

DATA AI BASICS

Using a computer as a computer scientist



Louis Jachiet
louis.jachiet@telecom-paris.fr
09 September 2025



IP PARIS

DIGITAL LIFE

Mails – some general and technical considerations

1.1.2

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

1.1.3

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.

1.1.4

THE GOOD

Mail is *somewhat standard* : sending text and files to almost anyone

Mail is *federated* : it does not belong to a single entity

Mail is *asynchronous* : generally instantaneous (but not always!)

1.1.5

THE BAD

Most 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...

1.1.6

THE UGLY

Big mail providers have created an oligopoly : small mails providers often end in spam

Companies are almost spammers : from a communication tool it became the de facto marketing tool

Spam and phishing are rampant : a lot of malicious actors are polluting mails and trying to scam you

1.1.7

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



1.1.8

USE A PROPER SUBJECT!

Keep it relatively short, but informative.

Don't hesitate to modify the subject when replying.

1.1.9

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 angriness (write the email, wait a few days)
- **Don't** use sarcasm

BE SPECIFIC!

Teachers receives dozens if not hundreds of mails a day : they spend 30s reading them sometimes doing something else in parallel (**caution** : made up statistics).

Each mail should be very specific in what you want!

GIVE CONTEXT!

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

It is always better to add slightly too much context!

TOO MUCH INFORMATION

Focus on what is *essential*.

Some advice :

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

TIMELY FOLLOW UP

People (teachers & students) loose tracks of mails, it is Ok.

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.

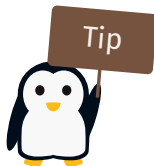
Non urgent : wait at least **one week**, ***urgent*** : go their office

PROOF-READ YOUR MAIL

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

PROOF-READ YOUR MAIL

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



To write a mail :

- Start by putting the attachment,
- then, write the mail
- then add recipients.

DIGITAL LIFE

How to use mail for Data AI

1.2.2

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
 - prefer standardized file types (e.g. PDF, ZIP) not PAGES
 - remove useless files
 - keep size small (if possible)

1.2.3

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

1.2.4

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!

However, lots of important communications will be sent to this address, check it regularly (or make a redirection)!

It comes with a calendar tool, don't hesitate to use such a tool!

DIGITAL LIFE

Security

1.3.2

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?



1.3.3

RÉSUMÉ

Confidentiality : Very low

Could be put on my webpage?

1.3.3

RÉSUMÉ

Confidentiality : Very low

Could be put on my webpage?

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

1.3.3

RÉSUMÉ

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**?

1.3.4

BANK ACCOUNT NUMBER / IDs

Confidentiality : Low

Who can use it?

1.3.4

BANK ACCOUNT NUMBER / IDs

Confidentiality : Low

Who can use it?

- Scammers.

1.3.5

PERSONAL MAIL

Confidentiality : High

What are the risks?

1.3.5

PERSONAL MAIL

Confidentiality : High

What are the risks?

- Scammers

1.3.5

PERSONAL MAIL

Confidentiality : High

What are the risks?

- Scammers
- Extortion

1.3.5

PERSONAL MAIL

Confidentiality : High

What are the risks?

- Scammers
- Extortion
- Intimate mails being leaked

1.3.5

PERSONAL MAIL

Confidentiality : High

What are the risks?

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

1.3.5

PERSONAL MAIL

Confidentiality : High

What are the risks?

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

1.3.6

HOW TO SECURE YOUR DATA ?

Passwords



- A different password for each website : **use a password manager!**
- Strong passwords (you are CS students, compute the entropy!)

Storage devices

- Encrypt all the drives containing data
- Don't use random USB keys

1.3.7

CRAFTING PASSWORDS?

