:
  - `itemize, center, abstract`

Group

- Text between { and }
- Many commands work until the end of the group
- Code
  - put {one word \bf in bold} here
- Result
  - put one word **in bold** here

Alignment

- Environments `center, flushleft, flushright`
- Example
  - `\begin{flushright}`
  - `Right aligned`
  - `\end{flushright}`
- Result

Right aligned

Font size

\tiny \scriptsize \footnotesize

\small \normalsize

\large \Large

\LARGE \huge

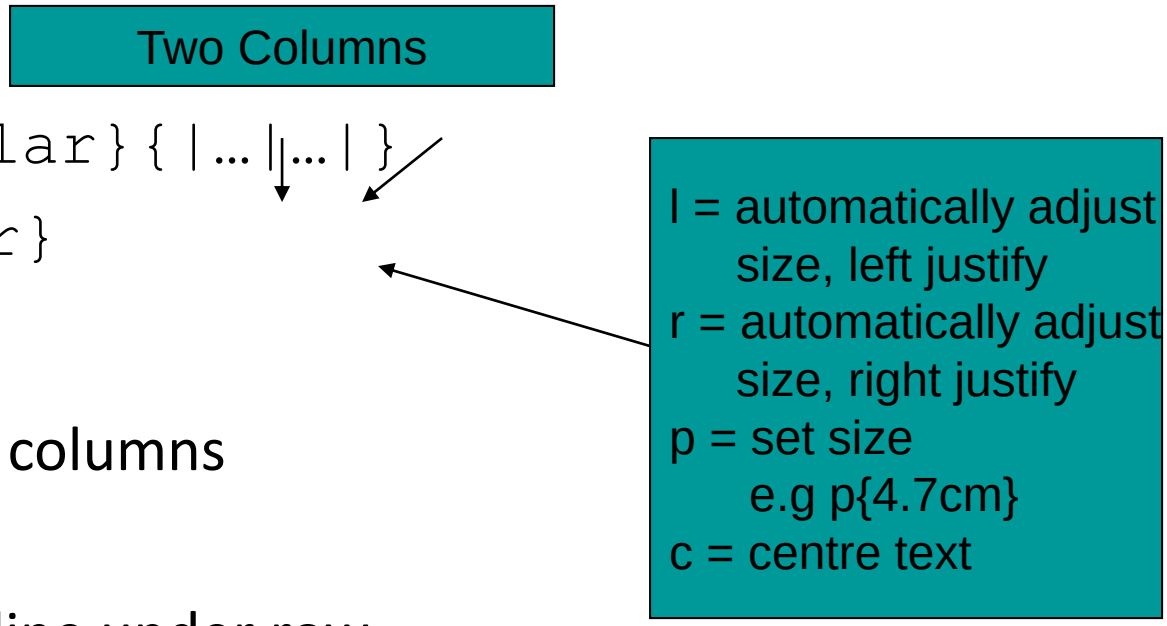
\Huge

Example of Latex document

```
\documentclass{article}
\title{Simple Example}
\author{Andrei Gurtov}
\date{March 2000}
\begin{document}
\maketitle
Hello world!
\end{document}
```

Tabular

- Columns
  - `\begin{tabular}{|...|...|}`
  - `\end{tabular}`
- Rows
  - `&` - Split text into columns
  - `\\` - End a row
  - `\hline` - Draw line under row
  - e.g. `123123 & 34.00\\ \hline`



Example of table

```
\begin{tabular}{|l|r|c|} \hline
Date & Price & Size \\ \hline
Yesterday & 5 & big \\ \hline
Today & 3 & small \\ \hline
\end{tabular}
```

Date	Price	Size
Yesterday	5	Big
Today	3	Small

Floating Bodies

- Floating bodies can stop splitting of tables and images over pages.

```
\begin{figure}[options]
\begin{table}[options]
```

Options (recommendations)
h = place table here
t = place at top of page
b = place at bottom of page

- They will now appear in the LOF and LOT.

Example of floating figure

- \begin{figure}[ht]
- \centering\epsfig{file=uni.ps, width=5cm}
- \caption{University of Helsinki}
- \label{uni}
- \end{figure}

Figure~\ref{uni} shows...

Images

- Use epsfig package
- `\usepackage{epsfig}`
- Including images in main body
- `\epsfig{file=filename.eps, width=10cm, height=9cm, angle=90}`
- Creating EPS - Use xv and/or xfig.
- MS Power Point, save as GIF and convert to EPS.

Bibliography by hand

```
\begin{thebibliography}{}
\bibitem[Come95]{Come95} Comer,
D. E., {\it Internetworking with TCP/IP:
Principles, Protocols and Architecture},
volume 1, 3rd edition. Prentice-Hall,
1995.
\end{thebibliography}
```

Bibliography using Bibtex

- Bibliography information is stored in a *.bib file, in Bibtex format.
- Include chicago package
  - \usepackage{chicago}
- Set referencing style
  - \bibliographystyle{chicago}
- Create reference section by
  - \bibliography{bibfile with no extension}

Bibliography using Bibtex

```
@book{Come95,
author="D. E. Comer",
title={Internetworking with TCP/IP: Principles,
   Protocols and Architecture},
publisher="Prentice-Hall",
year=1995,
volume=1,
edition="Third"}
```

Bibliography contd.

- Citing references in text
  - `\cite{cuc98}` = (Cuce 1998)
  - `\citeN{cru98}` = Crud (1998)
  - `\shortcite{tom98}` = (Tom, et. al. 1998)
- Creating Bibtex Files
  - Use Emacs with extensions.
  - or copy Bibtex entries from bibliography database.

Some Math

```
\begin{center}
{\large
$$ y=\frac{a^3+2c_{x}}{1+\sqrt{b_{x}}} $$ \\
\vspace{0.2in}
$$ Q=\sum_{i=1}^{j}\int_{\mu}^{\infty}f(x_{j})dx $$
      \\
\vspace{0.2in}
$$ \Psi = \oint_{- \infty}^{\infty}f_{xy}({\frac{\partial
Qx}{\partial Qy}})^{\Im_{\pi}^ \prime} $$ \\ }
```

$$y = \frac{a^3 + 2c_x}{1 + \sqrt{b_x}}$$

$$Q = \sum_{i=1}^{j} \int_{\mu}^{\infty} f(x_j) dx$$

$$\Psi = \oint_{-\infty}^{\infty} f_{xy} \left(\frac{\partial Qx}{\partial Qy}\right)^{\Im_{\pi}}$$

Tools

UNIX based systems

- xdvi, ghostview, fixps, emacs with latex/bibtex support.

Windows 98/NT

- Ghostview, Acrobat Distiller, Acrobat Reader, Scientific Workplace (not the best), the Bibtex viewer is good. Paint Shop Pro, Latex and Emacs

Web

- Overleaf : https://www.overleaf.com/

Exercise 2

- Create an account at Overleaf: https://www.overleaf.com/
- Create a new documents to try the use of Latex