Using a computer as a computer scientist

Louis Jachiet
Digital life
Digital life

Mails
Advantages

Mail is somewhat standard
sending text and files to almost anyone

Mail is federated
does not belong to a single entity

Mail is asynchronous
generally instantaneous
Disadvantages

Many extensions are non standard
encryption, receipt, colorful mails, etc.

Mail is federated
each provider runs its own infrastructure which can create friction

Mail is asynchronous
mail is not always fast...
From: duck42@spacelasers.com
To: Prof 3
Cc: Prof 2, Random person
Bcc: Prof 1
Subject: Mail

I had trouble with computers last week first I tried to reinstall my computer and it failed then I decided to use my friend computer but I had to start over and I was no able to send my work yesterday. Here is my file

Attachment: tp.rar

What is wrong?
How to write an email to teachers

- Use a proper mail address
- Select carefully the recipients
- Use a proper subject
- Be polite
- Be brief but describe:
  - the problem
  - your attempt
  - an eventual (short but VALID) excuse
- For attachments:
  - use meaningful names
  - use standardized file types (e.g. PDF, ZIP) not docx
  - remove useless files
  - keep size small (if possible)
Dear teacher,

I am firstname LASTNAME and I am in your course DATAAI 101. We were supposed to submit the TP 4 on moodle before the 5th but sadly I missed the deadline and now the submission on moodle is closed.

I am really sorry for this but would you consider the attached file? Or allow me to submit on moodle?

Warm Regards,

–

Firstname Lastname

Attachment: LASTNAME_firstname_TP4_DATAAI_101.zip
Your @telecom-paris.fr address is not forever!

It will disappear a few months after your departure,

be wary when communicating with this address!

Important communications will be sent to this address,

check it regularly!

It comes with a calendar tool, don’t hesitate to use such a tool!
Digital life

Security
Threat model

For each data storage determine:

- Who has access to it
- Against what I want to protect myself?
- How do I protect myself against it?
Confidentiality

Very low

Could be put on my webpage?
Confidentiality

Very low

Could be put on my webpage?

- If it includes your mail, what about spam?
Confidentiality

Very low

Could be put on my webpage?

- If it includes your mail, what about spam?
- If it includes a lot of personal infos how to protect “password lost” questions and identity theft?
Confidentiality

Low

Who can use it?
Confidentiality

Low

Who can use it?

- Scammers.
Confidentiality
High

What are the risks?

• Scammers.
• Extortion.
• Intimate mails being leaked.
• Stolen online accounts.
Confidentiality

High

What are the risks?

- Scammers.
**Confidentiality**

High

---

**What are the risks?**

- Scammers.
- Extortion.
Confidentiality

High

What are the risks?

- Scammers.
- Extortion.
- Intimate mails being leaked
Confidentiality
High

What are the risks?

- Scammers.
- Extortion.
- Intimate mails being leaked
- Stolen online accounts
Confidentiality

High

What are the risks?

- Scammers.
- Extortion.
- Intimate mails being leaked
- Stolen online accounts
- ...
How to secure your data?

Passwords
- A different password for each website (use a password manager!)
- Strong passwords

Storage devices
- Encrypt all the drives containing data
- Don’t use random USB keys
Crafting passwords?

**Tr0ub4dor & 3**
- **UNCOMMON** (non-gibberish) base word
- **ORDER** unknown
- **CAPS?**
- **COMMON SUBSTITUTIONS**
- **NUMERAL**
- **PUNCTUATION**

(You can add a few more bits to account for the fact that this is only one of a few common formats)

**~28 BITS OF ENTROPY**

$2^{29} = 3$ days at 1000 guesses/sec

(Plausible attack on a weak remote web service: Yes, cracking a stolen hash is faster, but it's not what the average user should worry about.)

**DIFFICULTY TO GUESS:** EASY

**DIFFICULTY TO REMEMBER:** HARD

**correct horse battery staple**
- **FOUR RANDOM COMMON WORDS**

**~44 BITS OF ENTROPY**

$2^{44} = 530$ years at 1000 guesses/sec

**DIFFICULTY TO GUESS:** HARD

**DIFFICULTY TO REMEMBER:**

Through 20 years of effort, we've successfully trained everyone to use passwords that are hard for humans to remember, but easy for computers to guess.

