L'Institut Des Humanités Numériques : outils et ressources

École d'hiver de l'IDHN - L'IA pour les SHS, les SHS pour l'IA : regards croisés

Mardi 16 décembre 2025

Maison de la recherche SHS Annie Ernaux, Cergy, France

Dimitra Niaouri
dimitra.niaouri@ensea.fr

Marine Delaborde
marine.delaborde@cyu.fr

Mathias Quoy
mathias.quoy@cyu.fr

IDHN

institut des humanités numériques

IDHN

- Rôle de l'IDHN à CY Cergy Paris Université
  - **Fédération** de la recherche en HN à CY
  - Environnement propice aux **collaborations**
  - **Partage de méthodes** et diffusion d'une culture scientifique numérique
  - Accompagnement et mise à disposition de **ressources**
- **Perspectives** :
  - Développement de la formation et montée en compétences (liens avec consortium)
  - Besoins de puissance de calcul
    - Investissement dans de nouveaux serveurs grâce à deux projets LT2D

L'IDHN depuis le 1er juillet 2025

5 laboratoires

AGORA

EMA
école, mutations, apprentissages

ETIS
Équipes Traitement de l'Information et Systèmes

LT2D
lexiques, textes, discours, dictionnaires

PLACES
laboratoire de géographie et d'aménagement

Conseil de direction

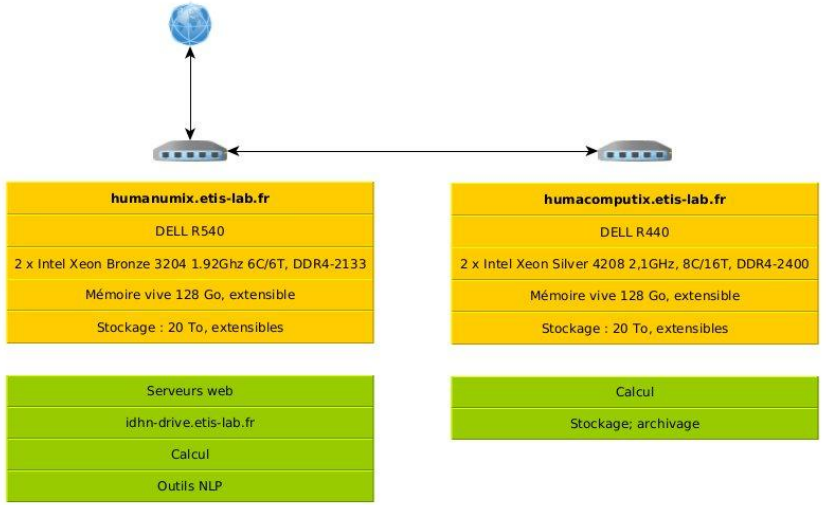
- **Mathias QUOY** (ETIS) - directeur
- **Marine DELABORDE** (LT2D) - directrice adjointe
- **Marie-Stéphanie ABOUNA** (AGORA)
- **Boris BORZIC** (pôle ingénierie ETIS)
- **Sonia CASTAGNET** (EMA)
- **Basile MICHEL** (PLACE)

Ressources

Matériel et logiciels

- **Matériel** :
  - **serveurs** :
    - Humanumix (stockage)
    - Humacomputix (calculs)
  - **ordinateurs fixes**

- **Outils** :
  - R / RStudio
  - IRaMuTeQ
  - QGIS
  - NVivo
  - Emotaix (Tropes)
  - PyDoct / Tesseract OCR
  - INCEpTION
  - TXM



Aide à l'utilisation des outils et ressources

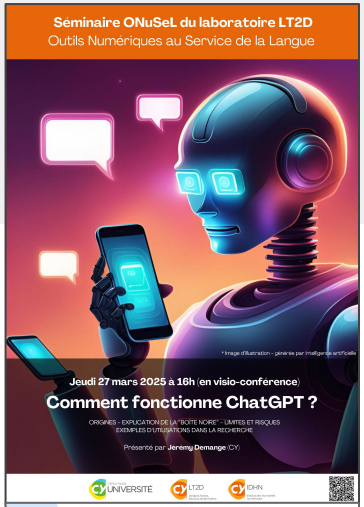
- **Objectif initial** : mutualiser l'effort de recrutement selon les compétences recherchées
  - Financement par le DIM : ingénieurs HN (Zakarya Despré, Jeremy Demange)
    - Développement d'outils pour des projets de recherche liés au DIM et à l'IDHN
      - API de textométrie
      - Modèles de détection de sentiments et d'ironie
      - Tests de culture politique
      - Interface graphique pour application de modèles de détection de la coréférence
      - VIKY.ai : partenariat de recherche avec la société Pertimm pour le développement d'une plateforme multi-agents pour le traitement de données linguistiques
      - https://github.com/cyidhn

Aide à l’utilisation des outils et ressources

- À l’heure actuelle (2025/2026) :
  - **Mission doctorale hors enseignement** : 32 jours (Dimitra Niaouri)
    - Production de documentation : rédaction de fiches pratiques et de tutoriels
    - Veille : suivi de l'évolution des outils et méthodes en HN
    - Accompagnement ponctuel personnalisé :
      - entretiens avec les chercheurs, conseils sur les outils adaptés, aide à la mise en application
      - Réalisation de rapports synthétiques des accompagnements
  - Contactez-nous si vous avez besoin d’un outil ou d’un accompagnement :
    - mathias.quoy@cyu.fr
    - marine.delaborde@cyu.fr

Découverte d'outils et de ressources

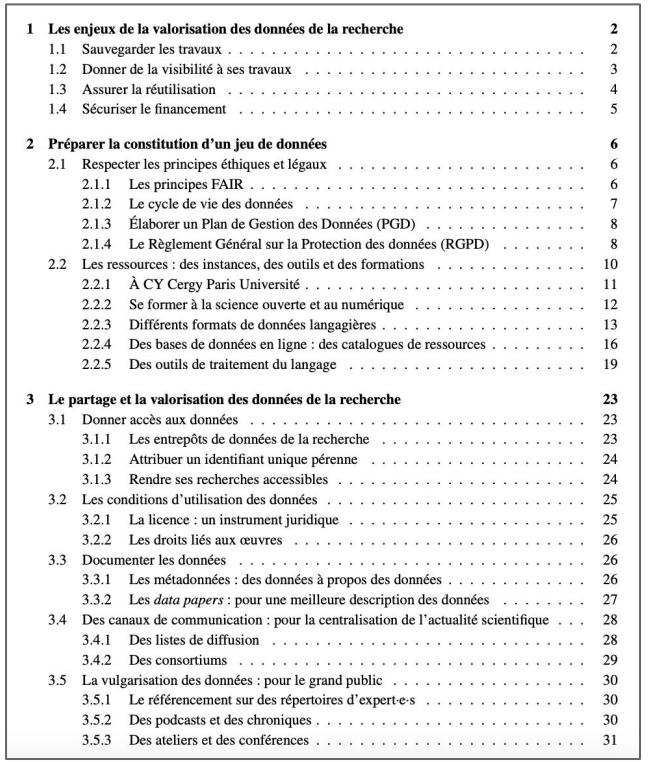
- Actions actuelles :
  - Séminaires
  - Conférences
  - Écoles thématiques
- En discussion → formations :
  - Ajout du séminaire **ONuSeL** (Outils Numériques au Service de la Langue) au Catalogue des formations de l'**ED AHSS Arts**, Humanités, Sciences Sociales
  - Organisation de **formations** - en collaboration avec des consortiums **Huma-Num** ou non



Ressources

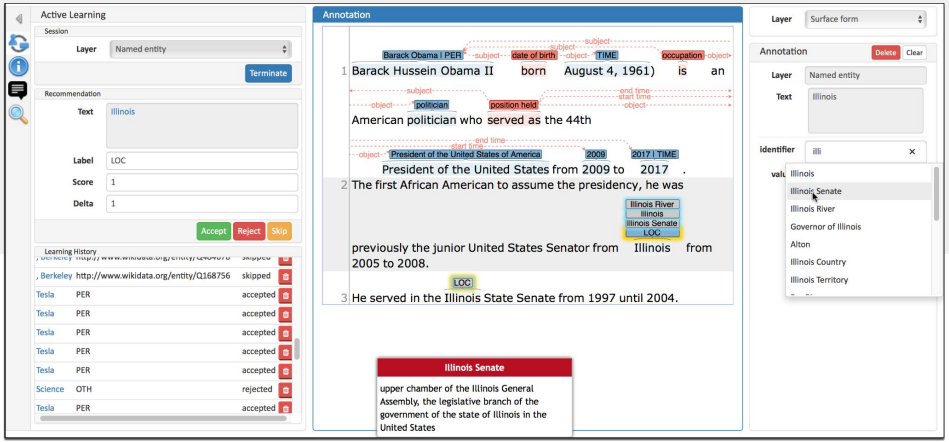
Méthodologie

- Guide de valorisation des données au LT2D :
  - **Faire le lien avec les chercheurs**
  - Cadre : CPJ « Ressources numériques en SHS »
    - Ojectif 1 : « Valorisation des données au LT2D »
  - Mission IGE : Thomas Gervais d'Aldin (2024)
  - **Données langagières** écrites ou orales (Lexiques, Textes, Discours, Dictionnaires)
  - Objectif : veille et amélioration régulière
    - Données issues des RSN (CER + DPO)
  - Séminaire ONuSeL en 2026



- MVD : Référencement d'ouvrages lexicographiques
  - Passage d'un site statique à une base de données : réalisé par Luc Massip (LT2D)
- Omeka S : Plateforme de publication Web
  - gestion de fonds documentaires
  - système de publication multi-sites
  - gestion des métadonnées pour décrire les ressources
  - utilisation de vocabulaires du web sémantique + standard IIIF
  - modules (de connectivité)
- Sur serveur IDHN : https://dictionnaires-lt2d.idhn.etis-lab.fr/
- Contact : marine.delaborde@cyu.fr

INCEpTION

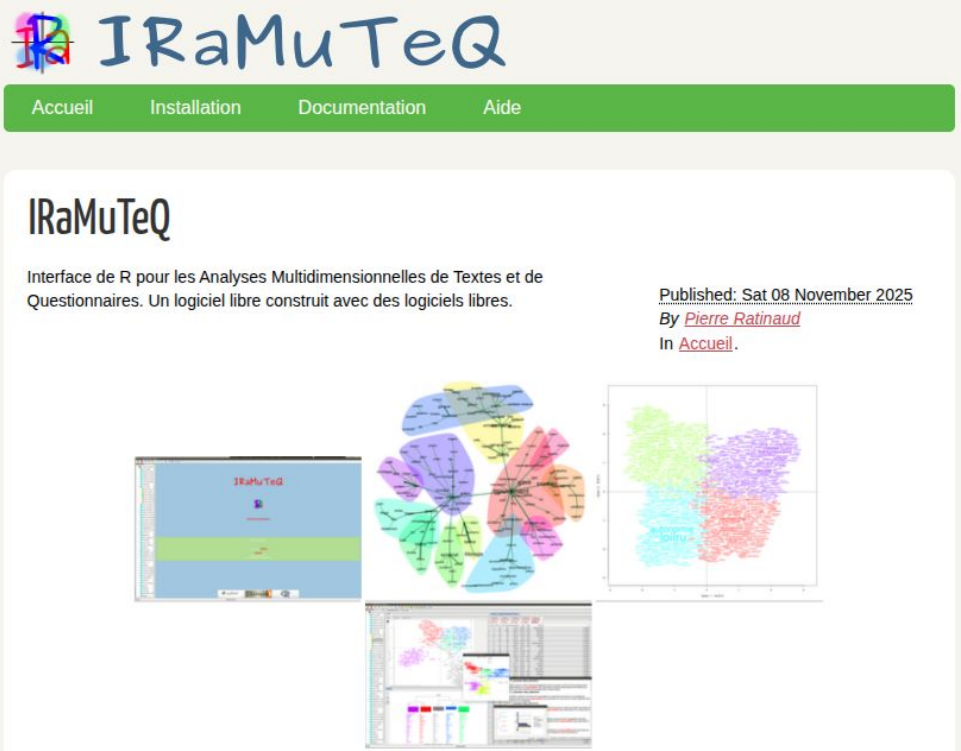


source : https://inception-project.github.io/

- Annotation multicouche
- Approche collaborative
  - gestion de cohortes d'annotateurs (rôles)
  - attribution de textes
  - adjudication et calcul d'accord inter-annotateur
- Active learning
- Serveur CORLI (consortium Huma-Num) hébergé par l'ATILF : https://corli.huma-num.fr/fr/inception/
- Serveur IDHN : https://inception.idhn.etis-lab.fr/
- Contact : marine.delaborde@cyu.fr

