

Using a computer as a computer scientist

Louis Jachiet

Digital life

Digital life

Mails – some general and technical considerations

What is a mail?

From a technical standpoint a mail is just a text file looking like this:

```
From: Louis Jachiet <test@jachiet.com>
To: Louis Jachiet <louis.jachiet@telecom-paris.fr>
Cc: cc-ed@jachiet.com
Subject: Re: This is a subject
Message-ID: <20220903164909.uksb6uca3o2zyyd4@moa.enst.fr>
MIME-Version: 1.0
Content-Type: text/plain; charset=us-ascii
Content-Disposition: inline
Content-Transfer-Encoding: quoted-printable
In-Reply-To: <20220903164842.fscuyipdyhc67pxk@moa.enst.fr>
```

Well received!

--

Louis Jachiet

What is a mail?

To send a mail, you contact a (mail) server that will take this file and send it to another server, that will send it to another server until it reaches the inbox of the recipients.

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

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...

How not to write an email

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?

Use a proper subject!

Keep it relatively short, but informative.

Be polite!

Unless you are replying to a recent mail:

- Add a salutation (Hello, Hi, Dear Teacher, Dear Ms X, etc.)
- Add an appropriate closing (Sincerely, Regards, Best Regards, Thank you)
- Use the correct language level
- Never write an email when you are angry / always down tone your angryness
- Don't use sarcasm

Be specific!

We you received hundreds of mails a day you spend 30s reading them (sometimes doing something else in parallel).

Each mail should be very specific in what you want!

Give context!

We often assume symmetric relation and that readers are our mails will have the same mental context as ours. *It is not the case!*

It is always better to add too much context!

Focus on what is *essential*.

Some advice

- The shape of your mail helps reading it!
- We (generally) don't need to know why you were late, sick, etc.
- Often, the body of a mail can fit the 280 chars tweet length

Timely follow up

We all loose tracks of mails sometimes, so you can send a reminder but:

- we have other things to do
 - we might take week-ends and vacation
 - the urgency for you is not a criterion if you are the reason for that urgency
-

Mails should ideally be sent during a week day before 5 pm.

Proof-read your mail

- Re-read everything you wrote
- Check all items above
- Check if attachments are needed

Digital life

How to use mail for Data AI

How to write an email

- 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 or pages
 - remove useless files
 - keep size small (if possible)

How to write an email

From: first.lastname@telecom-paris.fr

To: Main teacher

Cc: TAs

Subject: [DATAAI 101] Late submission for assignment XY

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

Using the Télécom mail

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

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?

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?

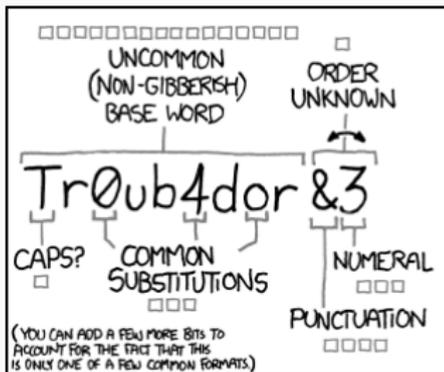
Passwords

- 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?



~28 BITS OF ENTROPY

$2^{28} = 3 \text{ DAYS AT } 1000 \text{ GUESSES/SEC}$

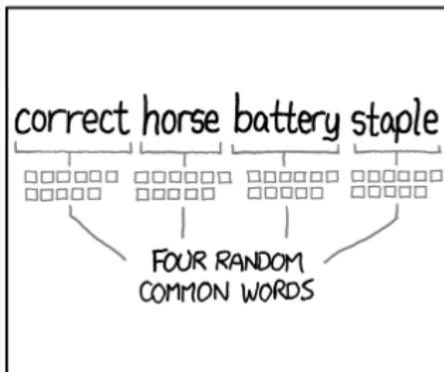
(PLAUSIBLE ATTACK ON A WEAK REMOTE WEB SERVICE. YES, CRACKING A STOLEN HASH IS FASTER, BUT IT'S NOT WHAT THE AVERAGE USER SHOULD WORRY ABOUT.)

DIFFICULTY TO GUESS: **EASY**

WAS IT TROMBONE? NO, TROUBADOR. AND ONE OF THE 0s WAS A ZERO?

AND THERE WAS SOME SYMBOL...