**WAS IT TROMBONE? NO, TROUBADOR. AND ONE OF THE O's WAS A ZERO?**

And there was some symbol...

**THAT'S A BATTERY STAPLE. CORRECT!**

You've already memorized it.
Crafting passwords

Algorithm

- Select 3 to 4 random words
- Separate them with random numbers or punctuation

Example

- mechanization / preference / outback / apologizing
- Separate them with numbers or punctuation

Result

mechanization4preference;outback_apologizing

Entropy

30 000 words $\rightarrow 30000^4 \approx 10^{18}$ possibilities
Digital life

Backups
Who is the biggest enemy of your data?
Who is the biggest enemy of your data?

YOU

VIRUS DETECTED
Who is the biggest enemy of your data?

![Hacker图标](image1.png)

![NSA图标](image2.png)

![病毒检测图标](image3.png)
Who is the biggest enemy of your data?

YOU

NATIONAL SECURITY AGENCY

ALL YOUR DATA

EVIL CORP

VIRUS DETECTED
Who is the biggest enemy of your data?

- YOU
- EVIL CORP
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VIRUS DETECTED
- VI...
Who is the biggest enemy of your data?

YOU
What can happen to your data?

- Loose it
- Being stolen
- Have an accident (fire, flood, etc.)
- Hard disk failure
- Data corruption
- Accidental deletion
- ...

One of these items happens every few years to everyone...
How to backup data

Golden rule: 3 times

- three different storage types,
- three different places,
- three types of backups.

Easy mistake: can you recover?

It is important to check that EVERYTHING is really backed-up from time to time and that the data is really there:

- Closed backup provider
- Corrupted USB key
- Some crucial files are missing
- etc.
Digital life

Online presence
Web

What is online will stay online...

but you can somewhat control it by adding content.
Creating a webpage can be done in a free and easy way, for instance:

- Github pages
- Gitlab
- Wordpress
- Netlify
- and many other possibilities...
Tools for computer scientists
Tools for computer scientists

Bibliography
Producing science requires to be up to date on recent advances.
Producing science requires to be up to date on recent advances. In computer science, science is first published in research papers (theses, journals, conferences, etc.).
Producing science requires to be up to date on recent advances. In computer science, science is first published in research papers (theses, journals, conferences, etc.).

To pursue science you need to find, get access, read, remember and cite the relevant article for your research.
Finding articles

Research articles are generally not easy to read. When starting a completely new subject, it is a good idea to read blog posts or books about it rather.

When starting an internship or a PhD most of the bibliography will be provided by your advisor.

When looking for a specific reference or with a specific question you can use a search engine, the most popular being Google Scholar.
Finding article with Google scholar

Google Scholar

Toutes les langues  Rechercher les pages en Français

Sur les épaules d'un géant
finding article with google scholar

succinct partial sums and fenwick trees
p bille, ar christiansen, n prezza... - ... symposium on string ..., 2017 - springer
... we present two succinct versions of the fenwick tree – which is known for its simplicity and practicality. our ... parallelization. keywords. partial sums fenwick tree succinct parallel. download conference paper pdf. 1 introduction ...

[PDF] fenwick tree and its application in solving the sum of subsequence
d pan, y chen - citeseer
the experimental program is executed and its result reveals that the consuming time of achieving the sum of subsequence by ft is far below the time of two commonly methods (denoted as, m1 and m2), and its average running velocity is about 182 times faster than ...

[PDF] psu.edu

a new data structure for cumulative frequency tables
pm fenwick - software: practice and experience, 1994 - wiley online library
a new method (the 'binary indexed tree') is presented for maintaining the cumulative frequencies which are needed to support dynamic arithmetic data compression. it is based on a decomposition of the cumulative frequencies into portions which parallel the binary ...

[PDF] psu.edu

segment and fenwick trees for approximate order preserving matching
r niquefa, j mendivelso, g hernández... - workshop on engineering ..., 2017 - springer
... we formally define the \(\delta \gamma\)-order-preserving matching problem. we designed two algorithms for it based on the segment tree and the fenwick tree, respectively. also ... 2.2 fenwick tree-based algorithm (bitpa). the
Finding article with Google scholar

Succinct partial sums and fenwick trees
P Bille, AR Christiansen, N Prezza... - ... Symposium on String ..., 2017 - Springer
... We present two succinct versions of the Fenwick Tree – which is known for its simplicity and practicality. Our ... parallelization. Keywords. Partial sums Fenwick tree Succinct Parallel. Download conference paper PDF. 1 Introduction ...