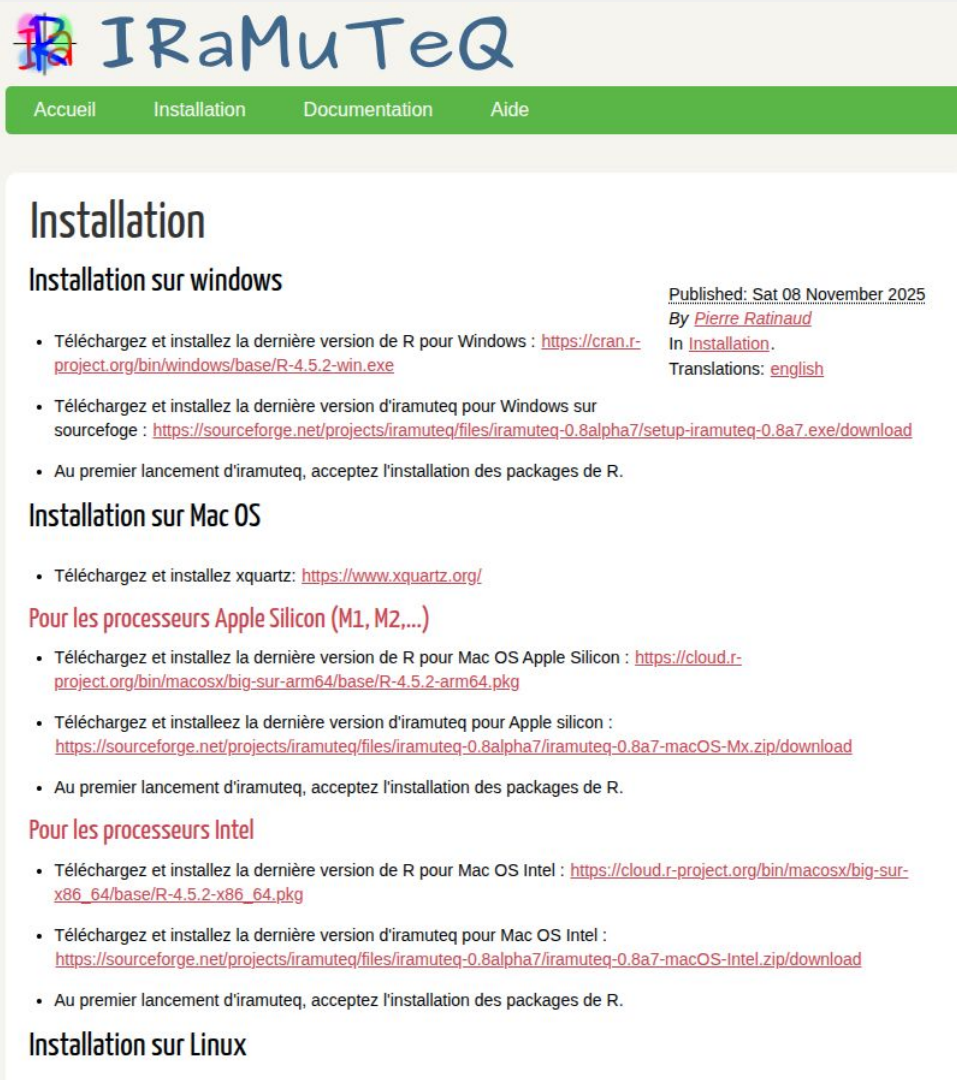
IRaMuTeQ

- Open-source software for statistical text analysis.
- It was created by French researcher Pierre Ratinaud at LERASS (Laboratoire d'Études et de Recherches Appliquées en Sciences Sociales).
- Based on **R** and **Python**.
- **Webpage:** https://pratinaud.gitpages.huma-num.fr/iramuteq-website/



IRaMuTeQ

Installation guidelines:
https://pratinaud.gitpages.huma-num.fr/iramuteq-website/category/installation.html



IRaMuTeQ

Documentation:
https://pratinaud.gitpages.huma-num.fr/iramuteq-website/category/installation.html

Preparing IRaMuTeQ input files:
https://pratinaud.gitpages.huma-num.fr/iramuteq-website/documentation/DataFilePreparation.pdf

IRaMuTeQ

Accueil Installation Documentation Aide

Installation

Installation sur windows

Published: Sat 08 November 2025
By Pierre Ratinaud
In Installation.
Translations: english

- Téléchargez et installez la dernière version de R pour Windows : https://cran.r-project.org/bin/windows/base/R-4.5.2-win.exe
- Téléchargez et installez la dernière version d'iramuteq pour Windows sur sourcefoge : https://sourceforge.net/projects/iramuteq/files/iramuteq-0.8alpha7/setup-iramuteq-0.8a7.exe/download
- Au premier lancement d'iramuteq, acceptez l'installation des packages de R.

Installation sur Mac OS

- Téléchargez et installez xquartz: https://www.xquartz.org/

Pour les processeurs Apple Silicon (M1, M2,...)

- Téléchargez et installez la dernière version de R pour Mac OS Apple Silicon : https://cloud.r-project.org/bin/macosx/big-sur-arm64/base/R-4.5.2-arm64.pkg
- Téléchargez et installeez la dernière version d'iramuteq pour Apple silicon : https://sourceforge.net/projects/iramuteq/files/iramuteq-0.8alpha7/iramuteq-0.8a7-macOS-Mx.zip/download
- Au premier lancement d'iramuteq, acceptez l'installation des packages de R.

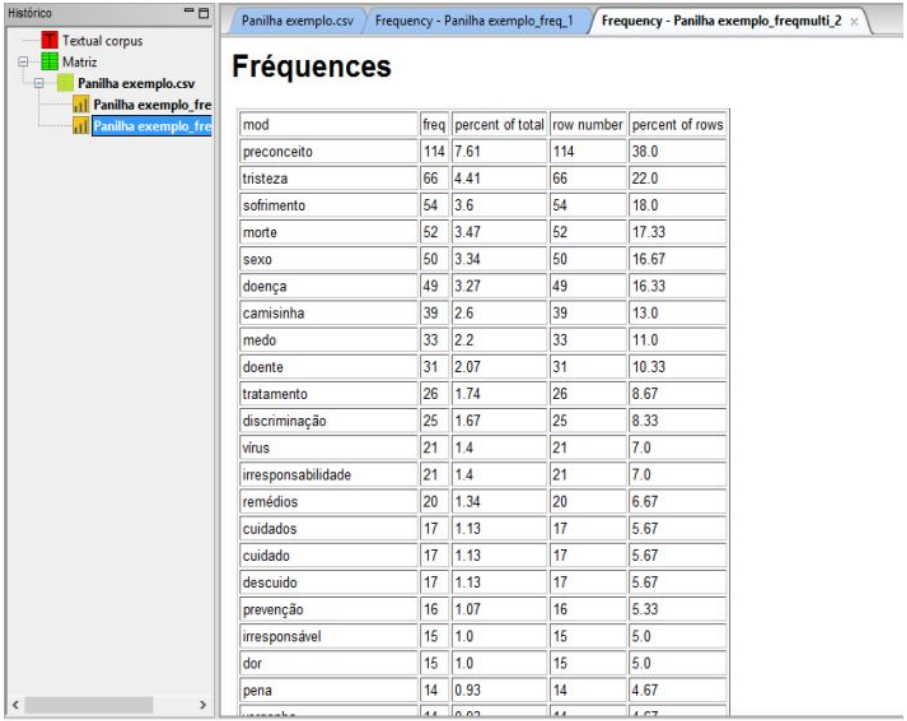
Pour les processeurs Intel

- Téléchargez et installez la dernière version de R pour Mac OS Intel : https://cloud.r-project.org/bin/macosx/big-sur-x86_64/base/R-4.5.2-x86_64.pkg
- Téléchargez et installez la dernière version d'iramuteq pour Mac OS Intel : https://sourceforge.net/projects/iramuteq/files/iramuteq-0.8alpha7/iramuteq-0.8a7-macOS-Intel.zip/download
- Au premier lancement d'iramuteq, acceptez l'installation des packages de R.

Installation sur Linux

IRaMuTeQ

1. **Classic Lexicographical Analysis**
- Counts word frequencies.
- Identifies rare words (hapax).
- Creates reduced word forms (lemmatization).
- Builds a dictionary of reduced forms.
- Useful for counting and interpreting data.

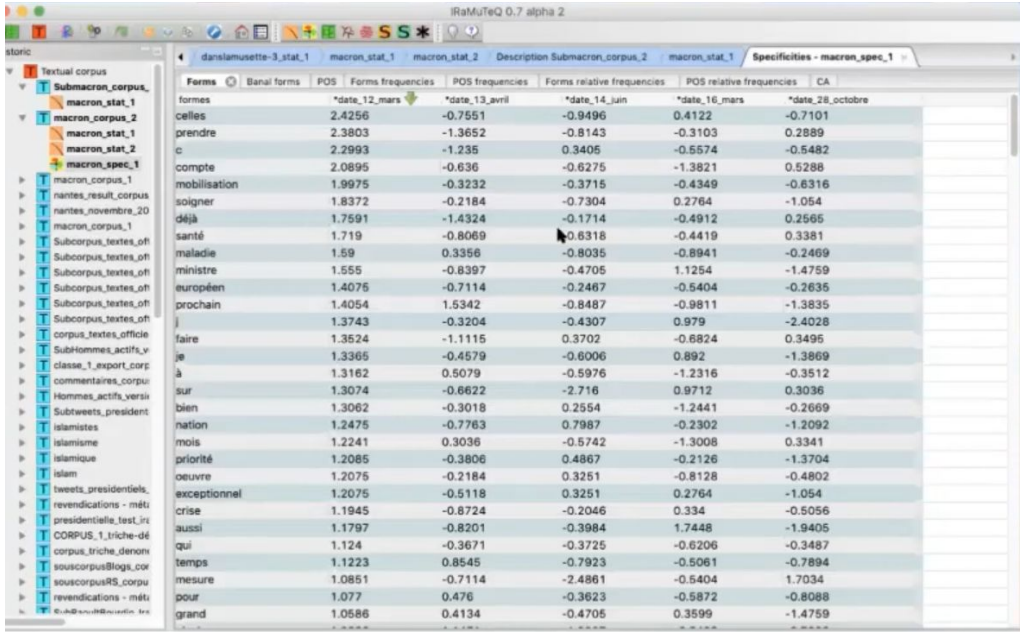


mod	freq	percent of total	row number	percent of rows
preconceito	114	7.61	114	38.0
tristeza	66	4.41	66	22.0
sofrimento	54	3.6	54	18.0
morte	52	3.47	52	17.33
sexo	50	3.34	50	16.67
doença	49	3.27	49	16.33
camisinha	39	2.6	39	13.0
medo	33	2.2	33	11.0
doente	31	2.07	31	10.33
tratamento	26	1.74	26	8.67
discriminação	25	1.67	25	8.33
vírus	21	1.4	21	7.0
irresponsabilidade	21	1.4	21	7.0
remédios	20	1.34	20	6.67
cuidados	17	1.13	17	5.67
cuidado	17	1.13	17	5.67
descuido	17	1.13	17	5.67
prevenção	16	1.07	16	5.33
irresponsável	15	1.0	15	5.0
dor	15	1.0	15	5.0
pena	14	0.93	14	4.67