DIFFICULTY TO REMEMBER: **HARD**



~44 BITS OF ENTROPY

$2^{44} = 550 \text{ YEARS AT } 1000 \text{ GUESSES/SEC}$

DIFFICULTY TO GUESS: **HARD**

THAT'S A BATTERY STAPLE.

CORRECT!

DIFFICULTY TO REMEMBER: **YOU'VE ALREADY MEMORIZED IT**

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.

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 $\rightarrow 30\,000^4 \approx 10^{18}$ possibilities

Digital life

Backups

Who is the biggest enemy of your data?



Who is the biggest enemy of your data?



Who is the biggest enemy of your data?



Who is the biggest enemy of your data?



Who is the biggest enemy of your data?



YOUR COMPUTER HAS BEEN LOCKED!

This operating system is locked due to the violation of the federal laws of the United States of America! (Article 1, Section 8, Clause 8; Article 202; Article 230 of the Criminal Code of U.S.A. provides for a deprivation of liberty for four to twelve years.)
Following violations were detected:
Your IP address was used to visit websites containing pornography, child pornography, zoophilia and child abuse. Your computer also contains video files with pornographic content, elements of violence and child pornography! Spam-messages with terrorist motives were also sent from your computer.
This computer lock is aimed to stop your illegal activity.

To unlock the computer you are obliged to pay a fine of \$200.

You have 72 hours to pay the fine, otherwise you will be arrested.

You must pay the fine through _____
To pay the fine, you should enter the digits resulting code, which is located on the back of your _____ in the payment form and press OK (if you have several codes, enter them one after the other and press OK).



Who is the biggest enemy of your data?



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

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



The image shows the Google Scholar homepage. At the top left is a hamburger menu icon. At the top right is the text "CONNEXION". The main heading is "Google Scholar" in its characteristic multi-colored font. Below the heading is a search bar with a magnifying glass icon on the right. Under the search bar are two radio buttons: "Toutes les langues" (selected) and "Rechercher les pages en Français". Below this is the slogan "Sur les épaules d'un géant" in green. At the bottom left is a globe icon and "FR". At the bottom right are links for "Aide", "Confidentialité", and "Conditions".

CONNEXION

Google Scholar

Toutes les langues Rechercher les pages en Français

Sur les épaules d'un géant

FR [Aide](#) [Confidentialité](#) [Conditions](#)

Finding article with Google scholar



fenwick tree

Scholar Environ 22 300 résultats (0,08 s) ANNÉE

Succinct partial sums and fenwick trees [\[PDF\] arxiv.org](#)
[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 ...
☆  Cité 10 fois [Autres articles](#) [Les 13 versions](#)

Fenwick Tree and its Application in Solving the Sum of Subsequence [\[PDF\] psu.edu](#)
[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 ...
☆  [Autres articles](#) [Les 2 versions](#) 

A new data structure for cumulative frequency tables [\[PDF\] psu.edu](#)
[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 ...
☆  Cité 264 fois [Autres articles](#) [Les 6 versions](#)

Segment and fenwick trees for approximate order preserving matching
[R Niquefa](#), [J Mendivelso](#), [G Hernández](#)... - [Workshop on Engineering ...](#), 2017 - Springer
... We formally define the (δ, γ) -order-preserving matching problem. We designed two algorithms for it based on the segment **tree** and the **Fenwick tree**, respectively. Also, ... **Fenwick Tree** Based Algorithm (hiPA). The

Finding article with Google scholar

The screenshot shows a Google Scholar search for 'fenwick tree'. The search bar at the top contains the text 'fenwick tree' and a magnifying glass icon. Below the search bar, the text 'Scholar Environ 22 300 résultats (0,08 s)' is displayed, along with a dropdown menu for 'ANNÉE' and a hamburger menu icon. The search results are listed below, each with a title, authors, a brief description, and a link to a PDF version. The first result is 'Succinct partial sums and fenwick trees' by P Bille, AR Christiansen, and N Prezza, published in 'Symposium on String ...' in 2017 by Springer. The second result is 'Fenwick Tree and its Application in Solving the Sum of Subsequence' by D Pan and Y Chen, published by Citeseer. The third result is 'A new data structure for cumulative frequency tables' by PM Fenwick, published in 'Software: Practice and experience, 1994' by Wiley Online Library. The fourth result is 'Segment and fenwick trees for approximate order preserving matching' by R Niquefa, J Mendivelso, and G Hernández, published in 'Workshop on Engineering ...' in 2017 by Springer. A red hand-drawn circle highlights the PDF link for the third result, '[PDF] psu.edu'.