Fenwick Tree and its Application in Solving the Sum of Subsequence
D Pan, Y Chen - Citeseer
The experimental program is executed and its result reveals that the consuming time of achieving the sum of subsequence by FT is far below the time of two commonly methods (denoted as, M1 and M2), and its average running velocity is about 182 times faster than ...

A new data structure for cumulative frequency tables
PM Fenwick - Software: Practice and experience, 1994 - Wiley Online Library
A new method (the 'binary indexed tree') is presented for maintaining the cumulative frequencies which are needed to support dynamic arithmetic data compression. It is based on a decomposition of the cumulative frequencies into portions which parallel the binary ...

Segment and fenwick trees for approximate order preserving matching
R Niquefa, J Mendivelso, G Hernández... - Workshop on Engineering ..., 2017 - Springer
... We formally define the \( \delta \gamma \)-order-preserving matching problem. We designed two algorithms for it based on the segment tree and the Fenwick tree, respectively. Also, ... 2 Fenwick Tree Based Algorithm (hitPA).

[PDF] arxiv.org
[PDF] psu.edu
[PDF] psu.edu
Finding article with Google scholar

Succinct partial sums and fenwick trees
P Bille, AR Christiansen, N Prezza... - ... Symposium on String ..., 2017 - Springer
We present two succinct versions of the Fenwick Tree – which is known for its simplicity and practicality. Our parallelization. Keywords. Partial sums Fenwick tree Succinct
Parallel. Download conference paper PDF. 1 Introduction ...

Fenwick Tree and its Application in Solving the Sum of Subsequence
D Pan, Y Chen - CiteSeer
The experimental program is executed and its result reveals that the consuming time of achieving the sum of subsequence by FT is far below the time of two commonly methods (denoted as, M1 and M2), and its average running velocity is about 182 times faster than ...

A new data structure for cumulative frequency tables
PM Fenwick - Software: Practice and experience, 1994 - Wiley Online Library
A new method (the 'binary indexed tree') is presented for maintaining the cumulative frequencies which are needed to support dynamic arithmetic data compression. It is based on a decomposition of the cumulative frequencies into portions which parallel the binary ...

Segment and fenwick trees for approximate order preserving matching
R Niquefa, J Mendivelso, G Hernández... - Workshop on Engineering ..., 2017 - Springer
... We formally define the \(\delta \gamma\)-order-preserving matching problem. We designed two algorithms for it based on the segment tree and the Fenwick tree, respectively. Also, ... 2 Fenwick Tree Based Algorithm (hitPA). The...
Finding article with Google scholar

Succinct partial sums and fenwick trees

P Bille, AR Christiansen, N Prezza... - ... Symposium on String ..., 2017 - Springer

We present two succinct versions of the Fenwick Tree – which is known for its simplicity and practicality. Our ... parallelization. Keywords. Partial sums Fenwick tree Succinct Parallel. Download conference paper PDF. 1 Introduction ...

Fenwick Tree and its Application in Solving the Sum of Subsequence

D Pan, Y Chen - Citeseer

The experimental program is executed and its result reveals that the consuming time of achieving the sum of subsequence by FT is far below the time of two commonly methods (denoted as, M1 and M2), and its average running velocity is about 182 times faster than ...

A new data structure for cumulative frequency tables

PM Fenwick - Software: Practice and experience, 1994 - Wiley Online Library

A new method (the 'binary indexed tree') is presented for maintaining the cumulative frequencies which are needed to support dynamic arithmetic data compression. It is based on a decomposition of the cumulative frequencies into portions which parallel the binary ...

Segment and fenwick trees for approximate order preserving matching

R Niquefa, J Mendivelso, G Hernández... - Workshop on Engineering ..., 2017 - Springer

... We formally define the \(\delta \gamma\)–order-preserving matching problem. We designed two algorithms for it based on the segment tree and the Fenwick tree, respectively. Also ... 3. 2 Fenwick Tree Based Algorithm (hitPA). The
Finding article with Google scholar

Segment and fenwick trees for approximate order preserving matching
R Niquefa, J Mendivelso, G Hernández... - Workshop on Engineering ..., 2017 - Springer
... We formally defined the (binary tree) order preserving matching problem. We designed two algorithms, respectively. Also ...