IRaMuTeQ

2. Specificities & Correspondence Factor Analysis

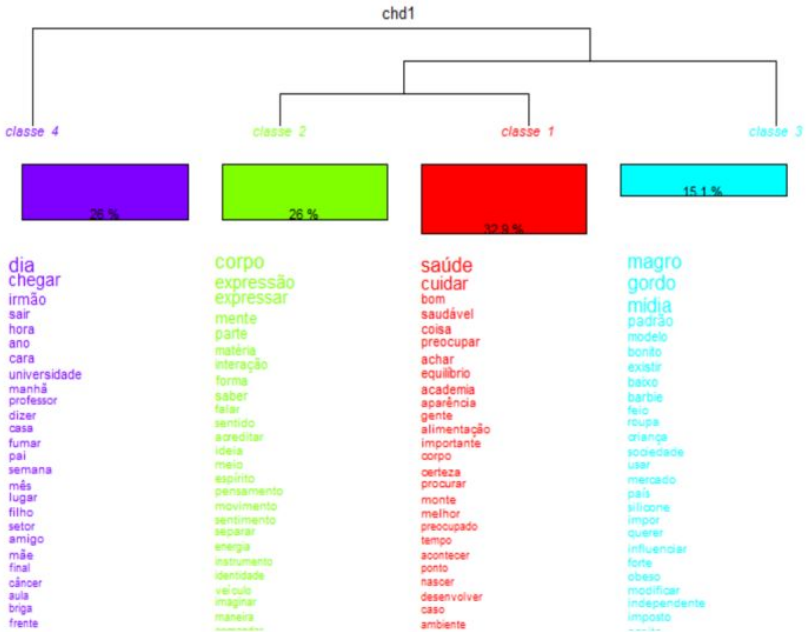
- Links text segments with variables (e.g., gender, age, topic).
- Shows which words are specific to each variable.
- Helps compare texs according to the characterization variables



IRaMuTeQ

3. Descending Hierarchical Analysis (DHA)

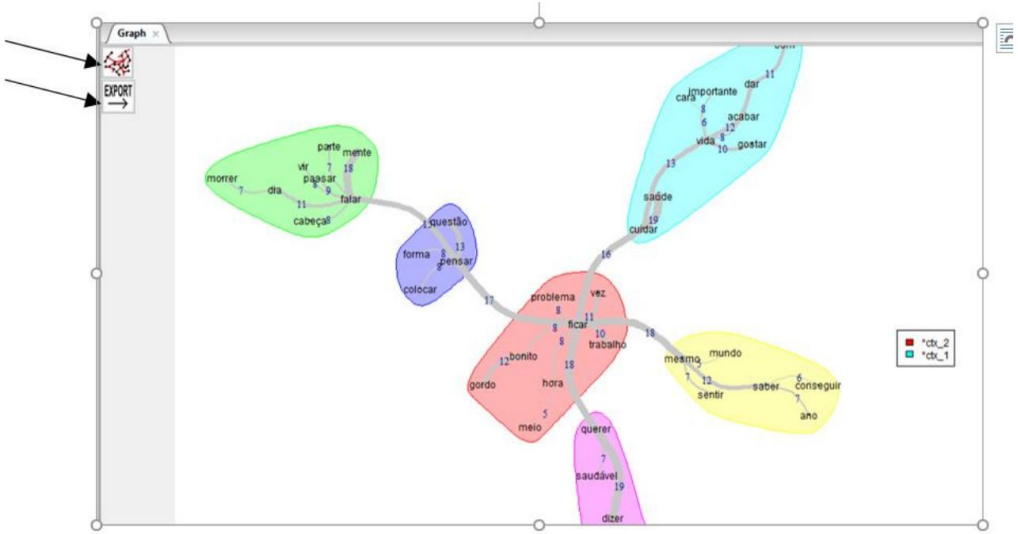
- Groups text segments based on shared vocabulary.
- Produces clusters with similar words.
- Shows the cluster structure through a dendrogram.
- Provides top words and key variables for each cluster.
- Used to identify themes, lexical fields, or social representations.



IRaMuTeQ

4. **Similarity Analysis**

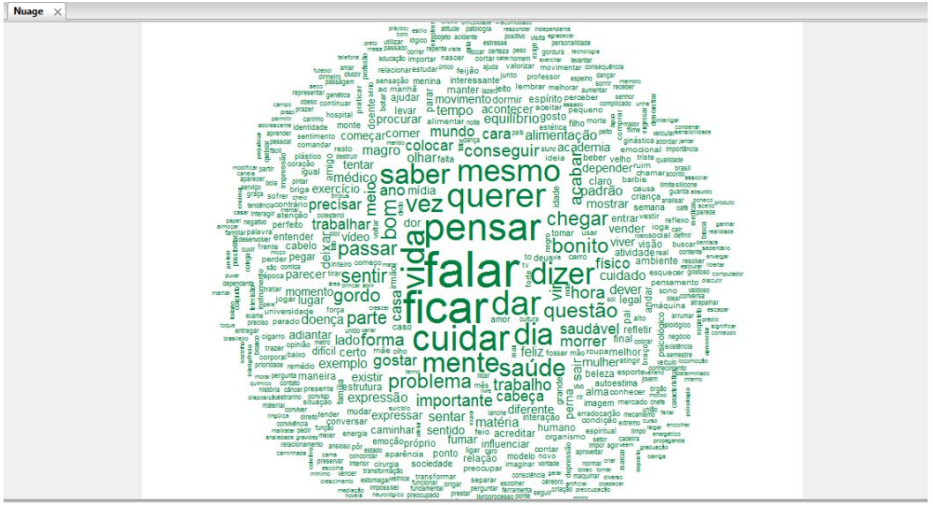
- Shows how words are connected through co-occurrences.
- Helps visualize the structure of the text (central vs. peripheral words).
- Allows to identify the word-cooccurrences providing information about the word connectivity in a corpus.
- Useful for identifying shared content across variables.



IRaMuTeQ

5. **Word Cloud**

- Displays words by size according to frequency.
- Simple and visual way to explore main vocabulary.



IRaMuTeQ


IRaMuTeQ
pratinaud.gitpages.huma-num.fr/iramuteq-website/
Nouvelle version de Chrome disponible
IRaMuTeQ
Accueil
Installation
Documentation
Aide
IRaMuTeQ
Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires. Un logiciel libre construit avec des logiciels libres.
Published: Sat 08 November 2025
By Pierre Ratinaud
In Accueil.
Présentation
Iramuteq est un logiciel libre distribué sous les termes de la licence GNU GPL (v2).
Il permet de faire des analyses statistiques sur des corpus texte et sur des tableaux individus/caractères.
Il repose sur le logiciel R (www.r-project.org) et le langage python (www.python.org)
Analyses sur des corpus texte
Iramuteq reproduit la méthode de classification décrite par Reinert (1983, 1991) (Classification Hiérarchique Descendante sur un tableau croisant les formes pleines et des segments de texte);
Recherche de spécificité à partir de segmentation définie;
ENG
14:35
11/12/2025

Tropes

- Open-source software for semantic analysis
- The initial release of the Tropes software dates to 1994

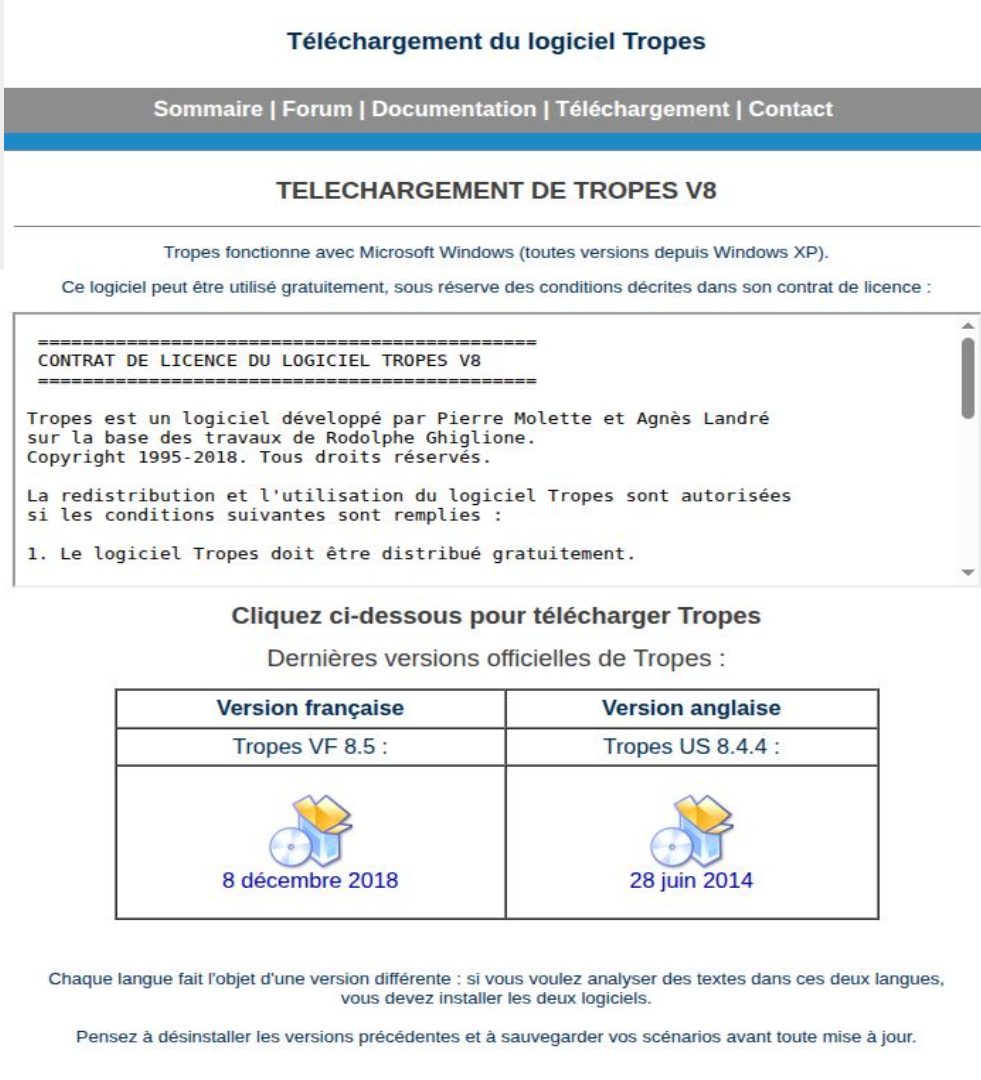
Webpage: https://www.tropes.fr/

Why Choose Tropes?

- Free Software
- No coding required
- No need for predefined conceptual categories or coding frames.
- Mainly a text-mining tool

Tropes

Installation guidelines:
https://www.tropes.fr/download.htm



Téléchargement du logiciel Tropes

Sommaire | Forum | Documentation | Téléchargement | Contact

TELECHARGEMENT DE TROPES V8

Tropes fonctionne avec Microsoft Windows (toutes versions depuis Windows XP).

Ce logiciel peut être utilisé gratuitement, sous réserve des conditions décrites dans son contrat de licence :

```
==========================================
CONTRAT DE LICENCE DU LOGICIEL TROPES V8
==========================================

Tropes est un logiciel développé par Pierre Molette et Agnès Landré
sur la base des travaux de Rodolphe Ghiglione.
Copyright 1995-2018. Tous droits réservés.

La redistribution et l'utilisation du logiciel Tropes sont autorisées
si les conditions suivantes sont remplies :

1. Le logiciel Tropes doit être distribué gratuitement.
```

Cliquez ci-dessous pour télécharger Tropes

Dernières versions officielles de Tropes :

Version française	Version anglaise
Tropes VF 8.5 :	Tropes US 8.4.4 :
8 décembre 2018	28 juin 2014

Chaque langue fait l'objet d'une version différente : si vous voulez analyser des textes dans ces deux langues, vous devez installer les deux logiciels.

Pensez à désinstaller les versions précédentes et à sauvegarder vos scénarios avant toute mise à jour.

