The global warming of the Web

Measuring the environmental footprint of HTTP requests (EcoIndex)

christophe.cerin@univ-grenoble-alpes.fr

March 27, 2023 – GreenDays

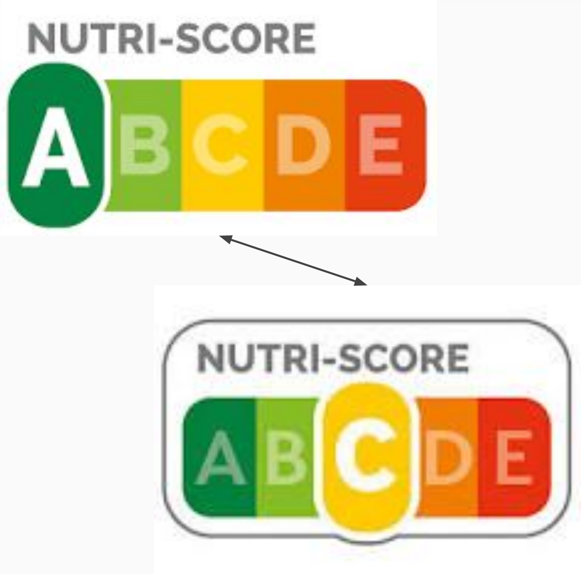
1. Problem statement

Focus on HTTP requests (not really the Web)

Can you quantify the evolution of the environmental footprint of HTTP requests (increase or decrease)?

You need:

- A metric: URL → score;
- A dataset of URLs.



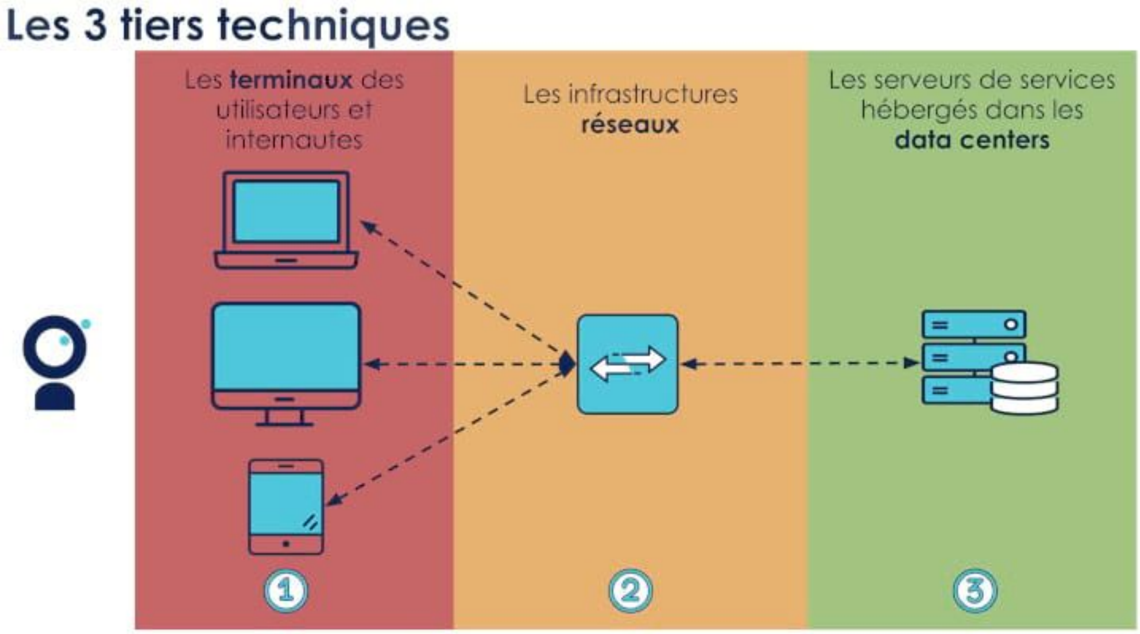
Programme national nutrition santé (PNNS), présidé par Serge Hercberg,

In the EcoIndex, it was chosen to represent:

An endpoint ⇒ by the number of DOM elements;

The network ⇒ by the "weight" of the page (KB returned) ;

The servers ⇒ by the number of http/https requests in the returned page.