Compact Fenwick tree patriot version
S Marchini, S Vigna - arXiv, 2019
Summary The Fenwick tree is a data structure for computing partial sums of a sequence of elements in such a way that range sum queries are efficiently answered. Its space and time requirements are low. The Fenwick tree is an powerful and versatile data structure with a wide range of applications. It is used in various fields such as computer science, engineering, and data analysis. The Fenwick tree is a popular choice for implementing cumulative frequency tables, which are commonly used in statistics, probability, and other fields.

Partial Sums on the Ultra-Wide Word RAM
P Bille, IL Görtz, FR Skjoldjensen - ... on Theory and Applications of Models ..., 2020 - Springer
... Our results are based on a simple and elegant in-place word RAM data structure, known as the Fenwick tree. Our main technical contribution is a new efficient parallel ultra-wide word RAM implementation of the Fenwick tree, which is likely of independent interest. Keywords ...
Getting access to articles

Succinct partial sums and fenwick trees

P. Bille, A.R. Christiansen, N. Prezza... - ... Symposium on String ..., 2017 - Springer

... We present two succinct versions of the Fenwick Tree – which is known for its simplicity and practicality. Our parallelization. Keywords: Partial sums Fenwick tree Succinct Parallel. Download conference paper PDF. 1 Introduction ...

Fenwick Tree and its Application in Solving the Sum of Subsequence

D. Pan, Y. Chen - Citeseer

The experimental program is executed and its result reveals that the consuming time of achieving the sum of subsequence by FT is far below the time of two commonly methods (denoted as, M1 and M2), and its average running velocity is about 182 times faster than ...

A new data structure for cumulative frequency tables

PM Fenwick - Software: Practice and experience, 1994 - Wiley Online Library

A new method (the 'binary indexed tree') is presented for maintaining the cumulative frequencies which are needed to support dynamic arithmetic data compression. It is based on a decomposition of the cumulative frequencies into portions which parallel the binary ...

Segment and fenwick trees for approximate order preserving matching

R. Niquefa, J. Mendivelso, G. Hernández... - Workshop on Engineering ..., 2017 - Springer

... We formally define the \( \delta \gamma \)-order-preserving matching problem. We designed two algorithms for it based on the segment tree and the Fenwick tree, respectively. Also...
Getting access to articles

Some articles are behind a paywall what can you do:

- See if Google scholar found an opened version
- Access through Télécom (Télécom is subscribed to several publishers)
- Use libgen / sci-hub (legality to discuss)
Getting access to articles with sci-hub

sci-hub

...to remove all barriers in the way of science

enter URL, PMID / DOI or search
Getting access to articles

SCI-HUB
...to remove all barriers in the way of science

10.1007/978-3-319-66963-2_13

open
Segment and Fenwick Trees for Approximate Order Preserving Matching

Rafael Niquefa\textsuperscript{1}, Juan Mendivelso\textsuperscript{2,3} (✉), Germán Hernández\textsuperscript{4}, and Yoan Pinzón\textsuperscript{5}

\textsuperscript{1} Facultad de Ciencias e Ingeniería, Politécnico Grancolombiano Institución Universitaria, Calle 57 # 3 - 00 Este, i Tower Bogotá, Bogotá, Colombia  
\texttt{rniquefa@poligran.edu.co}

\textsuperscript{2} Departamento de Matemáticas, Universidad Nacional de Colombia, Bogotá, Colombia  
\texttt{jmendivelso@unal.edu.co}

\textsuperscript{3} Facultad de Matemáticas e Ingenierías, Fundación Universitaria Konrad Lorenz, Bogotá, Colombia

\textsuperscript{4} Departamento de Ingeniería de Sistemas e Industrial, Universidad Nacional de Colombia, Bogotá, Colombia

\textsuperscript{5} Departamento de Electrónica y Ciencias de la Computación, Pontificia Universidad Javeriana, Cali, Colombia

Abstract. In this paper we combine two string searching related problems: the approximate string matching under parameters $\delta$ and $\gamma$, and the order preserving matching problem. Order-preserving matching regards
Read and remember articles

When you have read 3 articles in the past month, it is easy to remember:

- What each article says
- In what version of the article you can find the result
- The precise reference of the article

After three years (duration of PhD in France) and 100s of articles, it is much harder...
Read and remember articles

It is a very good thing to have *some system* to keep track of which article you read and what it contained as well as the BibTeX reference.