<p>□□□□□□□□□□□□□□ □</p> <p>UNCOMMON (NON-GIBBERISH) BASE WORD</p> <p>ORDER UNKNOWN</p> <p>Tr0ub4dor & 3</p> <p>CAPS? □ COMMON SUBSTITUTIONS □□□ NUMERAL □□□ PUNCTUATION □□□</p> <p>(YOU CAN ADD A FEW MORE BITS TO ACCOUNT FOR THE FACT THAT THIS IS ONLY ONE OF A FEW COMMON FORMATS.)</p>	<p>~28 BITS OF ENTROPY</p> <p>□□□□□□□ □ □□□□□□□ □ □□□ □□□ □□□□ □</p> <p>$2^{28} = 3 \text{ DAYS AT } 1000 \text{ GUESSES/SEC}$</p> <p>(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.)</p> <p>DIFFICULTY TO GUESS: EASY</p>	<p>WAS IT TROMBONE? NO, TROUBADOR. AND ONE OF THE 0s WAS A ZERO?</p> <p>AND THERE WAS SOME SYMBOL...</p>  <p>DIFFICULTY TO REMEMBER: HARD</p>
<p>correct horse battery staple</p> <p>□□□□□ □□□□□ □□□□□ □□□□□</p> <p>FOUR RANDOM COMMON WORDS</p>	<p>~44 BITS OF ENTROPY</p> <p>□□□□□□□□□□ □□□□□□□□□□ □□□□□□□□□□ □□□□□□□□□□</p> <p>$2^{44} = 550 \text{ YEARS AT } 1000 \text{ GUESSES/SEC}$</p> <p>DIFFICULTY TO GUESS: HARD</p>	<p>THAT'S A BATTERY STAPLE.</p> <p>CORRECT!</p>  <p>DIFFICULTY TO REMEMBER: YOU'VE ALREADY MEMORIZED IT</p>

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.

1.3.8

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

1.4.2

WHO IS THE BIGGEST ENEMY OF YOUR DATA?



1.4.2

WHO IS THE BIGGEST ENEMY OF YOUR DATA?



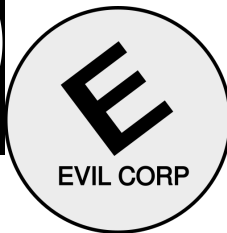
1.4.2

WHO IS THE BIGGEST ENEMY OF YOUR DATA?



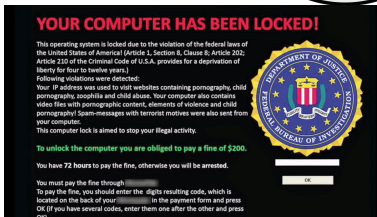
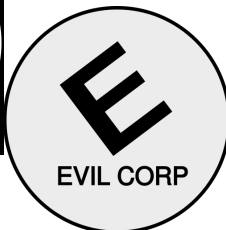
1.4.2

WHO IS THE BIGGEST ENEMY OF YOUR DATA?



1.4.2

WHO IS THE BIGGEST ENEMY OF YOUR DATA?



1.4.2

WHO IS THE BIGGEST ENEMY OF YOUR DATA?



YOU



1.4.3

WHAT CAN HAPPEN TO YOUR DATA?

- **Loosing** 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...



1.4.4

HOW TO BACKUP DATA

Golden rule : 3 times

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

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.

Mental exercise : you loose your laptop/google/usb key/etc. what do you really loose?

DIGITAL LIFE

Online presence

What is online will stay online...

but you can **somewhat** control it by adding content.



1.5.3

CREATING A WEB PAGE

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

Bibliography

2.1.2

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.

2.1.3

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**.



Google Scholar







☒ Toutes les langues ☐ Rechercher les pages en Français


Sur les épaules d'un géant



2.1.5


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

[\[PDF\] 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 ...
☆  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

2.1.6

FINDING ARTICLE WITH GOOGLE SCHOLAR

The screenshot shows the Google Scholar interface with the search term 'fenwick tree' entered in the search bar. The results list three articles. The first article, 'Succinct partial sums and fenwick trees', is by P Bille, AR Christiansen, and N Prezza, published in 2017. The second article, '[PDF] Fenwick Tree and its Application in Solving the Sum of Subsequence', is by D Pan and Y Chen. The third article, 'A new data structure for cumulative frequency tables', is by PM Fenwick, published in 1994. A red hand-drawn circle highlights the '[PDF] psu.edu' link for the third article.