Fundamental equation

$$Ecoindex = 100 - \frac{5 * [3 * Fd(Taille_{DOM}) + 1 * Fs(Poids_{Page}) + 2 * Fr(Nb_{Requêtes})]}{6}$$

Key points:

1. Normalization;
2. Weights 3, 2, and 1 based on macro studies;
3. Fd, Fs, and Fr – return ranks (from 0 to 20) in quantiles vectors. Check with https://github.com/cnumr/GreenIT-Analysis/blob/master/script/ecoIndex.js

EcoIndex is also…

1. A set of best practices;
2. A score from A to G;
3. A score for gas emission;
4. A score for water consumption.

```
function getEcoIndexGrade(ecoIndex)
{
if (ecoIndex > 80) return "A";
if (ecoIndex > 70) return "B";
if (ecoIndex > 55) return "C";
if (ecoIndex > 40) return "D";
if (ecoIndex > 25) return "E";
if (ecoIndex > 10) return "F";
return "G";
}

function computeGreenhouseGasesEmissionfromEcoIndex(ecoIndex)
{
        return (2 + 2 * (50 - ecoIndex) / 100).toFixed(2);
}

function computeWaterConsumptionfromEcoIndex(ecoIndex)
{
        return (3 + 3 * (50 - ecoIndex) / 100).toFixed(2);
}
```

2. Criticisms and new approaches

Limitations inherent to the 3-tier model

BAD – EcoIndex does not consider:

- environmental impact of the computer making the query or of a user browsing;
- environmental impact of servers in the classical sense of life cycle assessments (LCA), nor of the different network types of equipment.

GOOD –

- discussion of models and their attributes that would meaningfully characterize the environmental impact of digital at the HTTP requests level;
- loading, creation, and display of the page in the browser are not simulated.

Limitations inherent to the calculation itself

QUESTIONABLE –

- Weights (3,2,1) are static: they do not consider variations over time or country, and these are limits for the model that we can describe as temporal and geographical;
- Stability over time of quantiles? A priori, websites are regularly reviewed to adopt, over time, the best eco-design practices, and therefore there is no reason for quantiles to be seen as a static quantities;
- Websites are dynamic;
- The A-G scores correspond to the EcoIndex ranges of 100-81 for A and 10-0 for G. How were these different bounds determined? Do they fit the quantiles for the HTTParchive EcoIndex measures? This is still being determined.

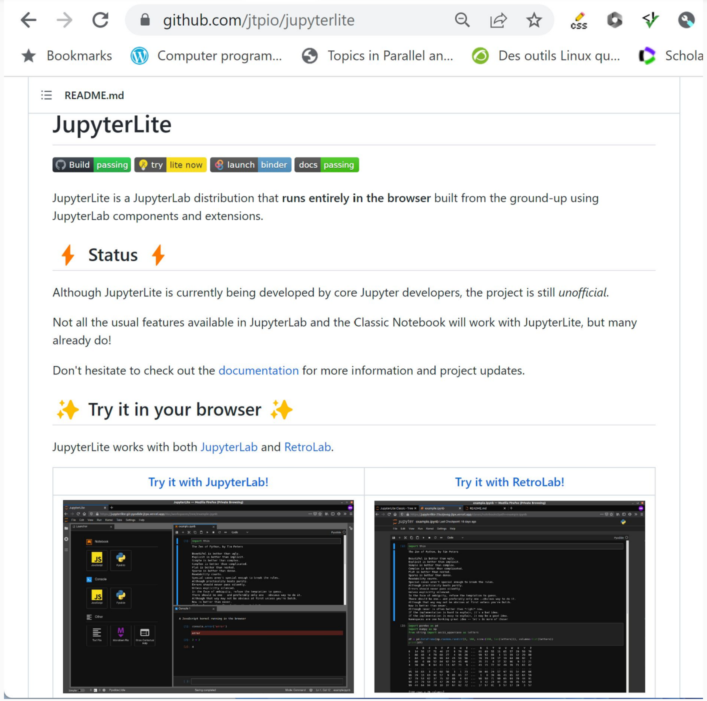
Limitations inherent in the attributes that make sense