Tropes

General Interface Structure

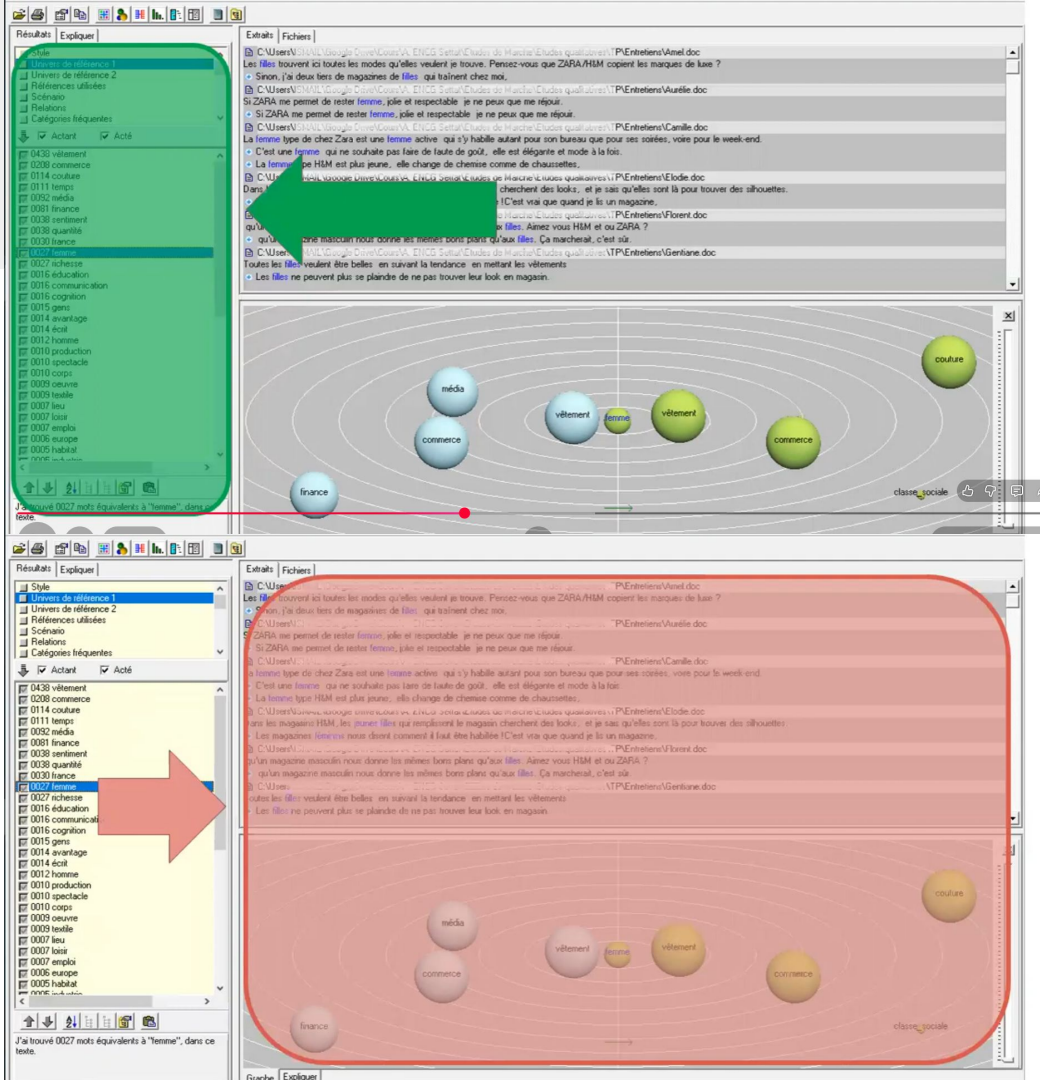
Trope's interface has two main areas:

Left Panel – Editing and Controls

- Controls overall operations.
- Allows selecting functions such as:
  - Edit
  - Analyse
  - Add or delete documents
  - Sort documents

Right Panel – Display of Results

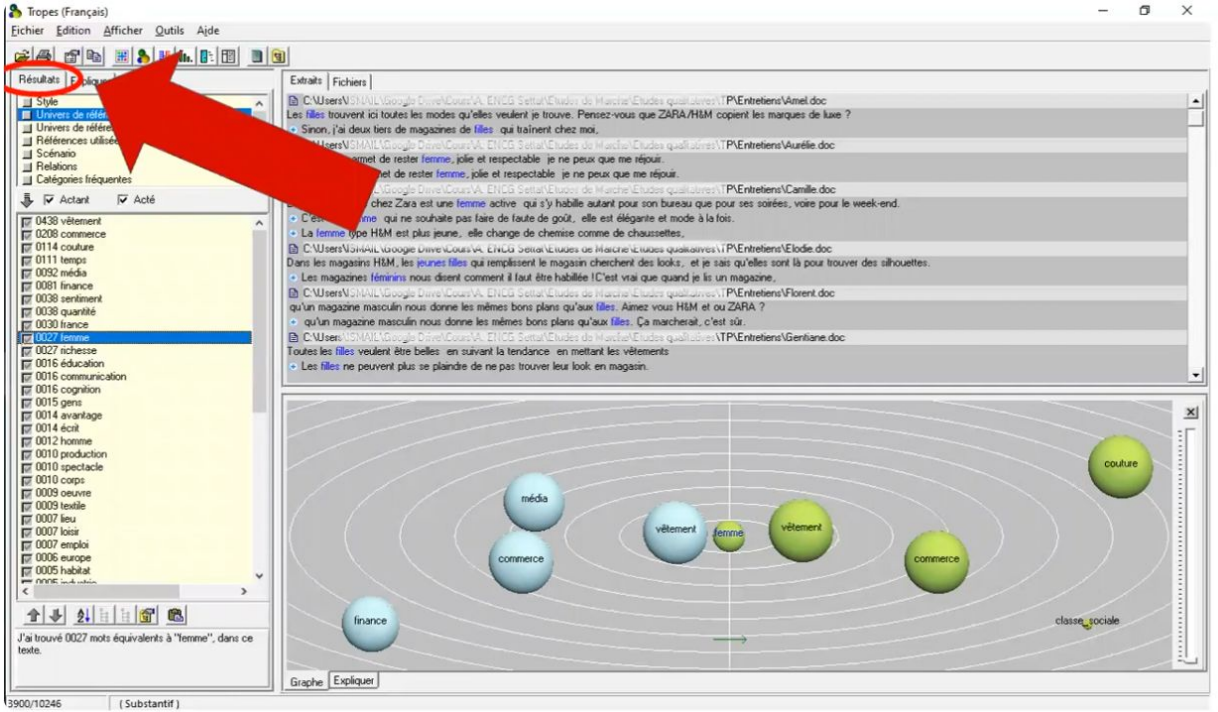
- Interactive visualisation.
- Shows:
  - Universes of reference
  - Text extracts
  - Graphical illustrations of results



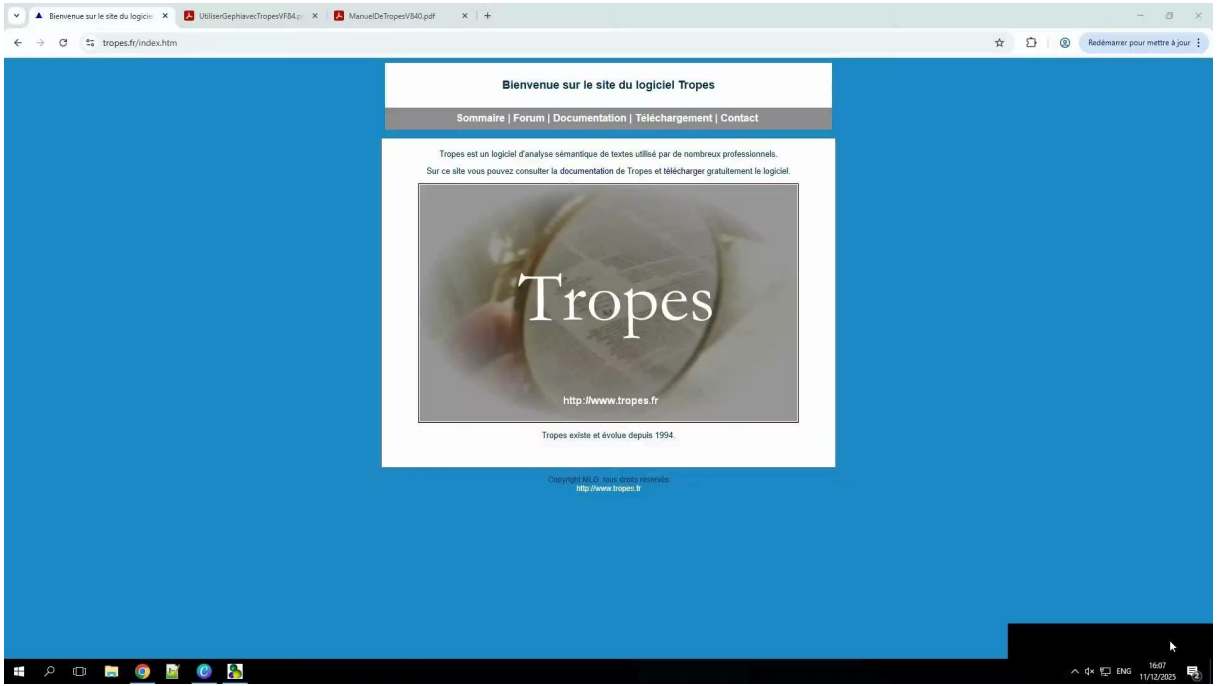
Tropes

Key Concept: Universes of Reference

- A core component of Tropes.
- A “universe of reference” is a classification or semantic universe grouping terms.
- Tropes organises words into:
  - Classes
  - Semantic universes
  - Levels (general to specific)

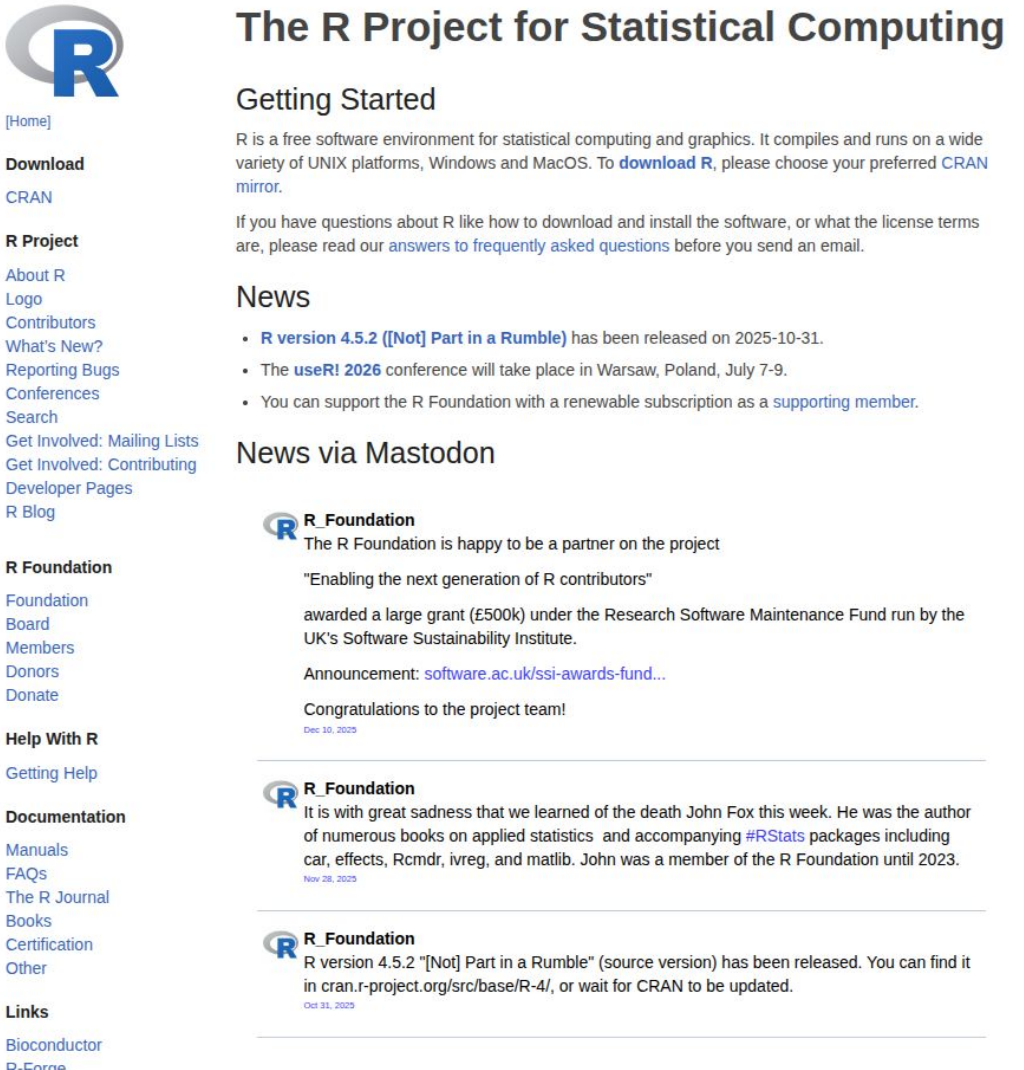


Tropes


Bienvenue sur le site du logiciel Tropes
tropes.fr/index.htm
Redémarrer pour mettre à jour
Bienvenue sur le site du logiciel Tropes
Sommaire | Forum | Documentation | Téléchargement | Contact
Tropes est un logiciel d'analyse sémantique de textes utilisé par de nombreux professionnels.
Sur ce site vous pouvez consulter la documentation de Tropes et télécharger gratuitement le logiciel.
Tropes
http://www.tropes.fr
Tropes existe et évolue depuis 1994.
ENG
16:07
11/12/2025