- A folder with the PDF of articles and notes
- A wiki or some other note taking tool
- Zotero
Read and remember articles with Zotero
Tools for computer scientists

Writing research
In CS, most conferences expect articles to be formatted through LaTeX which is a software layer over \TeX. 

\begin{frame}{LaTeX}
  In CS, most conferences expect articles to be formatted through \LaTeX which is a software layer over \TeX.

  \hline

  \inputmintedbox{LaTeX}{recslide.tex}
  {\LaTeX code for this slide}
\end{frame}
Pros

- Easy to write science
- Separation of styling and content
- Can be “programmed”
- Works well with git and so
- Good looking documents by default

Cons

- Learning curve is hard for first timers
- Requires help to do something new
- Requires installation of software to write
One big advantage of \LaTeX is that there exists templates for a multitude of use cases:

- research articles
- presentations
- résumé
- letters
- ...
Drawing figures with Tikz

Research articles often contains complex figures such as:

![Diagram of complexity classes]

- LOGTIME
- LOGSPACE
- PTIME
- NP
- co-NP
- PSPACE
- NP
Research articles often contain complex figures such as:

\begin{tikzpicture}
\tiny \pgftransformscale{.8}
\draw[very thick] (6,0) -- (-6,0);
\draw (-1.3,0) parabola bend (0,2.5) (1.3,0); \node at (0,1) { LOGTIME };
\draw (-2,0) parabola bend (0,3.5) (2,0); \node at (0,2.6) { LOGSPACE };
\draw (-3,0) parabola bend (0,4.5) (3,0); \node at (0,4) { PTIME };
\draw[dotted] (-4,0) parabola bend (2,6) (4.5,0); \node at (3,3.5) { NP };
\draw[dashed] (4,0) parabola bend (-2,6) (-4.5,0); \node at (-2.5,4) { co-NP };
\draw (-6,0) parabola bend (0,7.2) (6,0); \node at (0,6.5) { PSPACE };
\end{tikzpicture}
\vspace{-2em}

\textbf{Figure 1:} Example of Tikz use from texample.net
For articles, always prefer vector graphics to raster graphic:

- lighter
- scalable
- editable
Drawing figures with inkscape
Tools for computer scientists

Version control systems
Version Control Systems (VCS)

A system eventually capable of:

- Capture the evolution of files
- Synchronize files across multiple computers
- Allow multiple users to work in parallel on the same documents
Version Control Systems (VCS)

A system eventually capable of:

- Capture the evolution of files
- Synchronize files across multiple computers
- Allow multiple users to work in parallel on the same documents

It can be applied to a wide range of types of files:

- Source code
- Websites
- Reports
- Work environments
- ...
Popular VCS

The most popular ones are:

- Git
- Bazaar
- mercurial
- SVN (outdated)
- RCS (very outdated)
While Git works in a fully decentralized way but whole ecosystems have emerged around it with version control but also:

- access management
- file hosting
- wiki
- pull requests
- continuous integration
- ...
While Git works in a fully decentralized way but whole ecosystems have emerged around it with version control but also:

- access management
- file hosting
- wiki
- pull requests
- continuous integration
- ...
**K Gst: Knowledge Graph Summarization for Anomaly Detection & Completion**
Gitlab example

**Inkscape's CLI has changed in version 1.0**
Tarik Graba authored 1 year ago

**Makes the poster theme coherent with the presentation**
Tarik Graba authored 3 months ago

**Replace -- by \textendash for subsubitems**
Tarik Graba authored 9 months ago

**Merge branch 'master' of gitlab.enst.fr:latex/beamer-tpt-ng**
Cédric Ware authored 1 year ago

**Bump release version.**
Cédric Ware authored 1 year ago

**Fix default font search list.**
Cédric Ware authored 1 year ago

**Different conf logo in the poster example to avoid compilation errors**
Tarik Graba authored 1 year ago
Tools for computer scientists

SSH
**SSH principles**

**Shell**

Computers can be controlled with a *Graphical User Interface (GUI)* but historically they can also be controlled from the *Command line*.