fenwick tree

Scholar Environ 22 300 résultats (0,08 s) ANNÉE

Succinct partial sums and fenwick trees [PDF] arxiv.org
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 ...
☆ 99 Cité 10 fois Autres articles Les 13 versions

[PDF] **Fenwick Tree** and its Application in Solving the Sum of Subsequence [PDF] psu.edu
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 ...
☆ 99 Autres articles Les 2 versions

A new data structure for cumulative frequency tables [PDF] psu.edu
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 ...
☆ 99 Cité 264 fois Autres articles Les 6 versions

Segment and fenwick trees for approximate order preserving matching
R Niquefa, J Mendivelso, G Hernández... - Workshop on Engineering ..., 2017 - Springer
... We formally define the (δ, γ) -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 (hiPA). The

Finding article with Google scholar

The screenshot shows the Google Scholar search interface. At the top, the search bar contains the text 'fenwick tree' and a magnifying glass icon. Below the search bar, the text 'Scholar Environ 22 300 résultats (0,08 s)' is displayed. On the right side, there is a dropdown menu for 'ANNÉE' and a hamburger menu icon. The search results are listed below, each with a title, authors, a brief abstract, and a link to a PDF file. The third result, 'A new data structure for cumulative frequency tables', has the number '264' circled in red in the 'Cité 264 fois' text.

fenwick tree

Scholar Environ 22 300 résultats (0,08 s) ANNÉE

Succinct partial sums and fenwick trees [PDF] arxiv.org
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 ...
☆ 99 Cité 10 fois Autres articles Les 13 versions

[PDF] **Fenwick Tree** and its Application in Solving the Sum of Subsequence [PDF] psu.edu
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 ...
☆ 99 Autres articles Les 2 versions

A new data structure for cumulative frequency tables [PDF] psu.edu
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 ...
☆ 99 Cité 264 fois Autres articles Les 6 versions

Segment and fenwick trees for approximate order preserving matching
R Niquefa, J Mendivelso, G Hernández... - Workshop on Engineering ..., 2017 - Springer
... We formally define the (δ, γ) -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 (biPA). The

Finding article with Google scholar

The screenshot shows a Google Scholar search interface. At the top, there is a search bar with the text 'fenwick tree' and a magnifying glass icon. Below the search bar, the text 'Scholar Environ 22 300 résultats (0,08 s)' is displayed, along with a dropdown menu for 'ANNÉE' and a hamburger menu icon. The search results are listed below, each with a title, authors, a brief description, and a link to a PDF. The third result, 'A new data structure for cumulative frequency tables', has a red circle around the citation count 'Cité 264 fois'.

fenwick tree

Scholar Environ 22 300 résultats (0,08 s) ANNÉE

Succinct partial sums and fenwick trees [PDF] arxiv.org
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 ...
☆ 99 Cité 10 fois Autres articles Les 13 versions

[PDF] **Fenwick Tree** and its Application in Solving the Sum of Subsequence [PDF] psu.edu
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 ...
☆ 99 Autres articles Les 2 versions

A new data structure for cumulative frequency tables [PDF] psu.edu
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 ...
☆ 99 Cité 264 fois Autres articles Les 6 versions

Segment and fenwick trees for approximate order preserving matching
R Niquefa, J Mendivelso, G Hernández... - Workshop on Engineering ..., 2017 - Springer
... We formally define the (δ, γ) -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 (biPA). The

Finding article with Google scholar

P.M. 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 ...

☆ 99 Cité 264 fois Autres articles Les 6 versions

Segment and fenwick trees for approximate order preserving matching

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

... We formally define the 1D data structure, under approximate matching problem. We

designed two algorithms for the problem, respectively. Also ...

☆ 99 Cité 2 fois

A scalable asymptotically optimal

HF Yu, CJ Hsieh, H ...
dl.acm.org

... these problems.

modified Fenwick tree

we use an appropriate

☆ 99 Cité 73 fois

Compact Fenwick trees

S Marchini, S Vignati

Summary The Fenwick tree

in such a way that it

sum and performing

☆ 99 Cité 2 fois

Partial Sums on the Ultra-Wide Word RAM

P Bille, IL Gørtz, 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 ...