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 ...
☆ 10 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 ...
☆ 2 Cité 2 fois 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 ...
☆ 264 Cité 264 fois Autres articles Les 6 versions



2.1.7


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



[\[PDF\] 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 ...
☆  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



2.1.8


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](#)

[\[PDF\] 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 ...
☆  [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](#)

2.1.9

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 ...

☆ ⓘ 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 ϵ -data structure \mathcal{D} order preserving matching problem. We designed two algorithms for \mathcal{D} respectively. Also ...

☆ ⓘ Cité 2 fois

A scalable asymptotically

[HF Yu](#), [CJ Hsieh](#), [H](#)
[dl.acm.org](#)

... these problems. We modified **Fenwick** trees so that we use an appropriate ...


☆ ⓘ Cité 73 fois

Compact Fenwick

[S Marchini](#), [S Vigna](#)
Summary The **Fenwick** ...

in such a way that the sum and performing ...

☆ ⓘ Cité 2 fois

 Citer

APA	Fenwick, P. M. (1994). A new data structure for cumulative frequency tables. <i>Software: Practice and experience</i> , 24(3), 327-336.
ISO 690	FENWICK, Peter M. A new data structure for cumulative frequency tables. <i>Software: Practice and experience</i> , 1994, vol. 24, no 3, p. 327-336.
MLA	Fenwick, Peter M. "A new data structure for cumulative frequency tables." <i>Software: Practice and experience</i> 24.3 (1994): 327-336.

[BibTeX](#) [EndNote](#) [RefMan](#) [RefWorks](#)

[Partial Sums on the Ultra-Wide Word RAM](#)

[IPDEI arxiv.org](#)

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](#)

[\[PDF\] 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 ...

[☆](#) [🔗](#) [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](#)

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



GETTING ACCESS TO ARTICLES





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_3

url to share this paper:
sci-hub.se/10.1007/978-3-319-66963-2_3

Sci-Hub is fundraising

learn more →

created by: Alexandra Elbakyan

131 (1 of 13)
Automatic Zoom

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 Grancolombiano Institucion Universitaria,
Calle 57 # 3 - 00 Este, i 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 records

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 some **notes**
- A **wiki** or some other **note taking** tool
- A dedicated software, e.g. **Zotero**

The screenshot shows the Zotero application window. On the left is a sidebar with a file explorer view showing folders like 'My Library', 'Book Reviews', 'Colonial Medicine', 'Dissertation', 'Science and Empire', 'Teaching', 'Mapping', 'Open Access', 'Text Mining', 'Visualization', 'My Publications', 'Duplicate Items', and 'Trash'. Below these are 'Group Libraries' and a 'To Read' list. The main pane displays a list of articles with columns for Title, Creator, and Year. The article 'Circulation of Medicine in the Early Modern Atlantic World' by Cook and Walker (2013) is selected. On the right, the 'Info' tab shows details for this article, including its title, authors (Cook, Harold J. and Walker, Timothy D.), abstract, publication information (Social History of Medicine, Volume 26, Issue 3, Pages 337-351, Date 2013/08/01), and various identifiers (DOI, ISSN, URL, etc.).