R & R Studio

- R is a free software environment for statistical computing and graphics.
- **Webpage:** https://www.r-project.org/



R & R Studio

Installation:

https://cran.r-project.org/mirrors.html

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- Download R for Linux (Debian, Fedora/Redhat, Ubuntu)
- Download R for macOS
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2025-10-31, [Not] Part in a Rumble) R-4.5.2.tar.gz, read what's new in the latest version.
- The CRAN directory src/base-prerelease contains R alpha, beta, and rc releases as daily snapshots in time periods before a planned release.
- Between releases, the same directory src/base-prerelease contains snapshots of current patched and development versions. Please read about new features and bug fixes before filing corresponding feature requests or bug reports.
- Alternatively, daily snapshots are available here.
- Source code of older versions of R is available here.
- Contributed extension packages.

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

Supporting CRAN

- CRAN operations, most importantly hosting, checking, distributing, and archiving of R add-on packages for various platforms, crucially rely on technical, emotional, and financial support by the R community.

  Please consider making financial contributions to the R Foundation for Statistical Computing.

What are R and CRAN?

R is 'GNU S', a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. Please consult the R project homepage for further information.

CRAN is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R. Please use the CRAN mirror nearest to you to minimize network load.

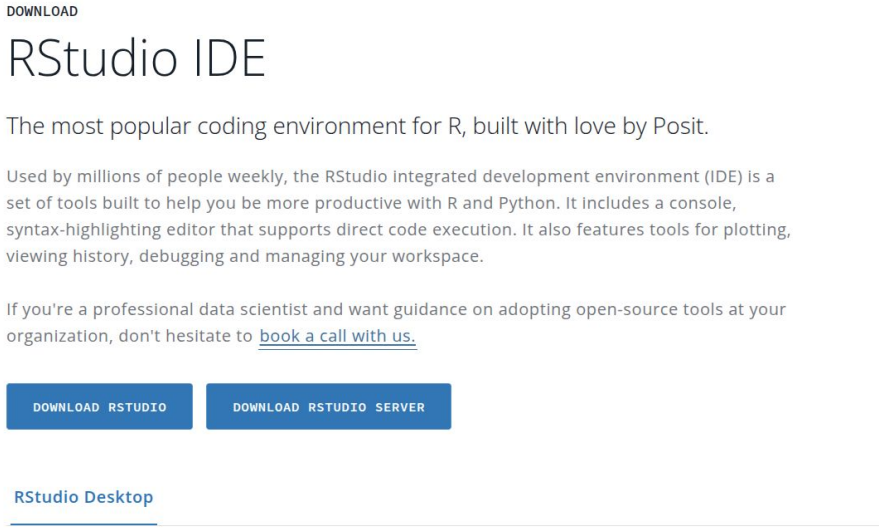
R & R Studio

Installation:

- https://posit.co/downloads/

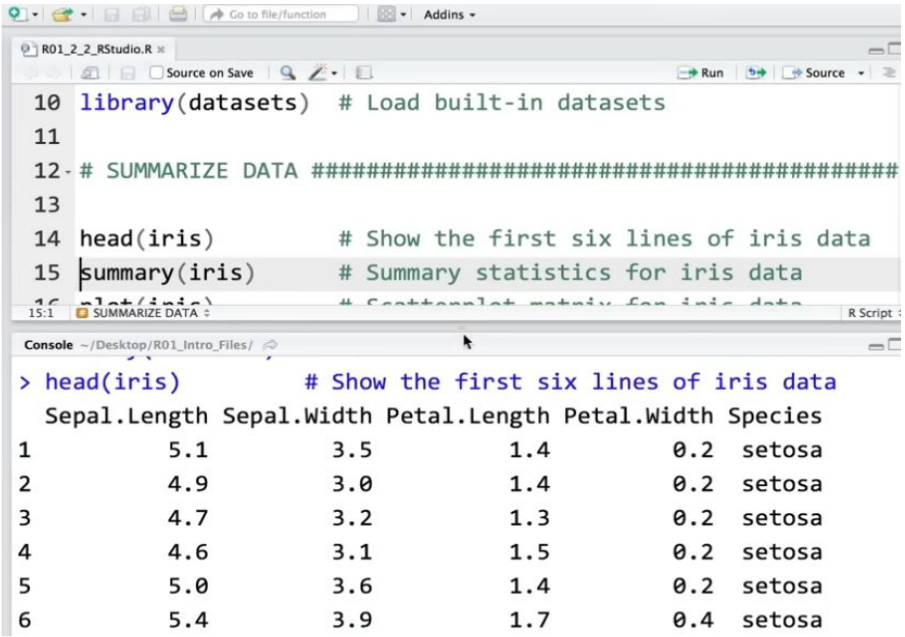
Why use R Studio?

- Makes working with R easier
- Standardizes keyboard shortcuts across Mac, Windows, and Linux
- Provides a consistent user experience

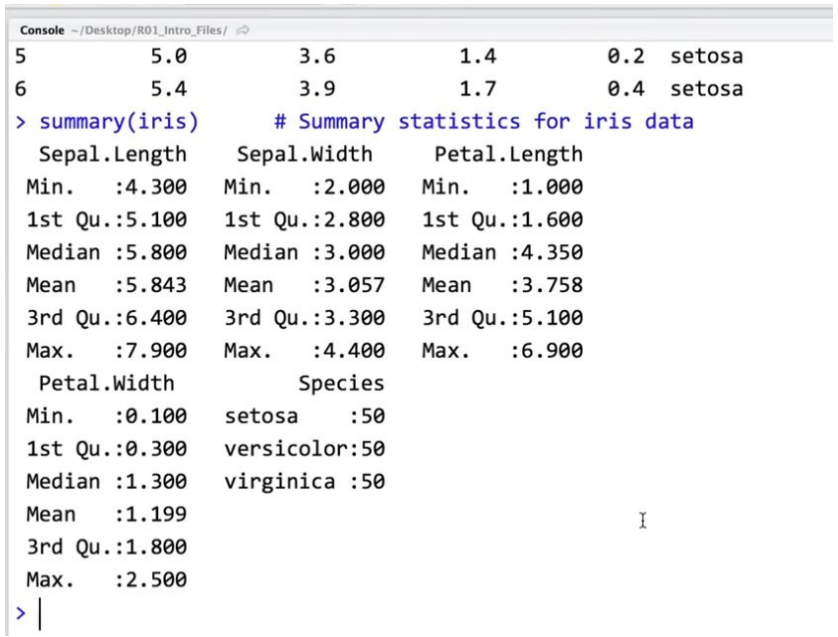


R & R Studio

Loading and printing datasets

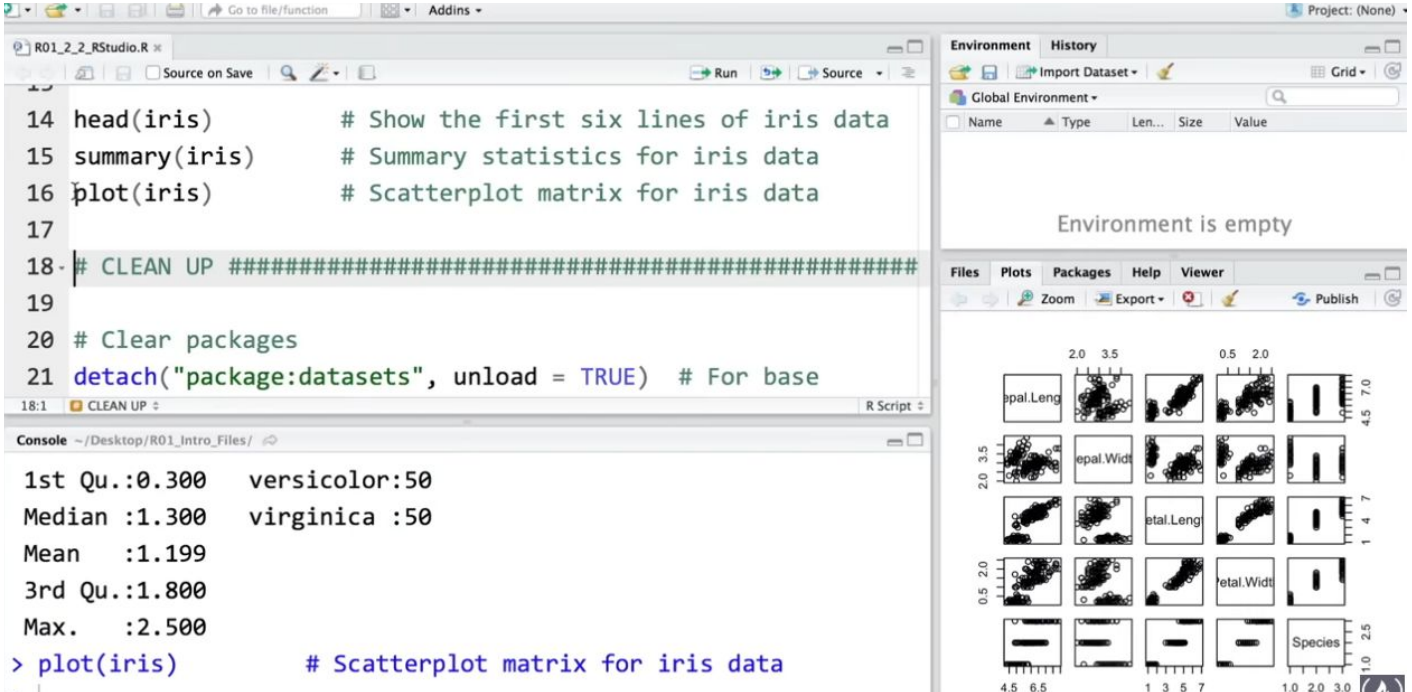


Descriptive statistics



R & R Studio

Plotting datasets



R & R Studio

Histograms

```
# Histograms for each species using options
hist(iris$Petal.Width [iris$Species == "setosa"],
  xlim = c(0, 3),
  breaks = 9,
  main = "Petal Width for Setosa",
  xlab = "",
  col = "red")
```

32:14 HISTOGRAM BY GROUP R Script

Console ~/Desktop/R01_Intro_Files/

```
6          5.4          3.9          1.7          0.4  setosa
> hist(iris$Sepal.Length)
> hist(iris$Sepal.Width)
> hist(iris$Petal.Length)
> hist(iris$Petal.Width)
> par(mfrow = c(3, 1))
> head(iris)
```