[PDF] arxiv.org

javascript:void(0) articles Les 7 versions

Getting access to articles

fenwick tree

Scholar Environ 22 300 résultats (0,08 s) ANNÉE

Succinct partial sums and fenwick trees [PDF] arxiv.org
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 ...
☆ 99 Cité 10 fois Autres articles Les 13 versions

[PDF] **Fenwick Tree** and its Application in Solving the Sum of Subsequence [PDF] psu.edu
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 ...
☆ 99 Autres articles Les 2 versions

A new data structure for cumulative frequency tables [PDF] psu.edu
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 ...
☆ 99 Cité 264 fois Autres articles Les 6 versions

Segment and **fenwick trees** for approximate order preserving matching
R Niquefa, J Mendivelso, G Hernández... - Workshop on Engineering ..., 2017 - Springer
... We formally define the (δ, γ) -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 (biFA). The

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



The image shows the Sci-Hub website interface. On the left, a black crow with a red key in its beak is perched against a light-colored brick wall. In the top right corner, there is a gold medal icon with a red ribbon and a small black box containing the text: "the first website in the world to provide mass & public access to research papers". The main title "SCI-HUB" is written in large, bold, red letters. Below the title, a red arrow-shaped banner contains the text "...to remove all barriers in the way of science". A search bar with the placeholder text "enter URL, PMID / DOI or search" is positioned below the banner. To the right of the search bar is a red button with a white key icon and the word "open". At the bottom of the page, there is a dark navigation bar with the links "about", "ideas", "community", and "donate".

the first website in the world to provide mass & public access to research papers

SCI-HUB

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

enter URL, PMID / DOI or search

open

about ideas community donate

Getting access to articles



The first website in the world to provide mass & open access to research papers

SCI-HUB

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

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

 open

[about](#) [ideas](#) [community](#) [donate](#)

Getting access to articles

The image shows a browser window displaying a Sci-Hub article page. The browser's address bar shows '131 (1 of 13)' and 'Automatic Zoom'. The article title is 'Segment and Fenwick Trees for Approximate Order Preserving Matching'. The authors listed are Rafael Niquefa¹, Juan Mendivelso^{2,3(✉)}, Germán Hernández⁴, and Yoan Pinzón⁵. The affiliations are: ¹ Facultad de Ciencias e Ingeniería, Politecnico Gran Colombiano Institucion Universitaria, Calle 57 # 3 - 00 Este, 1 Tower Bogotá, Bogotá, Colombia; ² Departamento de Matemáticas, Universidad Nacional de Colombia, Bogotá, Colombia; ³ Facultad de Matemáticas e Ingenierías, Fundación Universitaria Konrad Lorenz, Bogotá, Colombia; ⁴ Departamento de Ingeniería de Sistemas e Industrial, Universidad Nacional de Colombia, Bogotá, Colombia; ⁵ Departamento de Electrónica y Ciencias de la Computación, Pontificia Universidad Javeriana, Cali, Colombia. The abstract states: 'In this paper we combine two string searching related problems: the approximate string matching under parameters δ and γ , and the order-preserving matching problem. Order-preserving matching regards...'. The Sci-Hub logo is visible in the top left corner, and a 'save' button is present. A 'Sci-Hub is fundraising' banner is at the bottom left. The browser's page number '131' is visible in the top bar.

sci hub
to open science

save

Niquefa, R., Mendivelso, J., Hernández, G., & Pinzón, Y. (2017). *Segment and Fenwick Trees for Approximate Order Preserving Matching*. *Applied Computer Sciences in Engineering*, 131-143. doi:10.1007/978-3-319-66963-2_13

url to share this paper:
sci-hub.se/10.1007/978-3-319-66963-2_13

Sci-Hub is fundraising
learn more -

created by: Alexandra Elbakyan

Segment and Fenwick Trees for Approximate Order Preserving Matching

Segment and Fenwick Trees for Approximate Order Preserving Matching

Rafael Niquefa¹, Juan Mendivelso^{2,3(✉)}, Germán Hernández⁴, and Yoan Pinzón⁵

¹ Facultad de Ciencias e Ingeniería, Politecnico Gran Colombiano Institucion Universitaria, Calle 57 # 3 - 00 Este, 1 Tower Bogotá, Bogotá, Colombia
rniquefa@poligran.edu.co

² Departamento de Matemáticas, Universidad Nacional de Colombia, Bogotá, Colombia
jcmendivelsom@unal.edu.co

³ Facultad de Matemáticas e Ingenierías, Fundación Universitaria Konrad Lorenz, Bogotá, Colombia

⁴ Departamento de Ingeniería de Sistemas e Industrial, Universidad Nacional de Colombia, Bogotá, Colombia

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 δ and γ , 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...