Title	Creator	Year
Guerre, 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...	Snelders	2013
Medical Experimentation and Race in the Eighteenth-century ...	Schiebinger	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 Nicaraguan...	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 Practic...	Neill	2009
Battles of the Self: War and Subjectivity in Early Modern France	Pichichero	2008
The Experiments of Ramón M. Termeyer 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, Opomania, and Pederasty": Colonial Constructions ...	Proschan	2003
Choosing Scientific Patrimony: Sir Ronald Ross, Alphonse Lav...	Guillemin	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 Co...	Bynum	2000

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 y m d

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

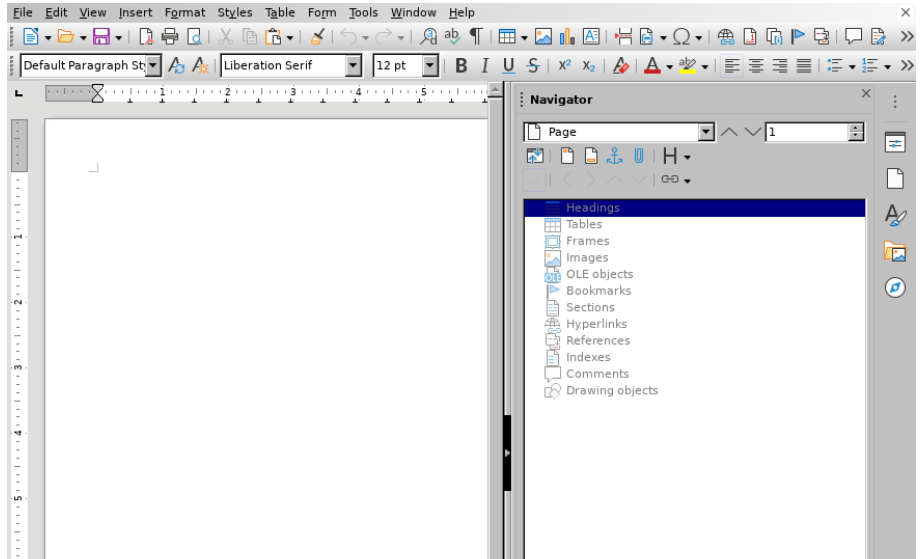
Modified 1/24/2018, 11:50:15 AM

TOOLS FOR COMPUTER SCIENTISTS

Writing research

2.2.2

OFFICE



2.2.3

LaTeX

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

LaTeXcode 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}{recslide.tex}  
                  {\LaTeX code for this slide}
```

```
\end{frame}
```

2.2.4



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

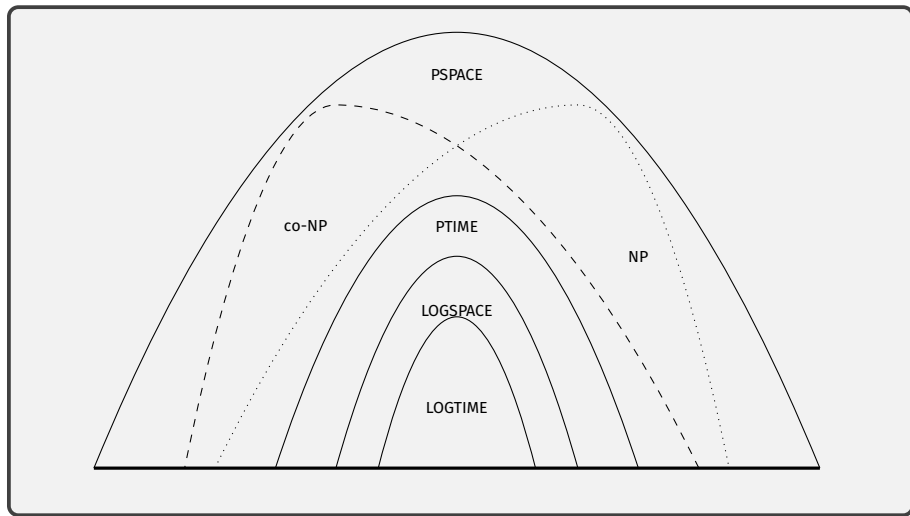
One big advantage of LaTeX is that there exists templates for a multitude of use cases :

- research articles
- presentations
- résumé
- letters
- ...

2.2.6

DRAWING FIGURES WITH TIKZ

Research articles often contains complex figures such as :



2.2.6

DRAWING FIGURES WITH TIKZ

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}
```

