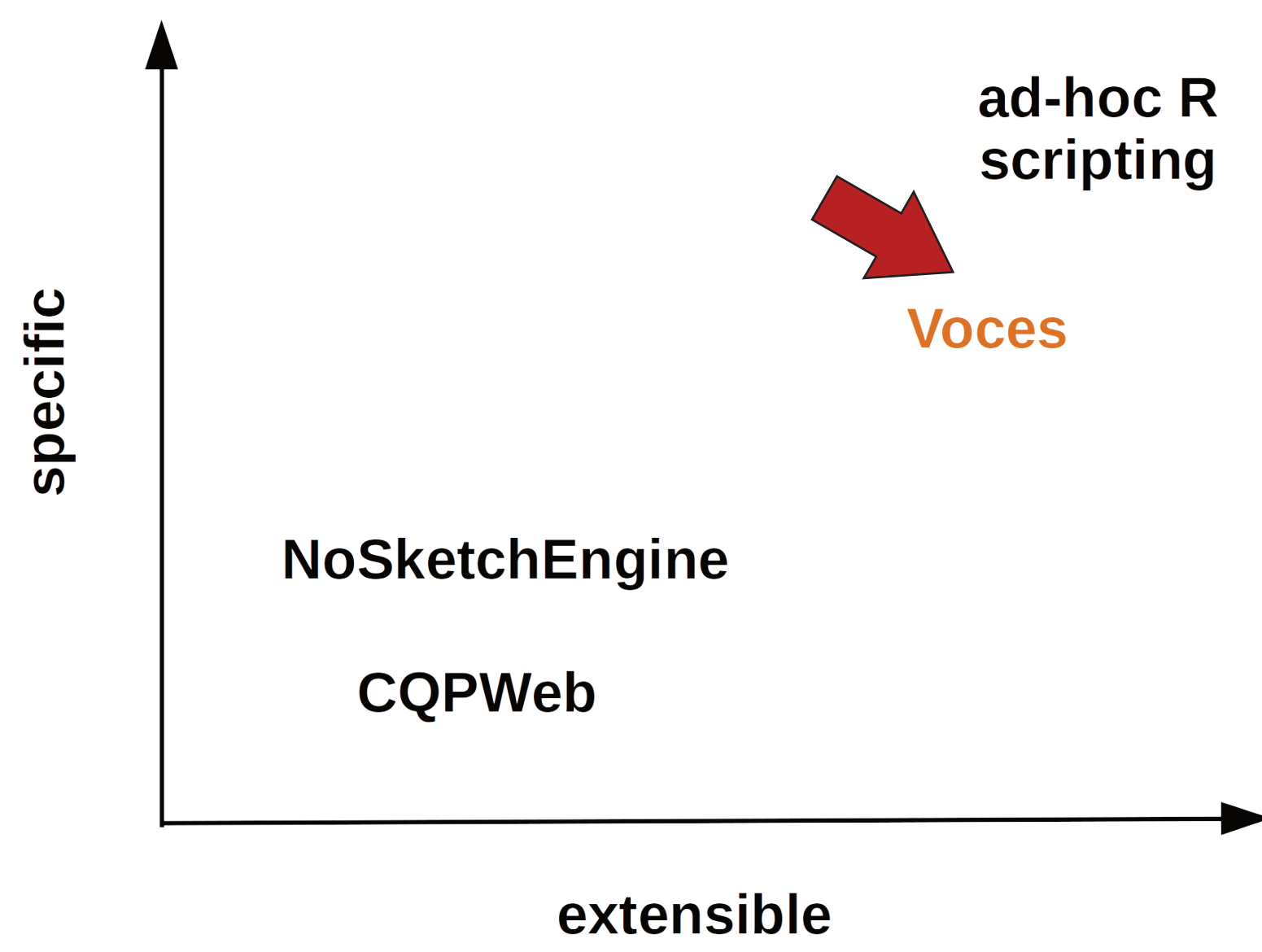# VOCES. An R-based Dashboard for Lexical Semantics

Krzysztof Nowak (krzysztof.nowak@ijp.pan.pl)

Institute of Polish Language (Polish Academy of Sciences)