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

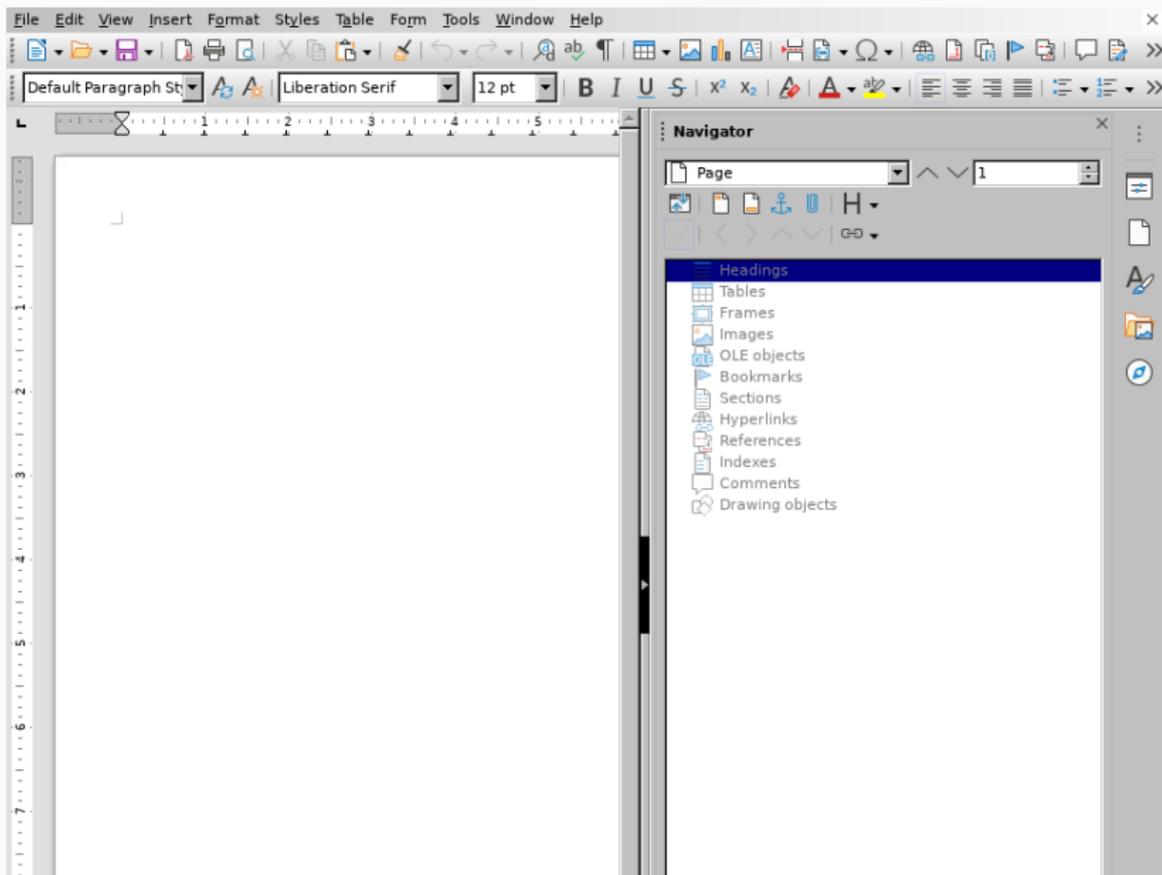
The screenshot shows the Zotero application window. On the left is a sidebar with a tree view of the library structure, including 'My Library' and 'Group Libraries'. The main area is a table of items with columns for Title, Creator, and Year. The item 'Circulation of Medicine in the Early Modern Atlantic World' by Cook and Walker (2013) is selected. On the right, a detailed view of this article is shown, including its title, author information, abstract, and publication details.