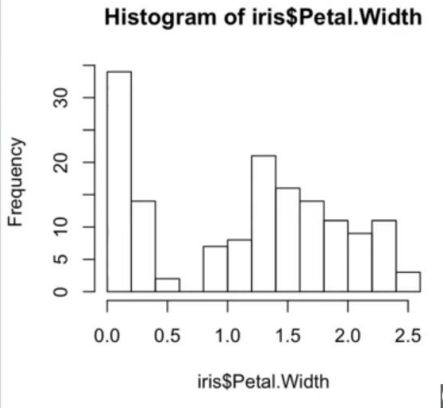
Global Environment

Name | Type | Len... | Size | Value

Environment is empty

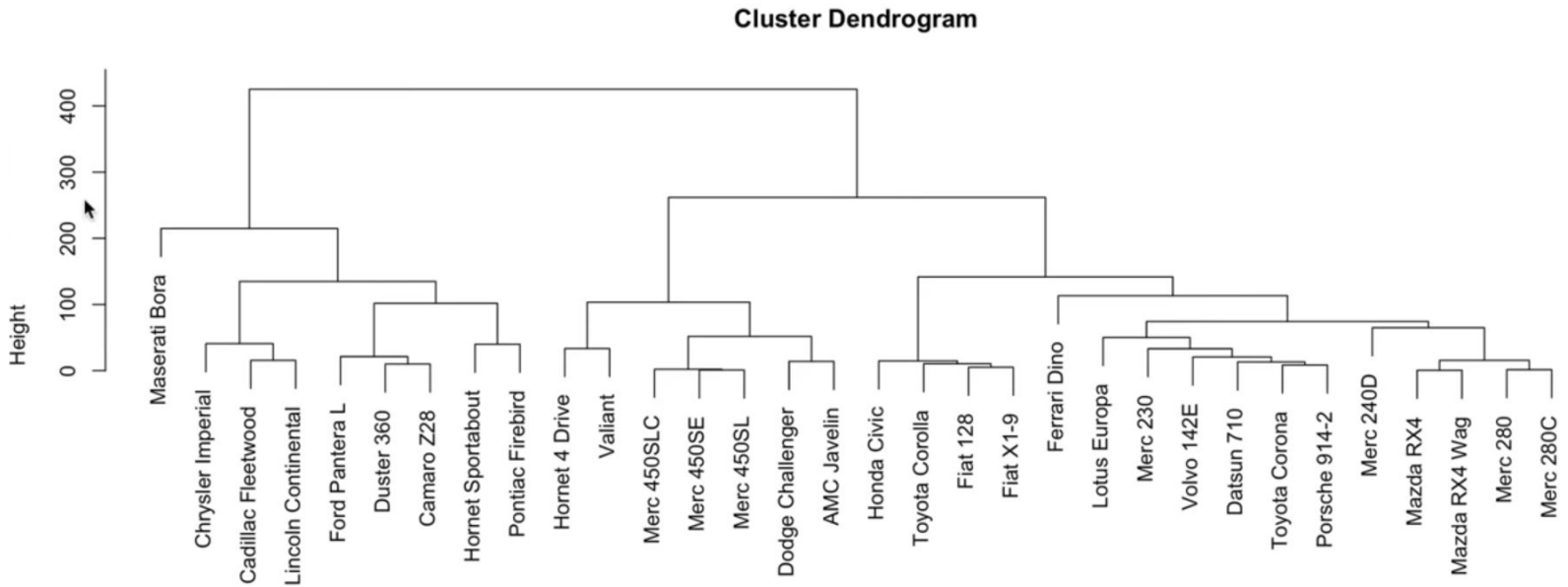
Files | Plots | Packages | Help | Viewer

Zoom | Export



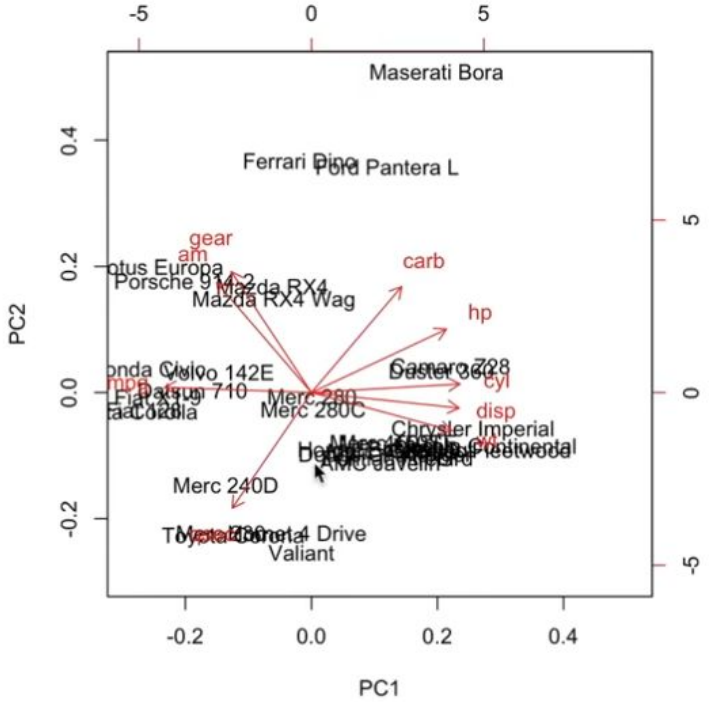
R & R Studio

Hierarchical Clustering

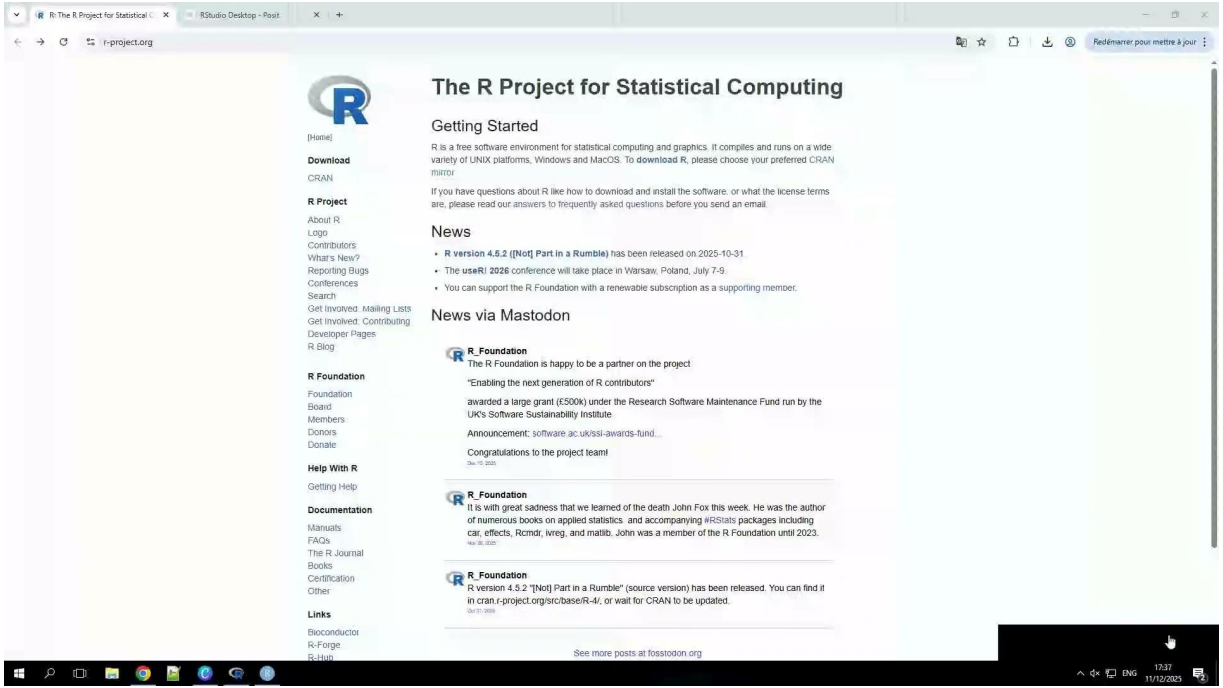


R & R Studio

Principal Component Analysis (PCA)

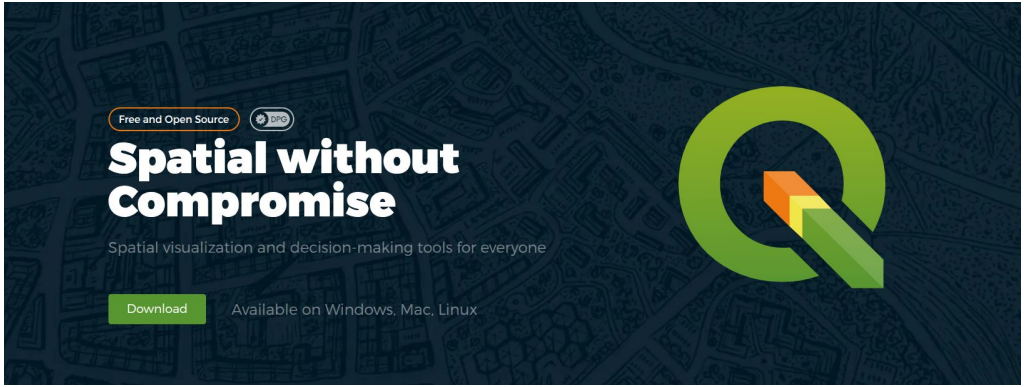


R & R Studio


R: The R Project for Statistical C
RStudio Desktop - Posit
r-project.org
Redémarrer pour mettre à jour
The R Project for Statistical Computing
[Home]
Download
CRAN
R Project
About R
Logo
Contributors
What's New?
Reporting Bugs
Conferences
Search
Get Involved: Mailing Lists
Get Involved: Contributing
Developer Pages
R Blog
R Foundation
Foundation
Board
Members
Donors
Donate
Help With R
Getting Help
Documentation
Manuals
FAQs
The R Journal
Books
Certification
Other
Links
Bioconductor
R-Forge
R-Hub
Getting Started
R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To download R, please choose your preferred CRAN mirror.
If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.
News
R version 4.5.2 ([Not] Part in a Rumble) has been released on 2025-10-31.
The useR! 2026 conference will take place in Warsaw, Poland, July 7-9.
You can support the R Foundation with a renewable subscription as a supporting member.
News via Mastodon
R_Foundation
The R Foundation is happy to be a partner on the project
"Enabling the next generation of R contributors"
awarded a large grant (£500k) under the Research Software Maintenance Fund run by the UK's Software Sustainability Institute
Announcement: software.ac.uk/ssi-awards-fund...
Congratulations to the project team!
R_Foundation
It is with great sadness that we learned of the death John Fox this week. He was the author of numerous books on applied statistics, and accompanying #RStats packages including car, effects, Rcmdr, ivreg, and matlib. John was a member of the R Foundation until 2023.
R_Foundation
R version 4.5.2 "[Not] Part in a Rumble" (source version) has been released. You can find it in cran.r-project.org/src/base/R-4/, or wait for CRAN to be updated.
See more posts at fosstodon.org
ENG
17:37
31/12/2025