Motivation

- quick inspection of word properties in lexicographic and lexicologist practice
- rapid integration of the code produced by a research group and external R packages (wordspace etc.)
- reducing distance between development and users
- filling the space between fully-fledged corpus query systems and ad-hoc scripting
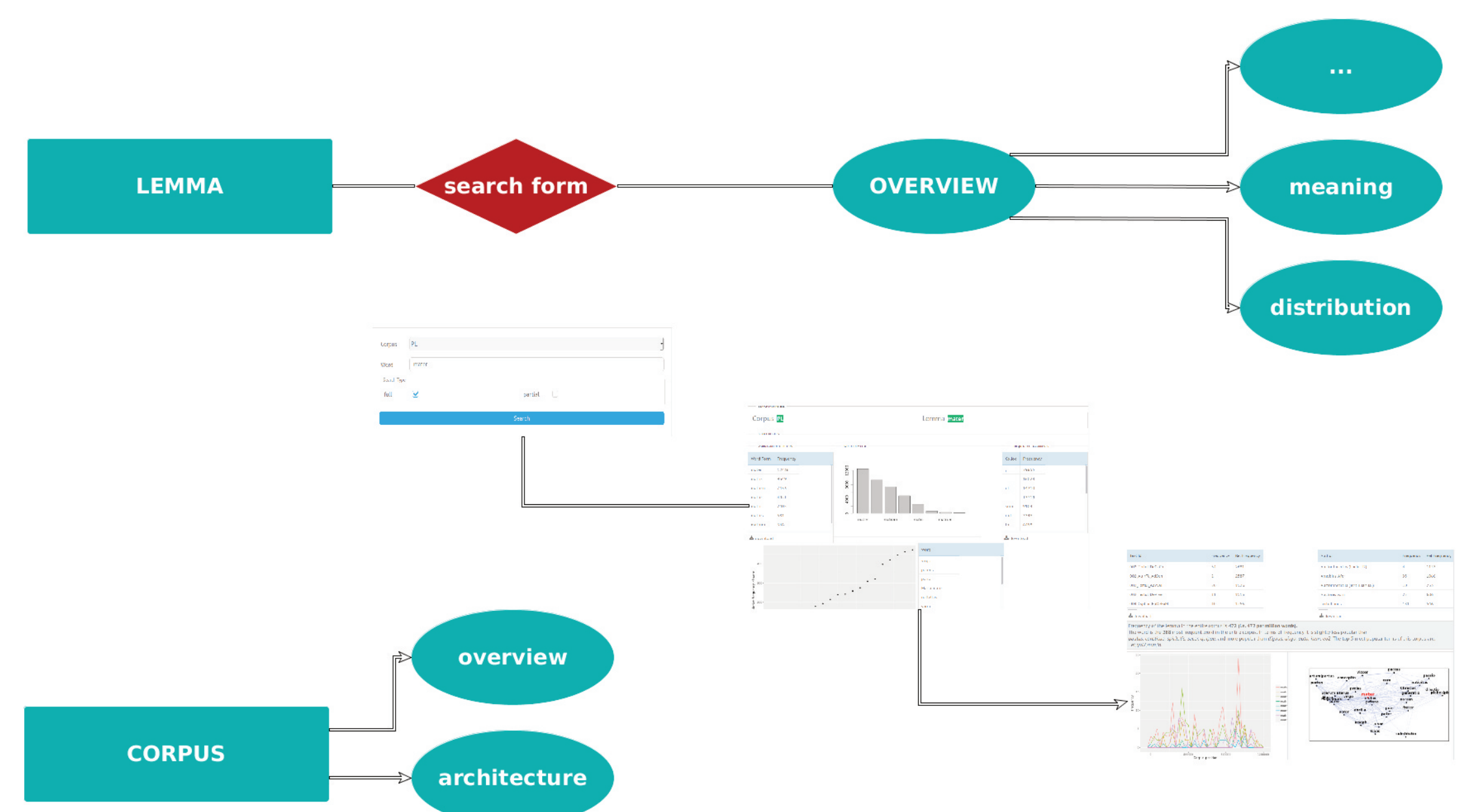
Points of Access

single lemma
- query starts with the simple search form
- on success the Overview page is populated with widgets presenting selected lemma properties
- on widget click the Details page opens