Title	Creator	Year
Guerra, maladie, empire. Les services de santé militaires en ...	Zaugg	2016
Officiers de santé et soignantes créoles face à la Fièvre jaune	Nobi	2016
The Emergence of Tropical Medicine in France	Osborne	2014
Colonial Disease, Translation, and Enlightenment: Franco-Briti...	Charters	2014
Trading in Drugs through Philadelphia in the Eighteenth Centu...	Wilson	2013
The Medicines Trade in the Portuguese Atlantic World: Acquisi...	Walker	2013
Leprosy and Slavery in Suriname: Godfried Schilling and the Fr...	Sneliders	2013
Medical Experimentation and Race in the Eighteenth-century ...	Schleibinger	2013
The Circulation of Bodily Knowledge in the Seventeenth-centu...	Gómez	2013
Circulation of Medicine in the Early Modern Atlantic World	Cook and Walker	2013
Synthesis of scholarship on "medicines" to restore focus o...		
Full Text PDF		
Colonial Medical Encounters in the Nineteenth Century: The Fr...	Thoral	2012
Networks in Tropical Medicine: Internationalism, Colonialism, a...	Neill	2012
Early Clinical Features of Dengue Virus Infection in Nicaragu...	Biswas et al.	2012
Medicine in an age of commerce and empire: Britain and its tr...	Harrison	2010
Finding the "Ideal Diet": Nutrition, Culture, and Dietary Practi...	Neill	2009
Battles of the Self: War and Subjectivity in Early Modern Fran...	Pichichero	2008
The Experiments of Ramón M. Terreyer SJ on the Electric Eel ...	de Asúa	2008
Psychiatry and Empire	Mahone and Vaughan	2007
Medicine and the Market in England and Its Colonies, C.1450-...	Jenner and Wallis	2007
Matters of exchange: commerce, medicine, and science in the...	Cook	2007
A Horrible Tragedy in the French Atlantic	Rothschild	2006
"Neither of meate nor drinke, but what the Doctor alloweth": ...	Chakrabarti	2006
Transnationalism in the colonies: Cooperation, rivalry, and rac...	Neill	2005
Variolation, Vaccination and Popular Resistance in Early Coloni...	Brimmes	2004
"Syphilis, Opiomania, and Pederasty": Colonial Constructions u...	Proschan	2003
Choosing Scientific Patrimony: Sir Ronald Ross, Alphonse Lav...	Gulieimin	2002
Madness and Colonization: Psychiatry in the British and Frenc...	Keller	2001
The Colonial Machine: French Science and Colonization in the ...	McClellan and Rego...	2000
From medical astrology to medical astronomy: sol-lunar and pl...	Harrison	2000
Disease and Empire: The Health of European Troops in the Col...	Bynum	2000
Climate & Consequences: Health, Race, Environment, and Bi...	Harper	1999

Info	Notes	Tags	Related
Item Type	Journal Article		
Title	Circulation of Medicine in the Early Modern Atlantic World		
Author	Cook, Harold J.		
Author	Walker, Timothy D.		
Abstract	The search for powerful drugs has caused people and commodities to move around the globe for many centuries, as it still does...		
Publication	Social History of Medicine		
Volume	26		
Issue	3		
Pages	337-351		
Date	2013/08/01		
Series			
Series Title			
Series Text			
Journal Abbr	Soc Hist Med		
Language	en		
DOI	10.1093/shm/hkt013		
ISSN	0951-631X		
Short Title			
URL	https://academic.oup.com/shm/article/26/3...		
Accessed	1/24/2018, 10:17:12 AM		
Archive			
Loc. in Archive			
Library Catalog			
Call Number			
Rights			
Extra			
Date Added	1/24/2018, 10:17:12 AM		
Modified	1/24/2018, 11:50:15 AM		

Tools for computer scientists

Writing research



In CS, most conferences expect articles to be formatted through L^AT_EX which is a software layer over T_EX.