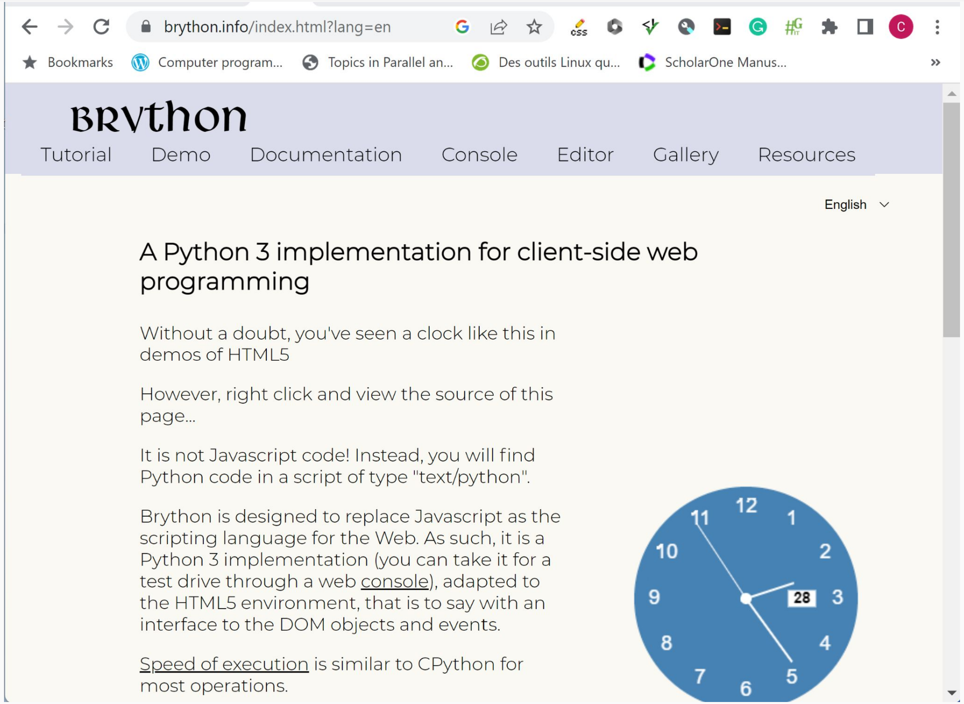
How to introduce new attributes into the model to capture:

- Energy mix;
- HTTP request goes through a 4/5G network or fiber;
- aggregate the CO_2 impact of the operator;
- …

A potential of 10, 100, … attributes, all related to “energy”. How to deal with a large number of attributes?

Limits inherent to the potential use of other forms of interaction models

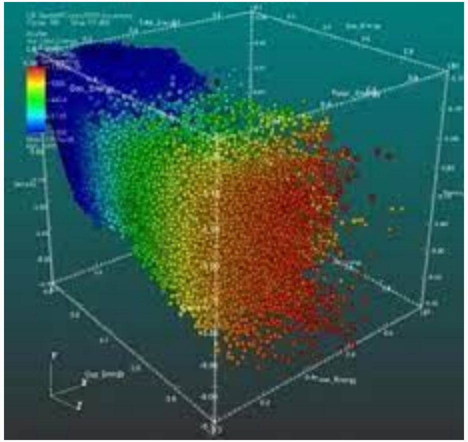




Towards new or extended definitions

Bypass the limits: (some new directions classified into two directions)

1. Forget weights and quantiles;
2. Consider, for a URL, N$\gg$1 attributes related to direct or indirect energy consumption:
   a. cluster URL (EcoIndex relates to the neighbors): relative measure;
   b. place/order URLs on a line: absolute measure.



Why?

- Observe, classify, and query URLs to better understand their similarities. Assume that N$\gg$1 attributes fully characterize the "System."