**Secure SHEll (SSH)**

Allow to get to the command line of another computers
SSH

SSH has many uses:

- Use computer from afar
  
  reserved for power users

- Run computation
  
  useful to run computation in e.g. lab rooms

- Move files
  
  useful to move files between the school system and your computer

- Protocol of connection (git)
  
  GIT often relies on ssh to “push” your changes

- Tunnel internet (much like a VPN)
  
  Access pay-walled articles from home
Good practices for programming
Good practices for programming

Writing code
There exists programming environments allowing you to:

- Have syntax coloring
- Automatic indenting or completion
- Shortcut for compilation/run and navigating through the errors
- ...

Don’t write code on basic notepads!
Good practices for programming

Virtualenv
Advanced classes will often require you to install python packages. To avoid “polluting” your computer you can install them in virtualenv.

```
# creating the virtual env
virtualenv nameOfTheVirtualenv

# using the virtualenv
source nameOfTheVirtualenv/bin/activate  # activation
which pip  # check which pip we are using
pip install package name  # install package in the venv
python  # run python in the venv
jupyter-notebook  # run jupyter equipped with the venv
```
Good practices for programming

Notebooks
Jupyter notebooks

In [44]: def make_numerical(df, cols):
   ...:     new_df = df.copy()
   ...:     for col in cols:
   ...:         new_df[col] = new_df[col].astype('category')
   ...:         new_df[col] = new_df[col].cat.codes
   ...:     return new_df

In [45]: raw_penguins = pd.read_csv('penguins.csv', sep=' ', header='infer')

In [46]: penguins = raw_penguins.dropna()

In [61]: num_data = make_numerical(penguins,['island','species','sex'])
num_data

Out[61]:
    species  island  bill_length_mm  bill_depth_mm  flipper_length_mm  body_mass_g  sex  year
0       0         2            39.1           18.7              181.0     3750.0   1  200
1       0         2            39.5           17.4              186.0     3800.0   0  200
2       0         2            40.3           18.0              195.0     3250.0   0  200
3       0         2            36.7           19.3              193.0     3450.0   0  200
4       0         2            39.3           20.6              190.0     3650.0   1  200
...      ...        ...              ...            ...                ...       ...   ...  ...
339     1         1            55.8           19.8              207.0     4000.0   1  200
340     1         1            43.5           18.1              202.0     3400.0   0  200
341     1         1            43.3           18.0              199.0     3750.0   1  200
Plotting bill length in function of flipper length (and show species)

```python
In [62]:
plt.figure(get_plot())
plt.scatter(num_data['bill_length_mm'], num_data['flipper_length_mm'],
plt.title('Scatter penguins')
plt.xlabel('bill length')
plt.ylabel('flipper length')
plt.show()
```

<IPython.core.display.Javascript object>
Notebooks are great for:

- Test quickly code
- Manipulate data
- Visualize data
- Present your work

Overall it is great for “write-only” code.

Notebooks have HUGE limitations:

- You need to know the order in which the cells were launched
- Some piece of code might be deleted
- Hard to test
- Hard to version
Notebooks are packaged in python:

```bash
# if using venv:
virtualenv ~/jupyterEnv
source ~/jupyterEnv/bin/activate

# installing them
pip install jupyter
jupyter-notebook
```
Good practices for programming

Writing code
Why write readable code?

Generally, code is written once and read many times:

- by other people
  - peer-review,
  - homework correction,
  - inclusion in other people code
- also by yourself from the future
  - debugging,
  - restarting a paused code session,
  - copying code from the past,
  - etc.

Even when writing “write-only” code, it is a good exercise to train to write readable code.
How write readable code?

The main tools to write readable code are:

- Use standardized code formatting
- Decompose your code into functions / modules
- Naming correctly variables and functions
- Comments
In general, be consistent on:

- indentation
- white spaces / tabulations
- capitalization
- style of functions calls
- style of comments

There are established conventions on code formatting

- see e.g. https://pep8.org/ for python
- also there are tools to auto-format
Decomposing code

Writing code is all about taking a complex problem and dividing it recursively into subproblems.