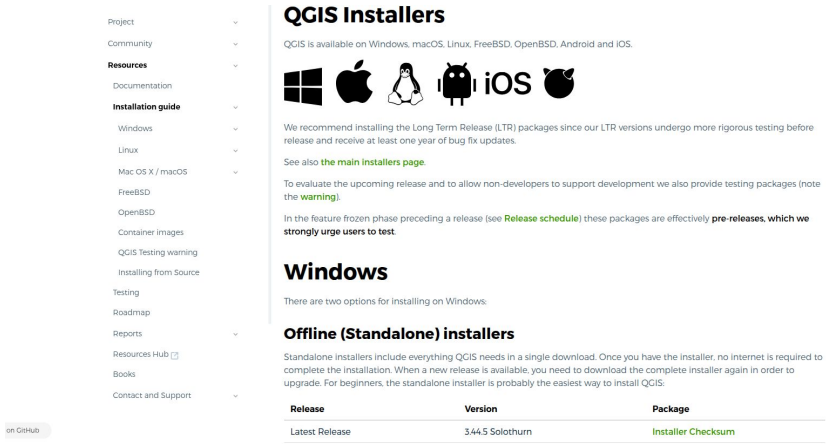
QGIS

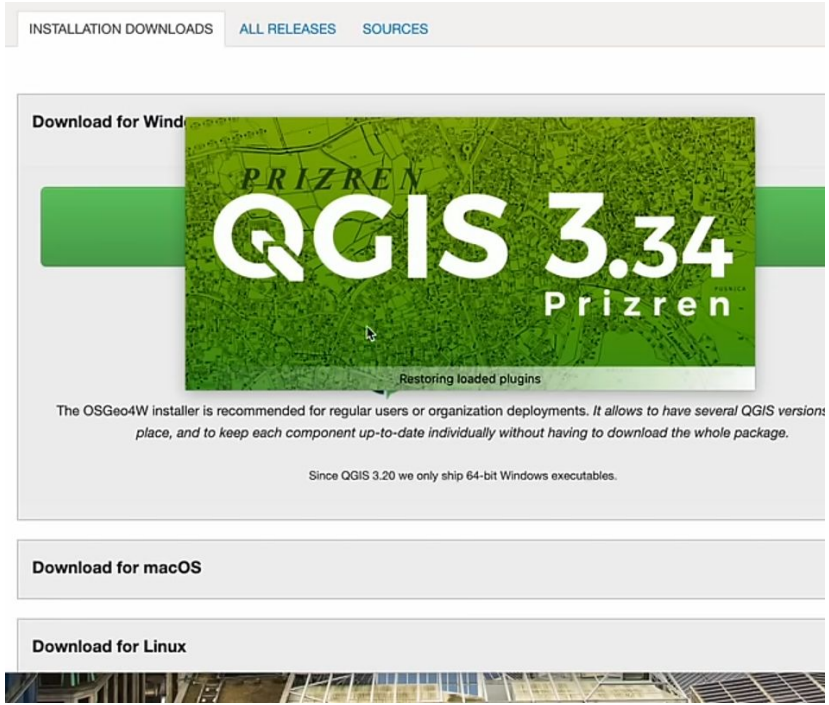
- QGIS is an open-source tool for creating, editing, visualizing, analyzing, and sharing geospatial data.
- Runs on Mac, Windows, Linux, and can be deployed in cloud environments.
- Offers a rich plugin ecosystem to extend its mapping, analysis, and data management capabilities.
- Supports Python scripting (PyQGIS) for automation and advanced customization.
- **Webpage:** https://qgis.org/



QGIS

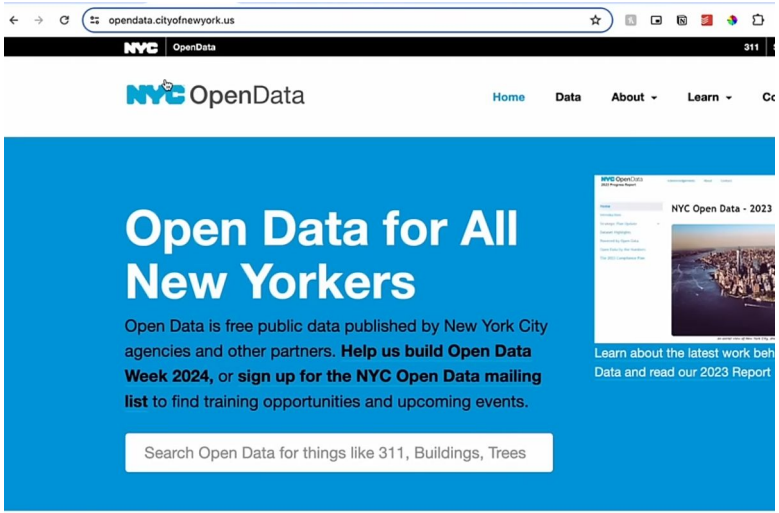
Installation

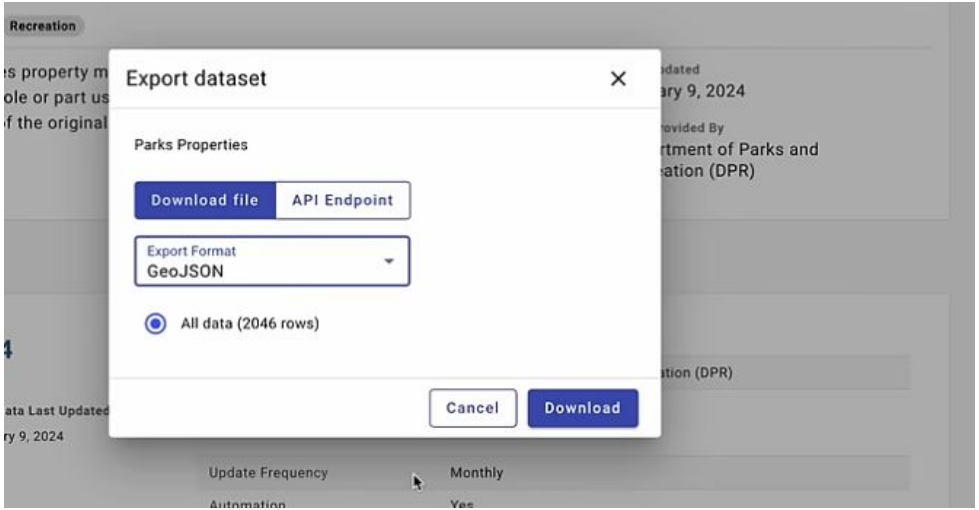




QGIS

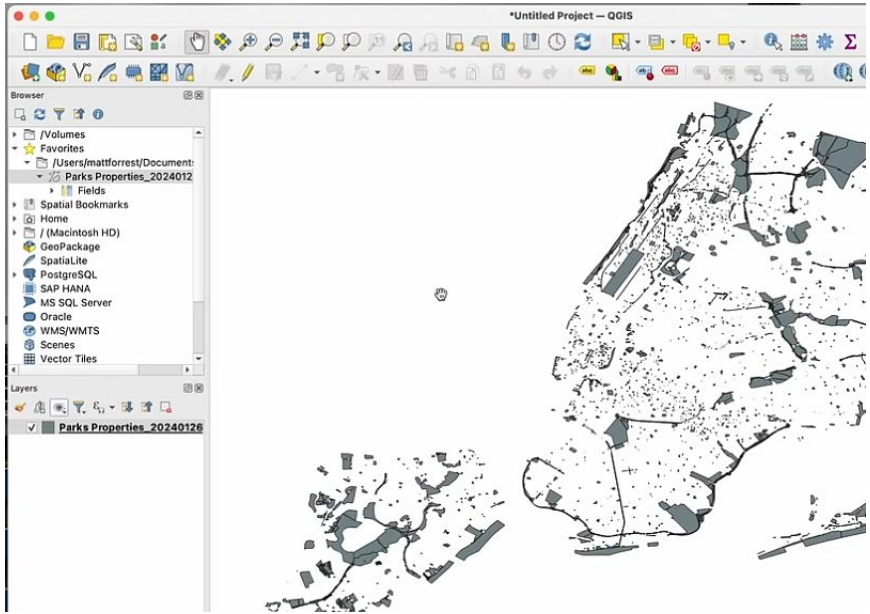
Exporting data



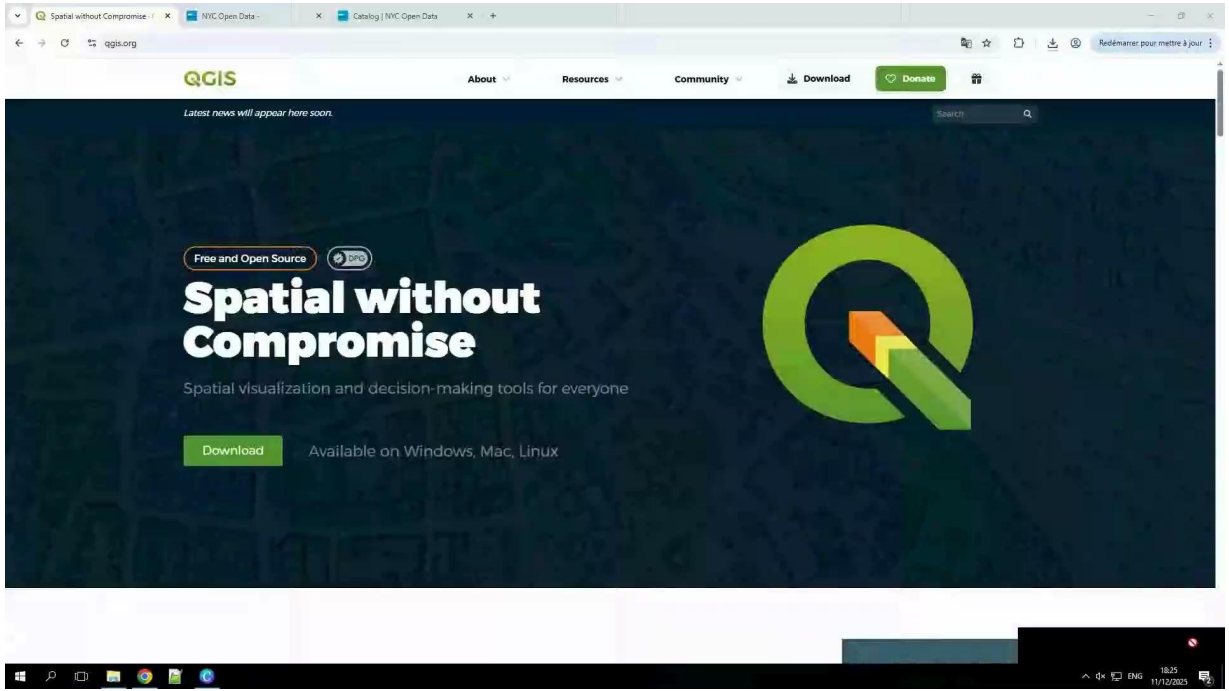


QGIS

Adding Layers

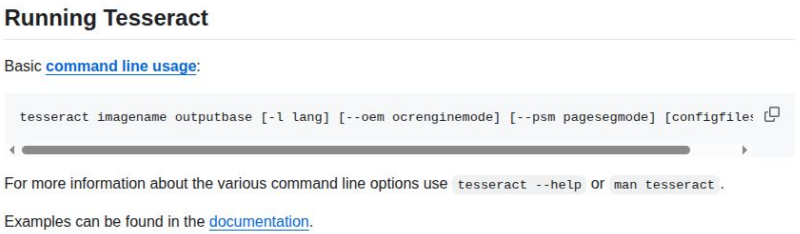


QGIS


qgis.org
QGIS
About
Resources
Community
Download
Donate
Latest news will appear here soon.
Search
Free and Open Source
Spatial without Compromise
Spatial visualization and decision-making tools for everyone
Download
Available on Windows, Mac, Linux

Tesseract

- Open-source Optical Character Recognition (OCR) engine.
- Provides libtesseract (library) and a command-line program.
- Supports 100+ languages and UTF-8.
- **Webpage:** https://github.com/tesseract-ocr/tesseract



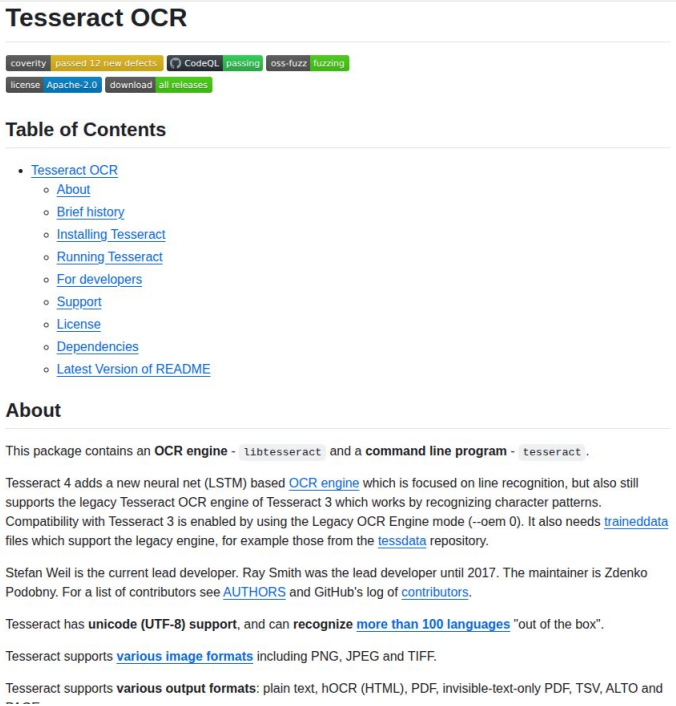
Running Tesseract

Basic **command line usage**:

```
tesseract imagename outputbase [-l lang] [--oem ocrenginemode] [--psm pagesegmode] [configfiles
```

For more information about the various command line options use `tesseract --help` or `man tesseract`.