L^AT_EXcode for this slide

```
\begin{frame}{\LaTeX}
```

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

```
\hline
```

```
\inputmintedbox{LaTeX}{reclslide.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

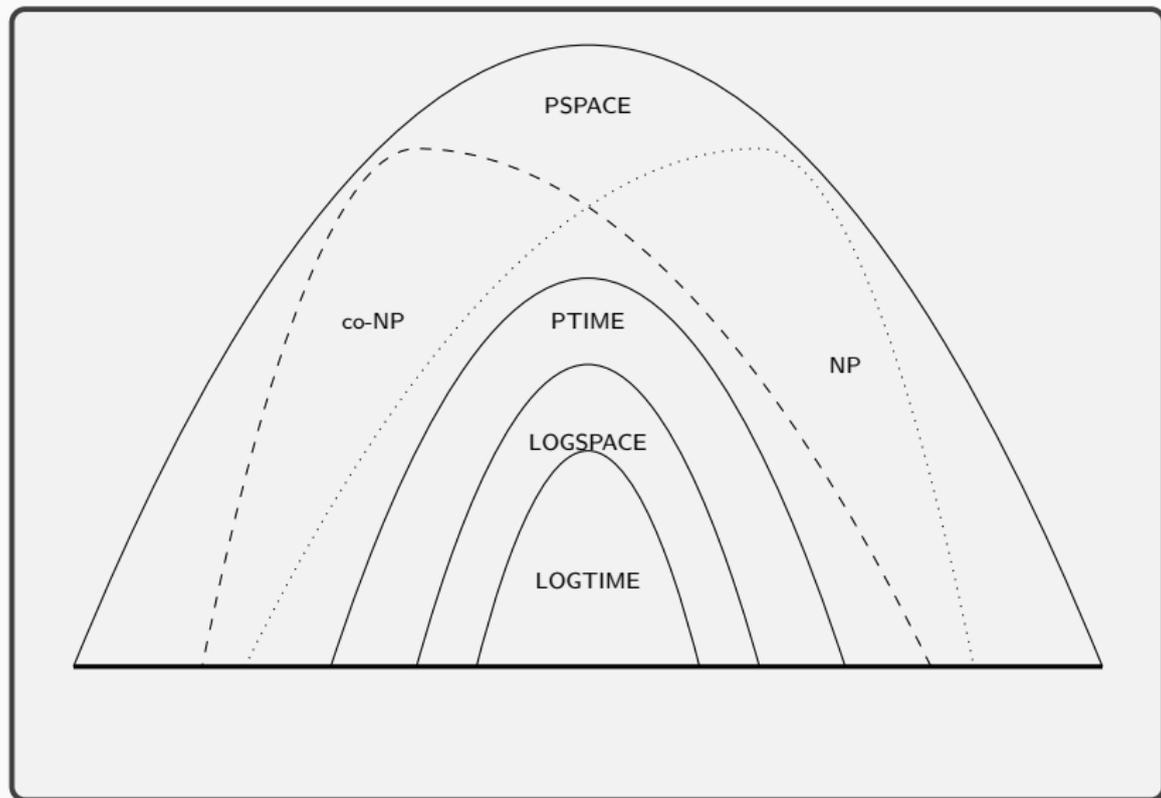
- Learning curve is hard for first timers
- Requires help to do something new
- Requires installation of software to write

One big advantage of L^AT_EX 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:



Research articles often contains 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}
```

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

File Edit View Layer Object Path Text Filters Extensions Help

Change: W: 76.606 - + H: 65.443 - + Rx: 0.000 - +

Fill: [redacted] Stroke: [redacted] 18.9

Align and Distribute (Shift+Ctrl+A)

Align

Relative to: Last selected

Distribute

Rearrange

Remove overlaps

H: 0.0 - + V: 0.0 - +

Fill and Stroke (Shift+Ctrl+F)

Fill Stroke paint Stroke style

Flat color

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
- ...

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
- ...

Github example

The screenshot shows the GitHub interface for the repository 'GemsLab / KGist'. At the top, there is a navigation bar with the GitHub logo, a 'Sign up' button, and a hamburger menu. Below this, the repository name 'GemsLab / KGist' is displayed, along with buttons for 'Notifications', 'Star' (15), and 'Fork' (5). A secondary navigation bar contains links for 'Code', 'Issues', 'Pull requests', 'Actions', 'Projects', 'Wiki', 'Security', and 'Insights'. The main content area is divided into three sections: a file browser, a commit history, and a README preview. The file browser shows a tree view with folders 'data', 'output', 'src', and 'test', and files '.gitignore', 'README.md', and 'license.txt'. The commit history shows a recent commit by 'cbelth' on Jun 12, 2020. The README preview shows the title 'KGist: Knowledge Graph Summarization for Anomaly Detection & Completion'. On the right side, there is an 'About' section with a description of the project and a list of related topics.