Consequences:

- Don’t reinvent the wheel each time
- Decompose your big projects into smaller projects
- Decompose your code into files
- Decompose your files into functions and classes
- Only a handful of variables should be global
- Files and functions should remain “small”

However, don’t overdo this...
Naming functions / variables

Functions should explain what they do from a high level perspective:

❌ doStuff
Naming functions / variables

Functions should explain what they do from a high level perspective:

- doStuff
- ComputeVAT
Naming functions / variables

Functions should explain what they do from a high level perspective:

- ✗ doStuff
- ✓ ComputeVAT
- ✗ SumMultiplyCategoryToPrices

Variables should explain what they store:

- ✗ ff
- ✗ nbFalsePositives
- ✓ numberOfWrongItems, GoodItemsCount (choose one)
- ✓ cleanAbort
- ✗ loadDataset
- ✗ readFile
Naming functions / variables

Functions should explain what they do from a high level perspective:

- doStuff
- ComputeVAT
- SumMultiplyCategoryToPrices
- cleanAbort
- loadDataset
- readFile

Variables should explain what they store:

- ff
Naming functions / variables

Functions should explain what they do from a high level perspective:

- doStuff 
- ComputeVAT
- SumMultiplyCategoryToPrices
- cleanAbort
- loadDataset
- readFile

Variables should explain what they store:

- ff
- nbFalsePositives
Naming functions / variables

**Functions should explain what they do from a high level perspective:**

- [x] doStuff
- [x] ComputeVAT
- [x] SumMultiplyCategoryToPrices
- [x] cleanAbort
- [x] loadDataset
- [x] readFile

**Variables should explain what they store:**

- [x] ff
- [x] nbFalsePositives
- [x] numberOfWrongItems, GoodItemsCount (choose one)
Naming functions / variables

Functions should explain what they do from a high level perspective:

✗ doStuff
✓ ComputeVAT
✗ SumMultiplyCategoryToPrices
✓ cleanAbort
✓ loadDataset
✗ readFile

Variables should explain what they store:

✗ ff
✓ nbFalsePositives
✗ numberOfWrongItems, GoodItemsCount (choose one)
A good comment is one that:

- provides some documentation on the code
- explain the structure of the code
- provides insights on some part of the code
- explain a piece of code from a high-level perspective
- Does not repeat the code
A good comment is one that:

- provides some documentation on the code
  
  *This function takes* $x$ *and* $y$ *and returns* $x \times y$

- explain the structure of the code
  
  *Here we load the data, here we train the model, etc.*

- provides insights on some part of the code
  
  $x+=1$ / *This accounts for the top square*

- explain a piece of code from a high-level perspective
  
  *Here we try to pre-compute the distance of all points to A*

- Does not repeat the code

  `computeShortestPath();` // will compute shortest path
You will mainly have write-once read-once code this year but you should start using good practices because writing does not take more time and you will need it someday!
Good practices for programming

Testing your code
Tests

Unit tests
Unit tests tries some very specific piece of code to make sure it works ok.

*e.g. test each function or each member of each class*

Functional tests
Runs the whole program and make sure the output is what was expected.

*easy to set up for programs without interfaces*

Regression tests
Each time a bug is discovered, we introduce tests to make sure this bug is covered.
Example of unittest package in Python

See https://docs.python.org/3/library/unittest.html

```python
import unittest

class TestStringMethods(unittest.TestCase):
    def test_upper(self):
        self.assertEqual('foo'.upper(), 'FOO')

    def test_isupper(self):
        self.assertTrue('FOO'.isupper())
        self.assertFalse('Foo'.isupper())

    def test_split(self):
        s = 'hello world'
        self.assertEqual(s.split(), ['hello', 'world'])
        # check that s.split fails when the separator is not a string
        with self.assertRaises(TypeError):
            s.split(2)

if __name__ == '__main__':
    unittest.main()
```

Louis JACHIET
Not using framework to test your code is **unthinkable** for professionals however for a short piece of code for a class you can test manually:

- check that the output has the correct format
- check on small examples it makes sense
- test the output is reasonable
- if possible run manually a few functional tests before submitting your work