corpus
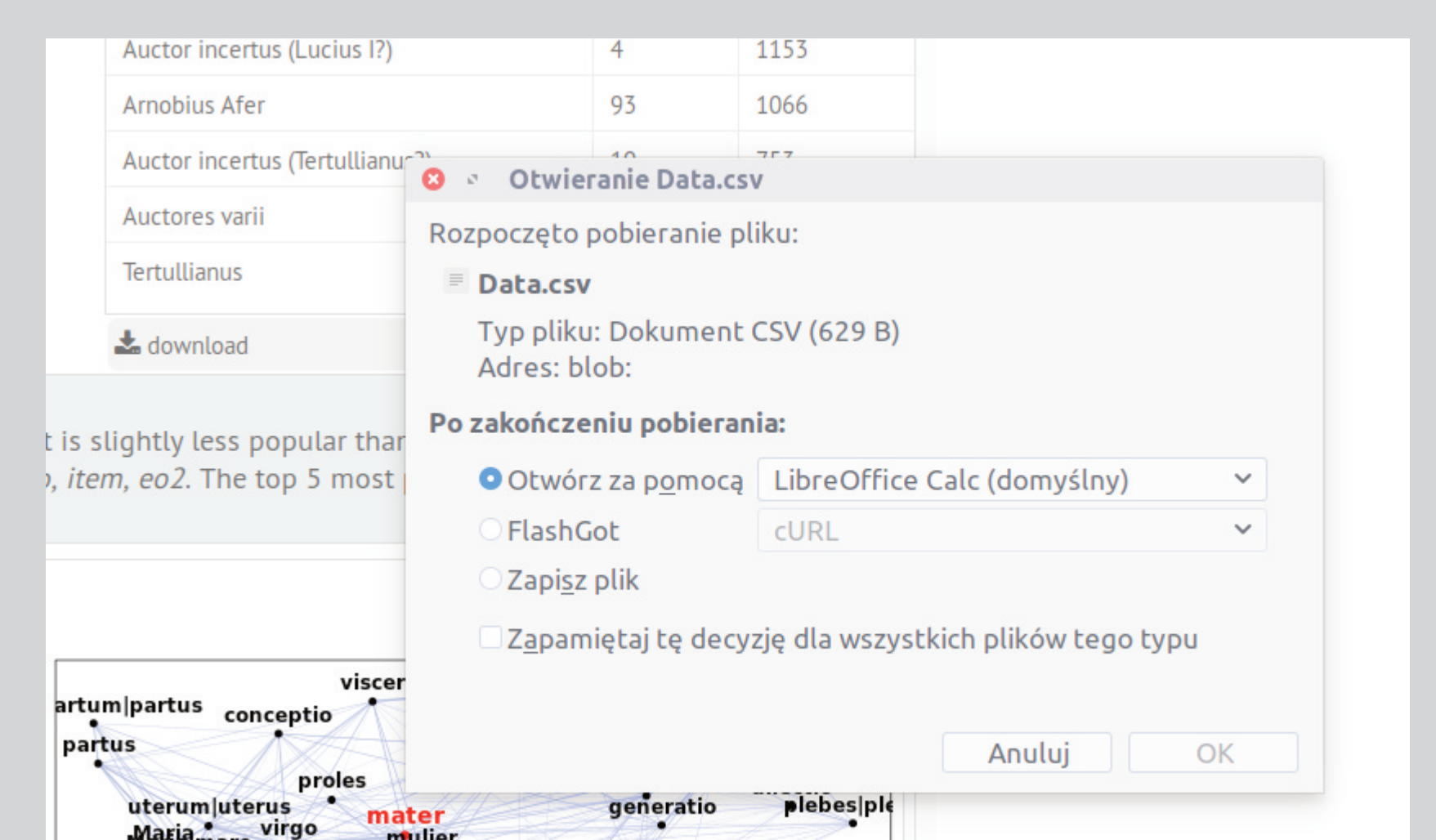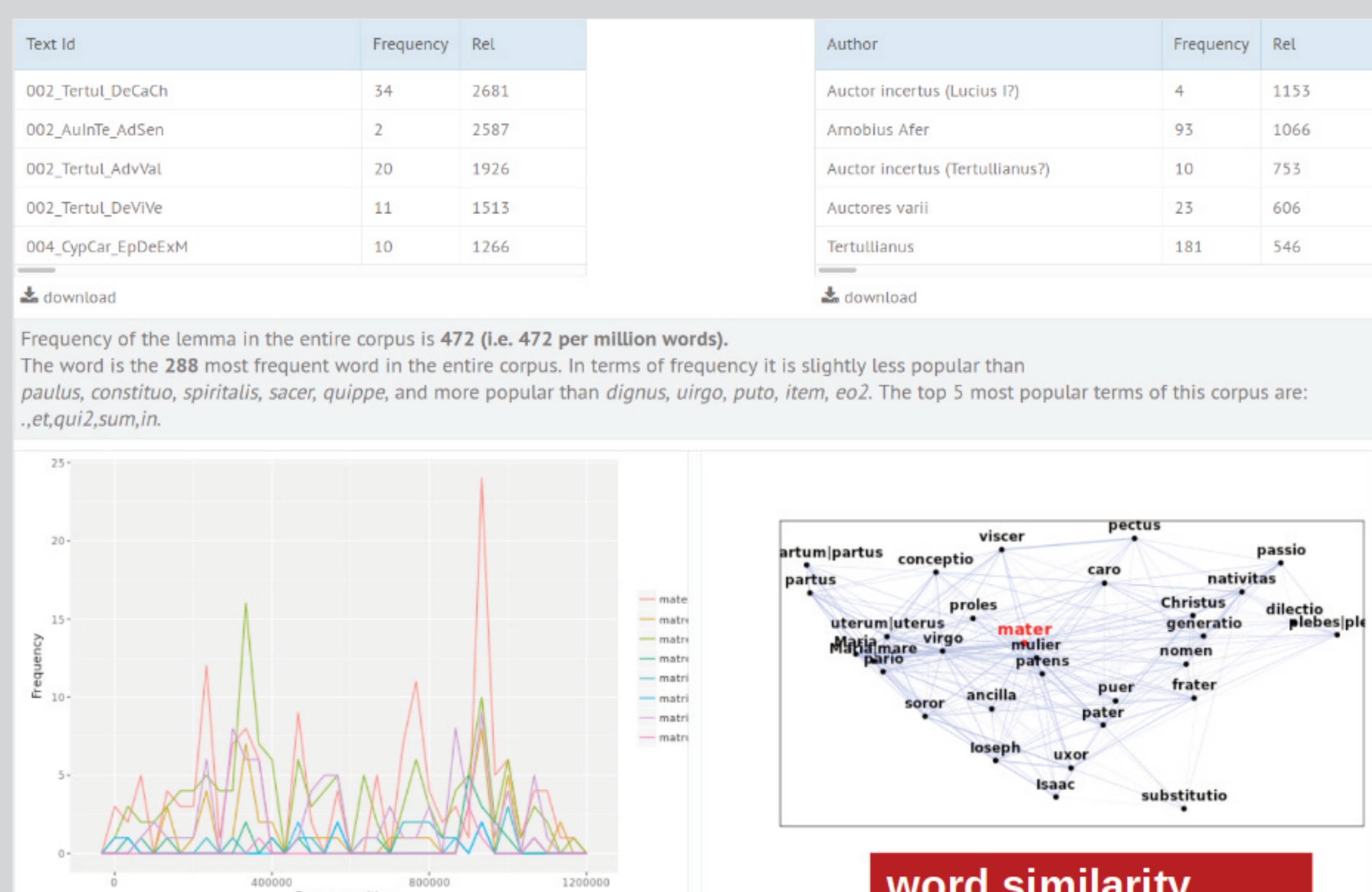- basic corpus stats may be inspected from the Corpus Info page