Example of Tikz use from texample.net

2.2.7

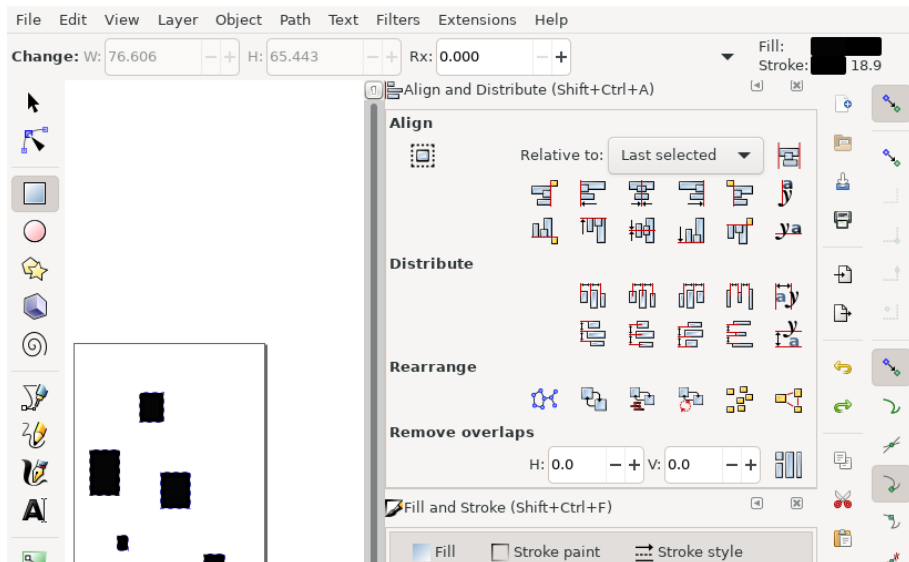
DRAWING FIGURES WITH INKSCAPE

For articles, always prefer **vector** graphics to **raster** graphic :

- lighter
- scalable
- editable

2.2.8

DRAWING FIGURES WITH INKSCAPE



TOOLS FOR COMPUTER SCIENTISTS

Version control systems

2.3.2

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

2.3.2

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
- ...

2.3.3

POPULAR VCS

The most popular ones are :

- **Git**
- Bazaar
- mercurial
- SVN (outdated)
- RCS (very outdated)

2.3.4

AROUND GIT

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
- ...

2.3.5

AROUND GIT

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
- ...

2.3.6

GITHUB EXAMPLE

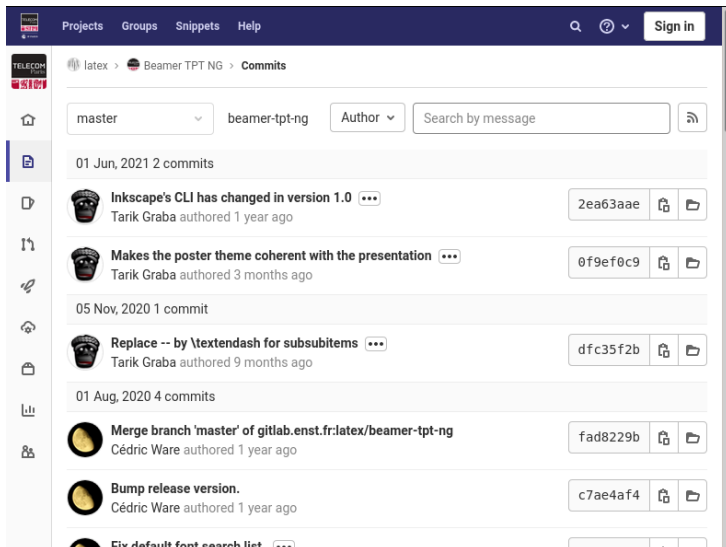
The screenshot shows the GitHub repository page for **GemsLab / KGist**. The repository is in the **master** branch. The main content area displays a list of files and folders, including **data**, **output**, **src**, **test**, **.gitignore**, **README.md**, and **license.txt**. The **README.md** file is selected, showing its commit history. The commit history for **README.md** is as follows:

File	Commit Message	Time Ago
data	added labels file to dbpedia.zip	15 months ago
output	updated documentation	2 years ago
src	add idify 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

The **About** section on the right provides a brief description of the project: "Knowledge Graph summarization for anomaly/error detection & completion (WebConf '20)". It also includes a link to the project's website: gemslab.github.io/papers/.... Below the link, there are several tags: **dbpedia**, **knowledge-graph**, **mdl**, **knowledge-graph-completion**, **rule-mining**, **minimum-description-length**, **webconf**, **nell**, **www2020**, and **graph-summarization**.

2.3.7

GITLAB EXAMPLE



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 to another computer

2.4.3

SSH

SSH has many uses :

- Use computers from afar

reserved for power users

- Run computation

useful to run computation in e.g. lab rooms

- Move files

useful e.g. between the school system and your computer

- Protocol of connection (git)

GIT can use ssh to “push” your changes

- Tunnel internet (much like a VPN)

Access pay-walled articles from home

3.1

RECENT EVOLUTION OF AI / LLM

GPT 3 released in 2020 and current models are much better, surely
General AI is around the corner!

3.1

RECENT EVOLUTION OF AI / LLM

GPT 3 released in 2020 and current models are much better, surely General AI is around the corner!

We don't know :

- they could plateau (i.e. not get any better than they are now)
 - they could slightly improve (but stay at roughly the current level)
 - they could improve a lot (but not a GAI)
 - GAI could be there tomorrow... or in 100s of years!
-

What is sure : models and the tooling around them are changing quickly.

3.2

MY TAKE ON AI AND GAI

We don't know what the future holds, therefore :

- there is no need to become an expert in using current systems

Do prompt engineer have a future?

- it is very likely that what we teach you will be useful **You know that $2*3=6$ even if we have calculators**
- you can use AI in class but try to learn how to do things without AI

My take : AI use might be positive if it helps you to learn not if it replaces learning!

3.2

MY TAKE ON AI AND GAI

We don't know what the future holds, therefore :

- there is no need to become an expert in using current systems
Do prompt engineer have a future?
- it is very likely that what we teach you will be useful **You know that $2*3=6$ even if we have calculators**
- you can use AI in class but try to learn how to do things without AI

My take : AI use might be positive if it helps you to learn not if it replaces learning!

Remark : For classes, using AI is like copy-pasting from any source : you have to reference it!

3.3

FURTHER POINTS ON AI

You are studying AI remember that AI has a lot of **ethical considerations :**

- ecological impacts
- jobs disruption
- cognitive development
- fake news / spam / *etc.*
- ...

3.3

FURTHER POINTS ON AI

You are studying AI remember that AI has a lot of **ethical considerations :**

- ecological impacts
- jobs disruption
- cognitive development
- fake news / spam / *etc.*
- ...

and of course some positive aspects :

- better science / health / education / *etc.*
- replacing human work
- ...

3.3

FURTHER POINTS ON AI

You are studying AI remember that AI has a lot of **ethical considerations :**

- ecological impacts
- jobs disruption
- cognitive development
- fake news / spam / *etc.*
- ...

and of course some positive aspects :

- better science / health / education / *etc.*
- replacing human work
- ...

