

Oțios

[2605 RO+](#)

[2025.11+](#)

[Initial / previous notes](#)

[Roadmap](#)

[Dev notes](#)

[Misc reference](#)

[gh:pax/otzios](https://github.com/pax/otzios)

2605 RO+

enhance language - flash cards

translate/obfuscate text → halfway duolingo. with LLM enhancements?

super meta navigator

create custom lists, publish online

manifest suveranist – să ne luăm limba înapoi

[video] song generation – viral contest, open stages

<https://clre.solirom.ro/>

2025.11+

get lingvist, publish paper

tools, quizzes, complică un text cu arhaisme/regionalisme

skibidi: Create a reverse, browse news and r/romania and find new words, used more than 3? times that are not in dictionary → alternative dictionary

Newsletter..choose frequency

Email, wa.. but somewhat randomly

Translator: make it weird

make it international? –with language ambassadors

publish package, reverse of wordfreq

ex:

- <https://anidescoala.ro/educatie/gramatica/stii-sa-scrii/otios/> (web archive)
- [Ce produc universitatile romanesti - Contributors](#) (web archive)
- https://www.tvr.ro/vedete/mirela-nagat-e-nevoie-sa-facem-front-comun-ca-sa-popularizam-valorile-sa-propunem-modele-sa-cream-inspira-ie_52919.html (web archive)
- [fb search: oțios](#)

see also: [google](#)

add project to perma.cc

Initial / previous notes

caută cuvinte nefolosite

get cuvinte rare

google them, elimină ce-i dicționar sau nu limba română
din dexonline elimină vocabularul de bază

listă de articole, last seen

Google search!

check with Google API - note exceptions, articles with those terms – search for and link web archive version!

<term> site:.ro -sinonim -dicționar -definiție -anagramă -rimează
&lr=lang_ro
&tbs=lr:lang_1ro,qdr:y

site:.ro -dicționar

[q1](#), [q2](#)

Also search SM, fb, X etc

Roadmap

- | | |
|---|---|
| ☐ get vocabular de bază | text %vocabular de bază |
| ☐ diff w dexonline - elimină arhaisme? | ☐ launch w gov2 sides.gov2.ro / playground.gov2.ro |
| ☐ search w google, bing, etc | ☐ stats on ro.wikipedia |
| ☐ make list | ☐ filter-out wikipedia (and other corpus - GDEL, commoncrawl) articles from dictionaries |
| ☐ last seen | |
| ☐ LLM trained to use old words as synonyms | |
| ☐ meta: compute lexical outstanding of a | |

Get vocabular de bază

how to open a .dict ?

Process recent RO news corpus?

Search for Ro LLM training datasets?

- [Counting Word Frequency in a File Using Python](#)
- [How to find word frequencies of each word from a large corpus?](#)
- [rspeer/wordfreq](#): Access a database of word frequencies, in various natural languages
- [orgtre/top-open-subtitles-sentences](#): Most common sentences and words for all languages in the OpenSubtitles2018 corpus with Python code
- [dexonline](#) Dictionare în formatul StarDict
- [Limited JSON API for Google searches using Programmable Search Engine](#)

Wikipedia stats

- [IlyaSemenov/wikipedia-word-frequency](#): Gather modern English word frequencies from all enwiki articles.
- [Numeri/wiki-wordlists](#): Generates Unicode word frequency lists from Wikipedia pages in any language.
- [kohjiaxuan/Wikipedia-Article-Scraper](#): A complete Python text analytics package that allows users to search for a Wikipedia article, scrape it, conduct basic text analytics and integrate it to a data pipeline without writing excessive code.
-

Dev notes

wikipedia-word-frequency

[IlyaSemenov/wikipedia-word-frequency](#) / Google Colab attempt

 [wikipedia-word-frequency.ipynb](#)

Misc reference

Involve dexonline

- [Listă cu cele mai folosite cuvinte din limba română \(vocabularul fundamental\)](#)
- [stopwords-ro.txt](#)
- [Green's Dictionary of Slang | Hacker News](#)
- [The Phrontistery](#)
- <https://www.dilema.ro/autor/rodica-zafiu>
- <https://www.facebook.com/Letopiset/>
- <https://clre.solirom.ro/>