The Details Page

- **word distribution by authors and texts**
- **basic prose description of the word distribution**
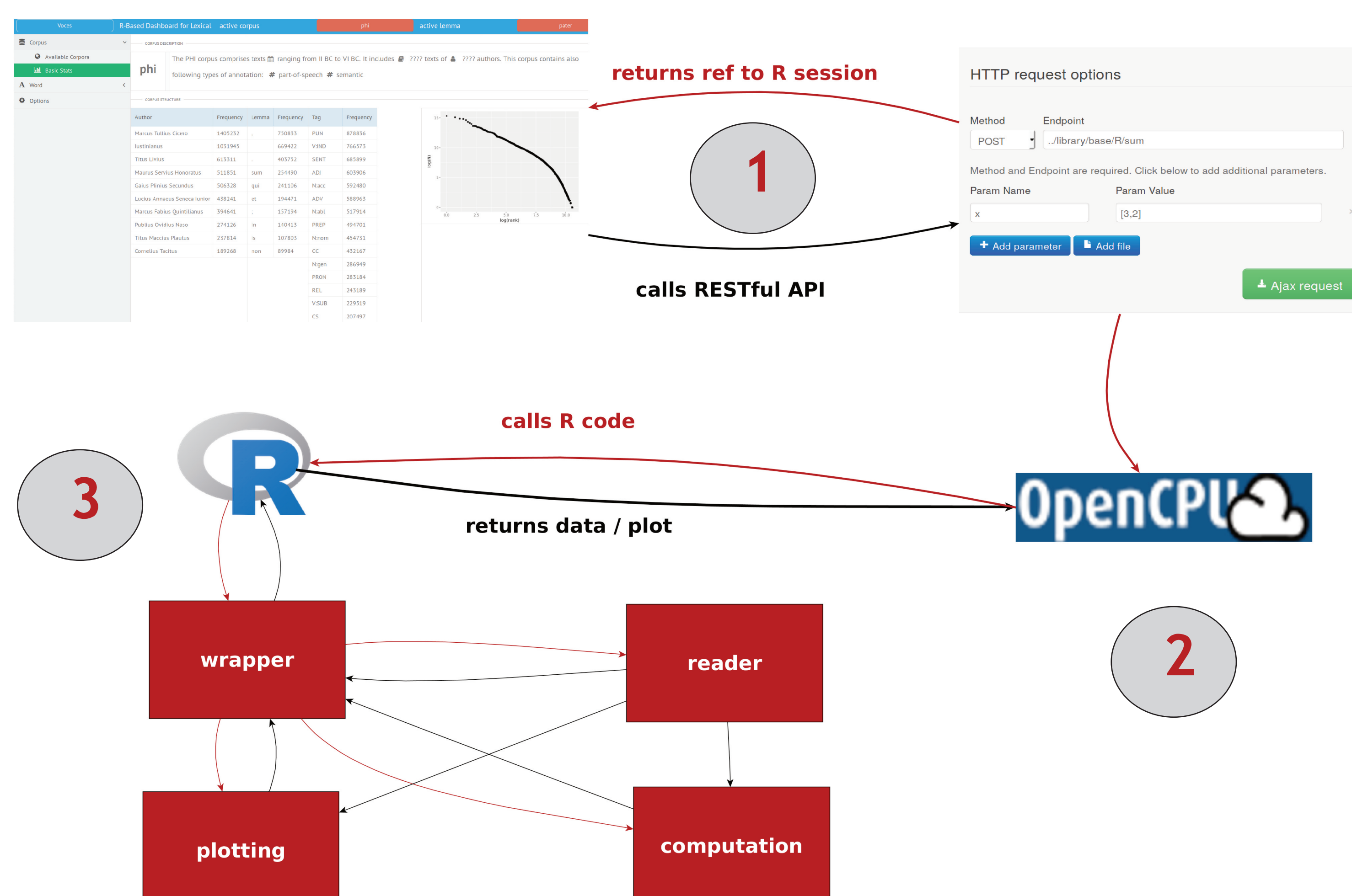- **wordform frequency by corpus position**
- **word similarity chart (pkg: wordspace)**

UNDERSTAND

text snippets explain where the data come from and how they have been computed

REUSE

both raw data and graphics can be downloaded

3. R Server

- the entire code is stored in a single package
- on session start data tables (word / lemma frequency etc.) are loaded into memory
- wrapper functions: process user's input, dispatch call to relevant functions, return results
- reader functions: communicate with NoSketchEngine, CWB etc. and process their output
- plotting and computation functions

How does it work?

returns ref to R session

① calls RESTful API

calls R code

returns data / plot

③ R

② OpenCPU

wrapper — reader — plotting — computation

1. Interface

- written in Webix (a JS UI library)
- communicates with a REST API
- processes output of the R session
- displays returned data as tables, d3.js charts etc.
- embeds R plots

2. OpenCPU

- communicates with the R server
- exposes an R session as REST API to the web application