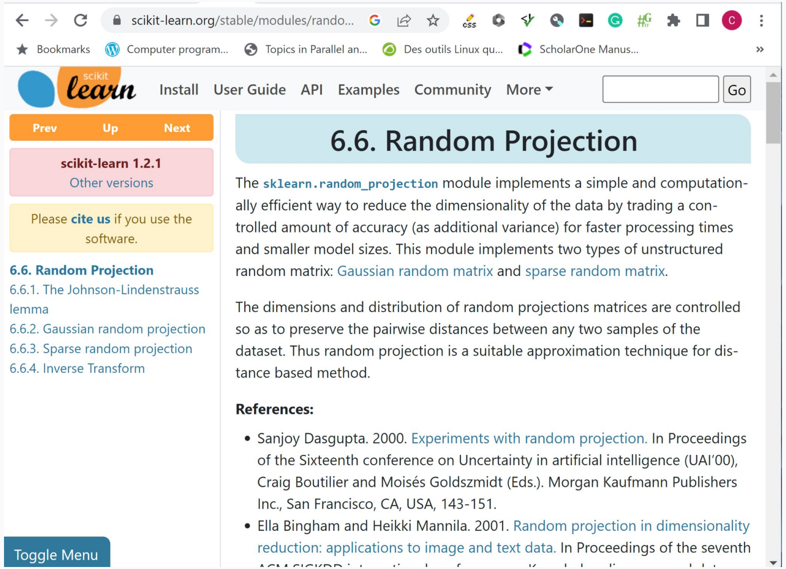
ML techniques help to deal with the explosion of attributes

Random Projection: a lemma states that if points in a vector space are of sufficiently high dimension, then they may be projected into a suitable lower-dimensional space in a way which approximately preserves the distances between the points.

Make a projection on the line and between 0 and 100 to bypass Fd **and** Fs **and** Fr

The “new EcoIndex” is the projection!

Note: *We still reduce to a single value.*



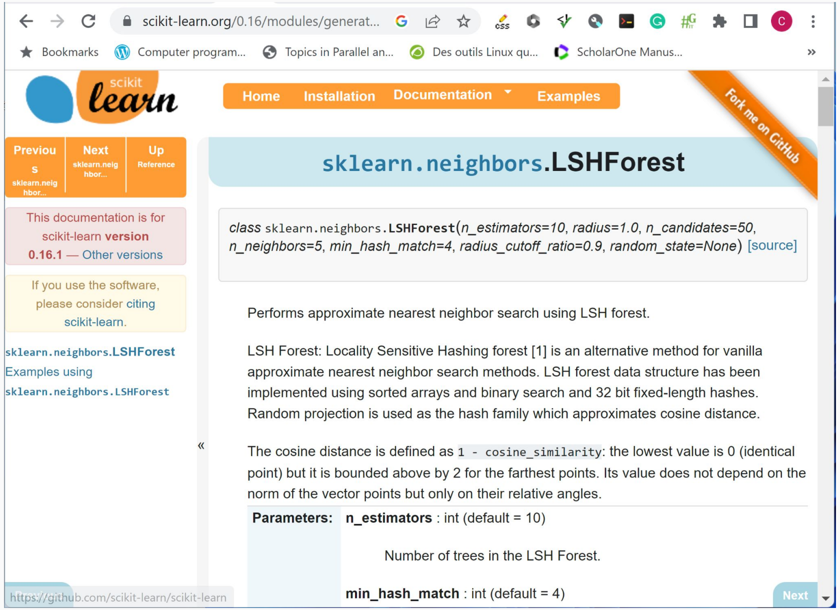
ML techniques help to deal with the explosion of attributes

Locality-sensitive hashing (LSH) is an algorithmic technique that hashes similar input items into the same "buckets" with high probability.

The “new EcoIndex” is the rank of the bucket!

One of the main applications of LSH is to provide a method for efficient approximate nearest neighbor search algorithms.