Examples can be found in the documentation.



Tesseract OCR

coverity passed 12 new defects | CodeQL passing | oss-fuzz fuzzing

license Apache-2.0 | download all releases

Table of Contents

About

This package contains an **OCR engine** - `libtesseract` and a **command line program** - `tesseract`.

Tesseract 4 adds a new neural net (LSTM) based OCR engine which is focused on line recognition, but also still supports the legacy Tesseract OCR engine of Tesseract 3 which works by recognizing character patterns. Compatibility with Tesseract 3 is enabled by using the Legacy OCR Engine mode (--oem 0). It also needs traineddata files which support the legacy engine, for example those from the tessdata repository.

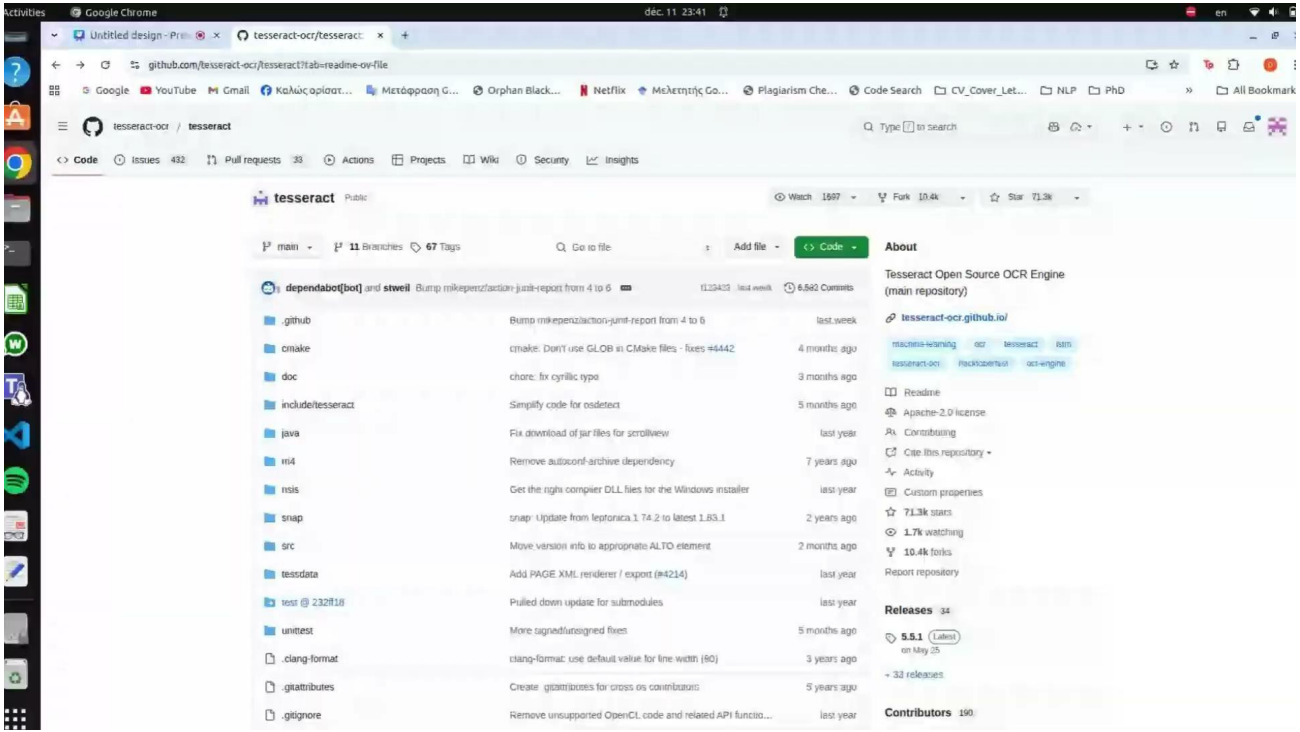
Stefan Weil is the current lead developer. Ray Smith was the lead developer until 2017. The maintainer is Zdenko Podobny. For a list of contributors see AUTHORS and GitHub's log of contributors.

Tesseract has **unicode (UTF-8) support**, and can **recognize more than 100 languages** "out of the box".

Tesseract supports **various image formats** including PNG, JPEG and TIFF.

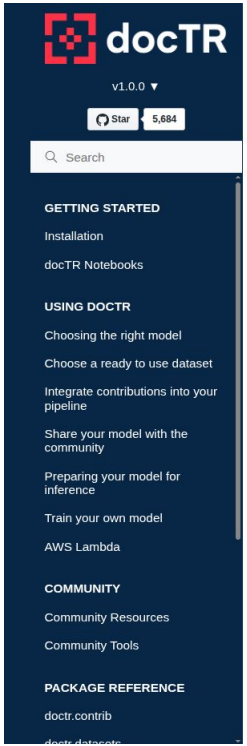
Tesseract supports **various output formats**: plain text, hOCR (HTML), PDF, invisible-text-only PDF, TSV, ALTO and

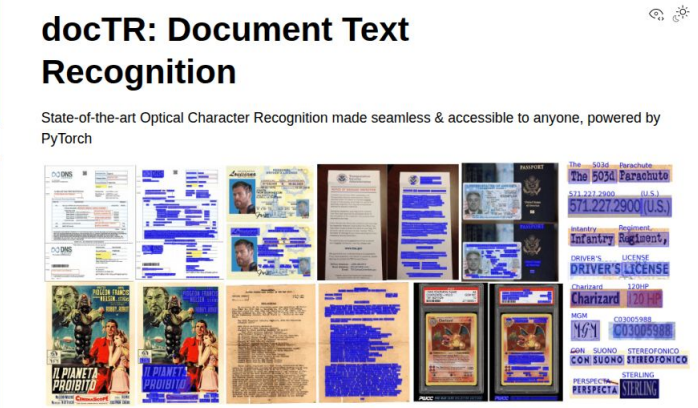
Tesseract


Activities
Google Chrome
déc. 11 23:41
Untitled design - Pri
tesseract-ocr/tesseract:
github.com/tesseract-ocr/tesseract?tab=readme-ov-file
Google
YouTube
Gmail
Καλώς ορίσατ...
Μετάφραση G...
Orphan Black...
Netflix
Μελετητής Go...
Plagiarism Che...
Code Search
CV_Cover_Let...
NLP
PhD
All Bookmark
tesseract-ocr / tesseract
Type / to search
Code
Issues 432
Pull requests 33
Actions
Projects
Wiki
Security
Insights
tesseract
Public
Watch 1697
Fork 10.4k
Star 71.3k
main
11 Branches
67 Tags
Go to file
Add file
Code
dependabot[bot] and stweil Bump mikepenz/action-junit-report from 4 to 6
6,582 Commits
.github
Bump mikepenz/action-junit-report from 4 to 6
last week
cmake
cmake: Don't use GLOB in CMake files - fixes #4442
4 months ago
doc
chore: fix cyrillic typo
3 months ago
include/tesseract
Simplify code for osdetect
5 months ago
java
Fix download of jar files for scrollview
last year
m4
Remove autoconf-archive dependency
7 years ago
nsis
Get the right compiler DLL files for the Windows installer
last year
snap
snap: Update from leptonica 1.74.2 to latest 1.83.1
2 years ago
src
Move version info to appropriate ALTO element
2 months ago
tessdata
Add PAGE XML renderer / export (#4214)
last year
test @ 232ff18
Pulled down update for submodules
last year
unittest
More signed/unsigned fixes
5 months ago
.clang-format
clang-format: use default value for line width (80)
3 years ago
.gitattributes
Create .gitattributes for cross os contributors
5 years ago
.gitignore
Remove unsupported OpenCL code and related API functio...
last year
About
Tesseract Open Source OCR Engine (main repository)
tesseract-ocr.github.io/
Readme
Apache-2.0 license
Contributing
Cite this repository
Activity
Custom properties
71.3k stars
1.7k watching
10.4k forks
Report repository
Releases 34
5.5.1 Latest
on May 25
+ 33 releases
Contributors 190

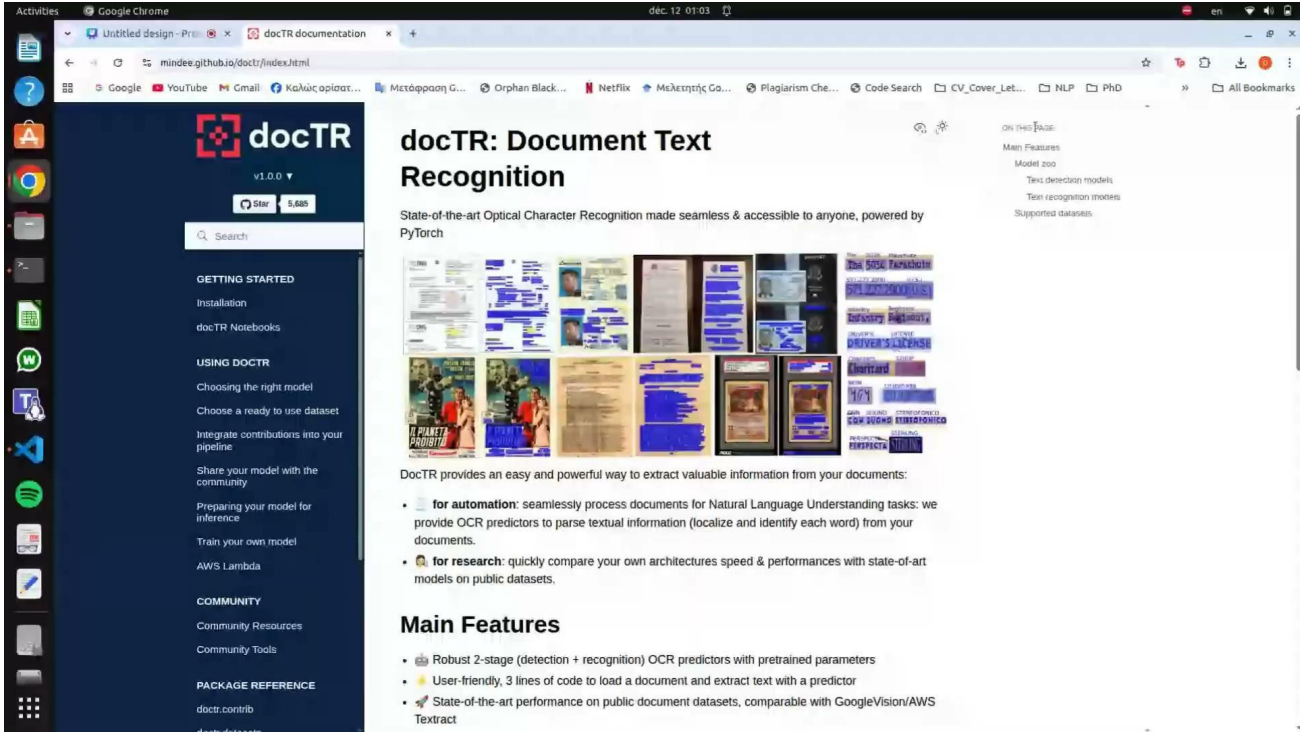
Python docTR

- Open-source Document Text Recognition library by Mindee.
- Built on PyTorch for modern deep-learning OCR.
- Provides text detection + text recognition in documents.
- Extracts and structures textual information from scanned or photographed documents.
- **Webpage:** https://mindee.github.io/doctr/index.html





Python docTR



Merci pour votre attention !

https://cyidhn.cyu.fr/accueil