GemsLab / KGist Notifications Star 15 Fork 5

<> Code Issues Pull requests Actions Projects Wiki Security Insights

master Go to file Code

cbelth Update README.md with requirements and FAQ ... on Jun 12, 2020 30

data	added labels file to dbpedia.zip	15 months ago
output	updated documentation	2 years ago
src	add identify for anomaly detection	16 months ago
test	updated naming	16 months ago
.gitignore	initial commit	2 years ago
README.md	Update README.md with requirements and FAQ	15 months ago
license.txt	added license	16 months ago

☰ README.md

KGist: Knowledge Graph Summarization for Anomaly Detection & Completion

About

Knowledge Graph summarization for anomaly/error detection & completion (WebConf '20)

[gemslab.github.io/papers/...](https://gemslab.github.io/papers/)

dbpedia knowledge-graph mdl knowledge-graph-completion rule-mining minimum-description-length webconf neil www2020 graph-summarization

Readme View license

Releases

Gitlab example

The screenshot displays the GitLab interface for the 'Beamer TPT NG' project. The top navigation bar includes 'Projects', 'Groups', 'Snippets', and 'Help', along with a search icon, a help icon, and a 'Sign in' button. The breadcrumb path is 'latex > Beamer TPT NG > Commits'. The left sidebar contains navigation icons for home, repository, search, and user profile. The main content area shows a list of commits, grouped by date. Each commit entry includes a commit icon, a title, the author's name and time since authored, a commit hash, and icons for cloning and downloading files.

Navigation: Projects Groups Snippets Help Sign in

Path: latex > Beamer TPT NG > Commits

Filters: master beamer-tpt-ng Author Search by message

01 Jun, 2021 2 commits

- Inkscape's CLI has changed in version 1.0** ...
Tarik Graba authored 1 year ago 2ea63aae
- Makes the poster theme coherent with the presentation** ...
Tarik Graba authored 3 months ago 0f9ef0c9

05 Nov, 2020 1 commit

- Replace -- by \textendash for subsubitems** ...
Tarik Graba authored 9 months ago dfc35f2b

01 Aug, 2020 4 commits

- Merge branch 'master' of gitlab.enst.fr:latex/beamer-tpt-ng**
Cédric Ware authored 1 year ago fad8229b
- Bump release version.**
Cédric Ware authored 1 year ago c7ae4af4
- Fix default font search list.** ...
Cédric Ware authored 1 year ago 271da1d4
- Different conf logo in the poster example to avoid compilation errors**
Tarik Graba authored 1 year ago 76066f04

21 Jul 2020 2 commits

Tools for computer scientists

SSH

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 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

Package python

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

Jupyter Untitled (unsaved changes)



Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted

Python 3

Run Code

```
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
4	0	2	36.7	19.3	193.0	3450.0	0	200
5	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

Jupyter notebooks

Jupyter Untitled (unsaved changes)



Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted

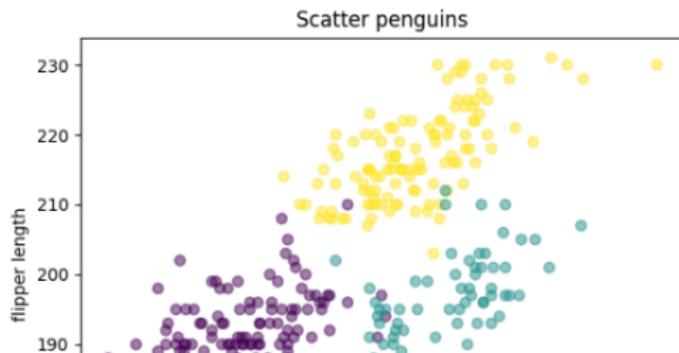
Python 3

Run Code

Plotting bill length in function of flipper length (and show species)

```
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

Installing Jupyter notebooks

Notebooks are packaged in python:

#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

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:

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

Variables should explain what they store:

- ff
- nbFalsePositives
- 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 expects two integers 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

Is it useful for masters?

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

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>

```
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()
```

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

Exercises!

Exercises

- Clone the following repository and create a short presentation:
`https://gitlab.telecom-paris.fr/latex/beamer-tpt-ng`
- Read the pep8 (<https://pep8.org/>) and read and do some of the exercises here <http://web.mit.edu/6.005/www/fa15/classes/04-code-review/>
- Create a repository on Télécom's gitlab (use the sign-in with Shibboleth and use your Télécom's account)
- Install a password manager (e.g. KeepassX)
- Start doing backups
- Make sure your disk is encrypted (if not consider doing it someday)
- Write yourself a résumé using LaTeX (see <https://www.overleaf.com/latex/templates/tagged/cv> for templates)