Note: *We still reduce to a single value.*

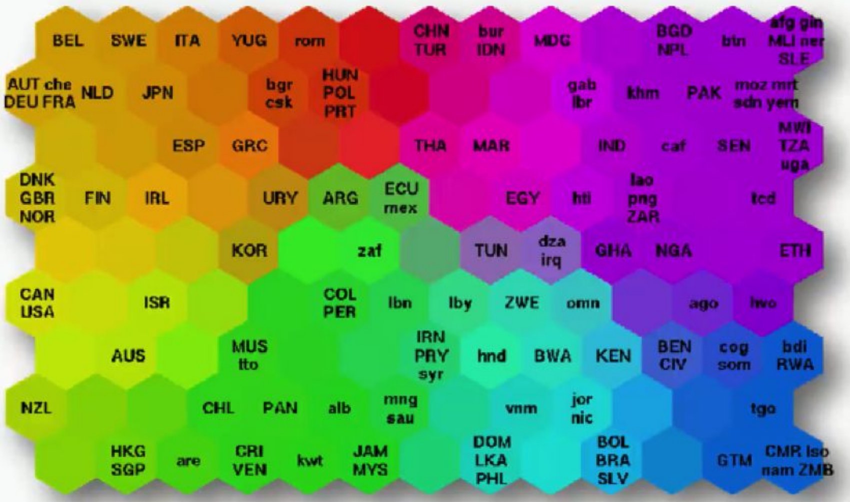


ML techniques help to deal with the explosion of attributes

A **self-organizing map (SOM)** is an unsupervised machine learning technique used to produce a low-dimensional (typically two-dimensional) representation of a higher dimensional data set while preserving the topological structure of the data.

A **set** of URLs corresponds to a SOM.

Good: position of one URL in the map (color) gives my neighbors - clustering (for free)



***Note:** We get a map not a single value.*

Self Organizing Map (Open Data exemple)

=== ARCEP data 2022_QoS_Metropole_data_habitations.csv ===

Column names of ARCEP data:

- lieu
- situation
- date
- heure
- operateur
- Profil
- rsrp
- latitude
- longitude
- protocole
- url
- file_name

=== ENEDIS data consommation-electrique-par-secteur-dactivite-commune.csv ===

Column names of ENEDIS (consommation):

- Année
- Code Commune
- Nom Commune
- ….
- CODE GRAND SECTEUR
- CODE SECTEUR NAF2
- Nb sites
- Conso totale (MWh)
- Conso moyenne (MWh)
- …..

=== ENEDIS data production-electrique-par-filiere-a-la-maille-commune.csv ===

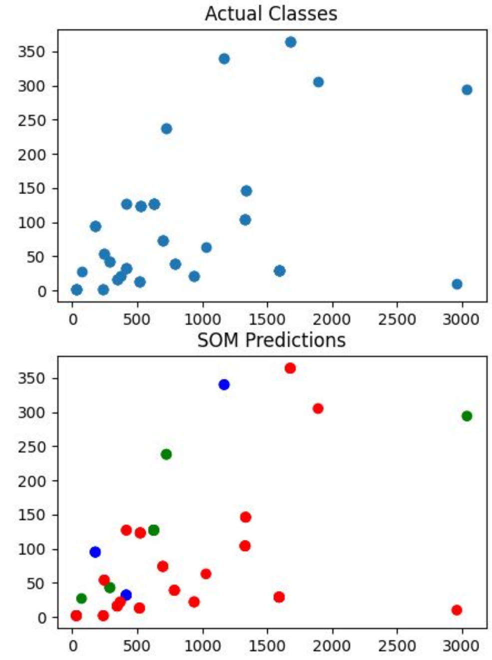
Column names of ENEDIS data (production):

- Année
- Nom commune
- Code commune
- Energie produite annuelle Photovoltaïque Enedis (MWh)
- Energie produite annuelle Eolien Enedis (MWh)
- Energie produite annuelle Hydraulique Enedis (MWh)
- Energie produite annuelle Bio Energie Enedis (MWh)
- Energie produite annuelle Cogénération Enedis (MWh)
- Energie produite annuelle Autres filières

Self Organizing Map (exemple for 2021)