More consideration on AI ethics in the AI ethics class :)

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

4.2.2

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

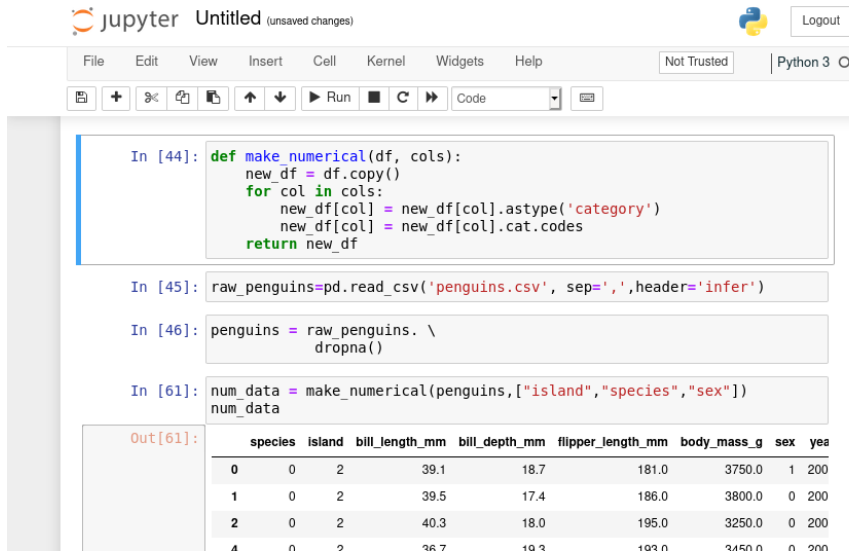
# using the virtualenv
source nameOfTheVirtualenv/bin/activate #activation
which pip #check which pip we are using
pip install package name # install package in the venv
python # run python in the venv
jupyter-notebook # run jupyter equipped with the venv
```

GOOD PRACTICES FOR PROGRAMMING

Notebooks

4.3.2

JUPYTER NOTEBOOKS



The screenshot shows a Jupyter Notebook interface with the title "Untitled (unsaved changes)". The top bar includes a menu (File, Edit, View, Insert, Cell, Kernel, Widgets, Help), a "Not Trusted" status, and a "Python 3" kernel selection. The toolbar contains icons for saving, adding cells, undo, redo, running, and other actions.

The notebook contains the following code cells:

```
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
```

The output of the last cell (In [61]) is a DataFrame:












	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

4.3.3

JUPYTER NOTEBOOKS

 jupyter Untitled (unsaved changes)  Logout

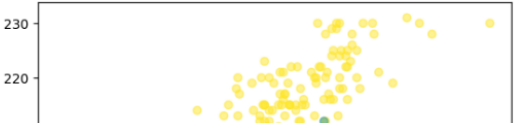
File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3

       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>



4.3.4

JUPYTER NOTEBOOKS

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

4.3.5

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

4.4.2

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.

4.4.3

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

CODE FORMATTING

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

4.4.5

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...

4.4.6

NAMING FUNCTIONS / VARIABLES

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

❌ doStuff

4.4.6

NAMING FUNCTIONS / VARIABLES

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

 doStuff

 ComputeVAT

4.4.6

NAMING FUNCTIONS / VARIABLES

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

- ☐ doStuff
- ☒ ComputeVAT
- ☐ SumMultiplyCategoryToPrices

4.4.6

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

4.4.6

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

4.4.6

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)

4.4.6

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)
- ☒ cleanAbort

4.4.7

COMMENTS

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

COMMENTS

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

4.4.8

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 Each unit test tests some very specific piece of code to make sure it works ok.

e.g. test a function or a member of a class

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

perfect to test algorithms

Regression tests Each time a bug is discovered, introduce tests to make sure this bug will not reappear.

4.5.3

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

5.1

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 e.g. using \LaTeX (see <https://www.overleaf.com/latex/templates/tagged/cv> for templates)

5.2

THIS AFTERNOON!

Logic is used in :

- Maths
- Electronics
- Computer science (theorem proving, proofs of programs)
- Knowledge Representation & semantic Web
- high-level NLP, Law, explanation, argumentation
- Database management

Logic does not hallucinate



Louis Jachiet

09 September 2025

louis.jachiet@telecom-paris.fr