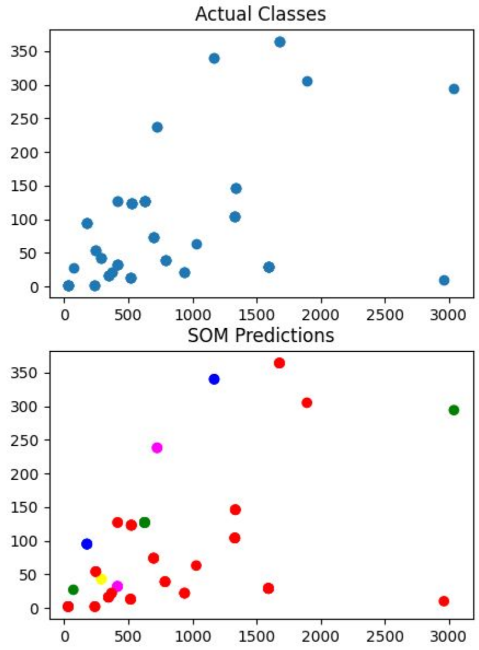
ORANGE ; 46.47044 ; -1.02718 ; 85121 ; FONTENAY-LE-COMTE ; https://www.allocine.fr/ ; 1593 ; 29 ; 1921444 ; 40.78 ; 2.18 ; 3.28 ; 1.0 ; 2.9 ; 3651.1647229005457 ; 6.799189428120197 ; 151.0991898335209 ; 0.0 ; 0.0 ; 0.0 ; 0.0 ; 0.0
BOUYGUES TELECOM ; 42.69109 ; 2.92053 ; 66028 ; PERPIGNAN ; ; 1.0 ; 4.25 ; 27254.2314006874 ; 5.425887199021979 ; 437.42442519542254 ; 0.0 ; 0.0 ; 0.0 ; 0.0 ; 0.0
FREE ; 50.40982 ; 2.96885 ; 62427 ; LENS ; ; nan ; 10 ; 54767.62509277385 ; 4.428171498445493 ; 167.0849727571493 ; 0.0 ; 0.0 ; 0.0 ; 0.0 ; 0.0
ORANGE ; 42.71066 ; 2.88409 ; 66136 ; PERPIGNAN ; https://www.instagram.com/accounts/login/ ; 178 ; 95 ; 29041638 ; 55.95 ; 1.88 ; 2.82 ; 1.0 ; 3.46 ; 249164.689899692 ; 3.4576912601780707 ; 2704.200723216868 ; 0.0 ; 0.0 ; 0.0 ; 0.0 ; 0.0
FREE ; 42.71066 ; 2.88409 ; 66136 ; PERPIGNAN ; https://www.jeuxvideo.com/ ; 2961 ; 10 ; 974937 ; 43.90 ; 2.12 ; 3.18 ; 1.0 ; 5.25 ; 249164.689899692 ; 3.4576912601780707 ; 2704.200723216868 ; 0.0 ; 0.0 ; 0.0 ; 0.0 ; 0.0
ORANGE ; 47.28475 ; -1.55131 ; 44201 ; NANTES ; https://fr.m.wikipedia.org/wiki/Wikip%C3%A9dia:Accueil_principal ; 1677 ; 364 ; 355418 ; 21.56 ; 2.57 ; 3.85 ; 1.0 ; 2.1 ; 22694.704171610498 ; 7.1254958152623225 ; 115.771 ; 0.0 ; 0.0 ; 0.0 ; 0.0 ; 0.0
FREE ; 50.43486 ; 3.05734 ; 62497 ; LENS ; https://www.allocine.fr/ ; 1593 ; 29 ; 1921751 ; 40.78 ; 2.18 ; 3.28 ; 1.0 ; 2.68 ; 13576.048222272546 ; 4.43082513781741 ; 0.0 ; 0.0 ; 0.0 ; 429.376 ; 0.0 ; 0.0
SFR ; 46.15188 ; -1.14069 ; 17300 ; LA ROCHELLE ; https://www.service-public.fr/ ; 524 ; 123 ; 1115732 ; 47.31 ; 2.05 ; 3.08 ; 1.0 ; 3.3 ; 174586.3200622886 ; 3.279724978627303 ; 4028.48056147116 ; 0.0 ; 0.0 ; 0.0 ; 0.0 ; 0.0

Self Organizing Map (exemple for 2021)



3 clusters

5 clusters



Interpretation seems difficult!

Scientific question: how such map can help in defining a 'weather situation'?

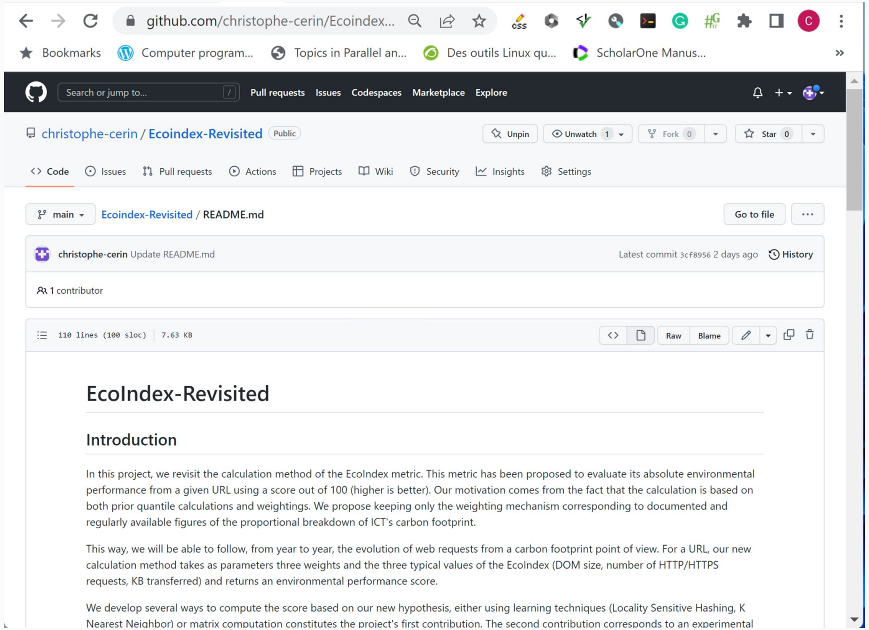
3. Conclusion

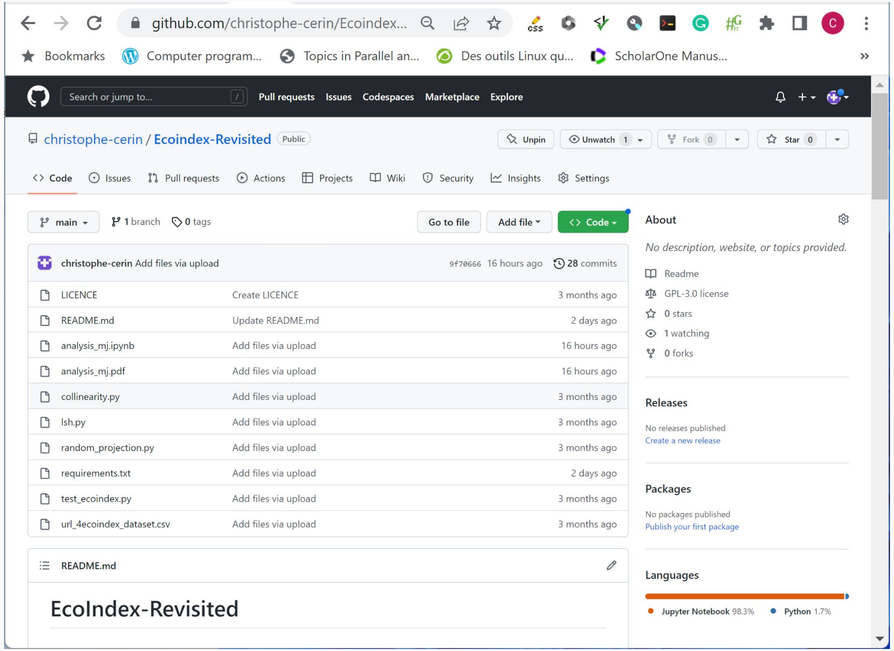
Future directions

- The EcoIndex: a good starting point understanding of the phenomena and issues related to the role of digital technology in global warming;
- We need to understand better the relationships between the different high-level models of 3-tier architecture, server or client, and the field analysis of the life cycle of a digital product or equipment (LCA) or the field of ”best practices”;
- Introducing many attributes in environmental impact models: **immerse the problem in a machine learning formulation ⇆ weather of web requests** ;
- Context of the "client-side web programming" vision.

Contribute:

https://github.com/christophe-cerin/Ecoindex-Revisited





Thank you…
and mind your HTTP requests!

Special thanks: Mathilde, Laurent, and Denis

christophe.cerin@univ-paris13.fr
http://lipn.univ-paris13.fr/~cerin/