RELATE

A portal for resources and technologies for Romanian Language

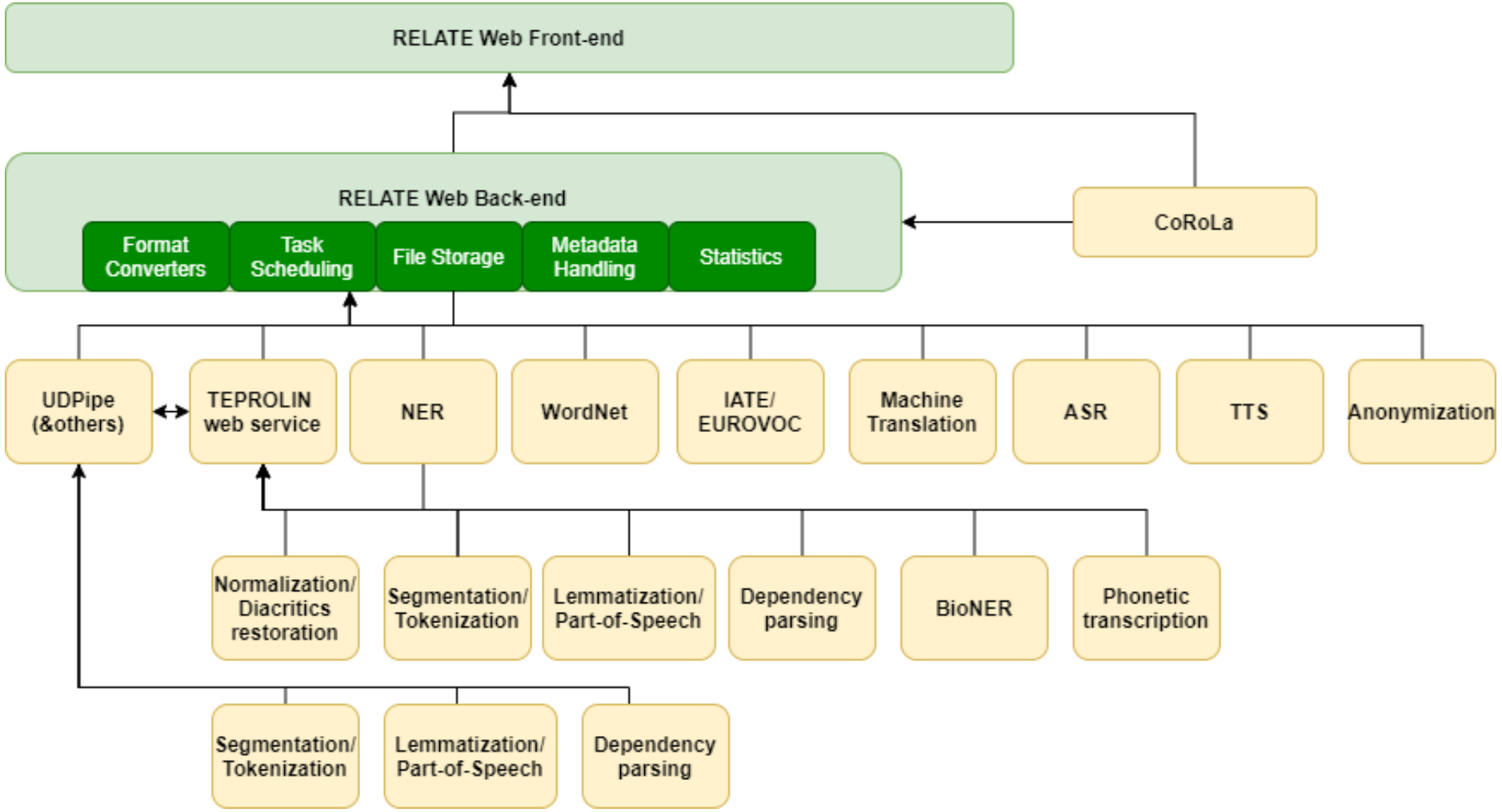
General characteristics

- Includes technologies and language resources developed by ICIA and its partners in several projects: COROLA, RETEROM, ROBIN, PRESIDENCY, MARCELL, CURLICAT
- It is aligned with the development philosophy of European Language Grid:

- WEB services, REST API, DOCKER
- The services may be distributed on multiple network nodes
- The services may be consumed directly from the partners

- RELATE has been used lately (2021) for deep processing of large and very large corpora (more than 200,000 documents) in our international running projects.
- It is a robust infrastructure, allowing for both CPU and GPU processing
- open source (https://github.com/racai-ai/RELATE)

Structural view of the RELATE



RELATE functional architecture

- **Web Front-end** (freely accessible).

- offers full set of 18 (for now) processing modules for text and speech data and various visualization modes *for single document only*

- **Web back-end** (accessible WITH Free access credentials)

- Corpora management: create, upload, download, archive, annotate, statistics, visualize, converting formats (CoNLL-U, CoNLL-U Plus, XML, JSON, RDF)

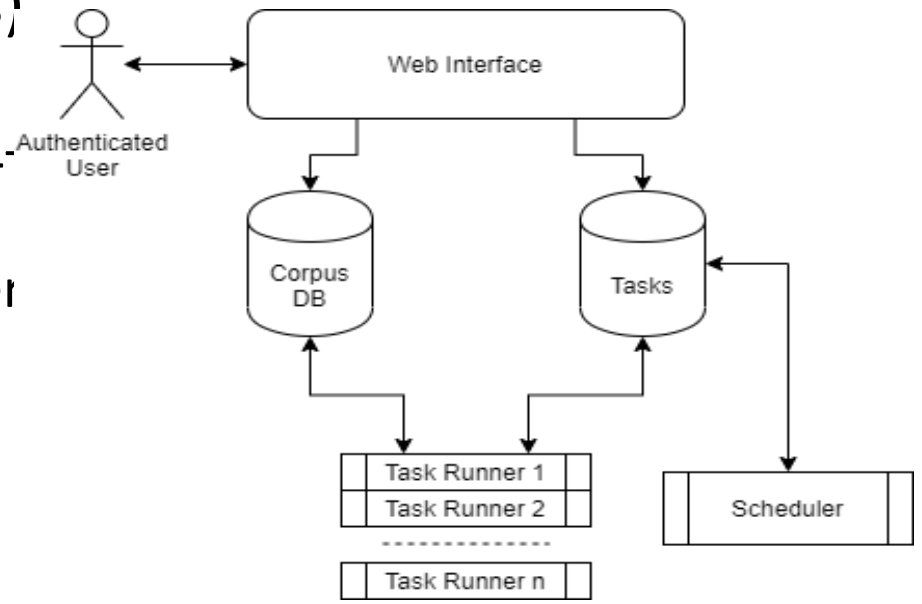
- Creation of gold corpora: integrates BRAT for NER, speech recorder for speech-text aligned corpora

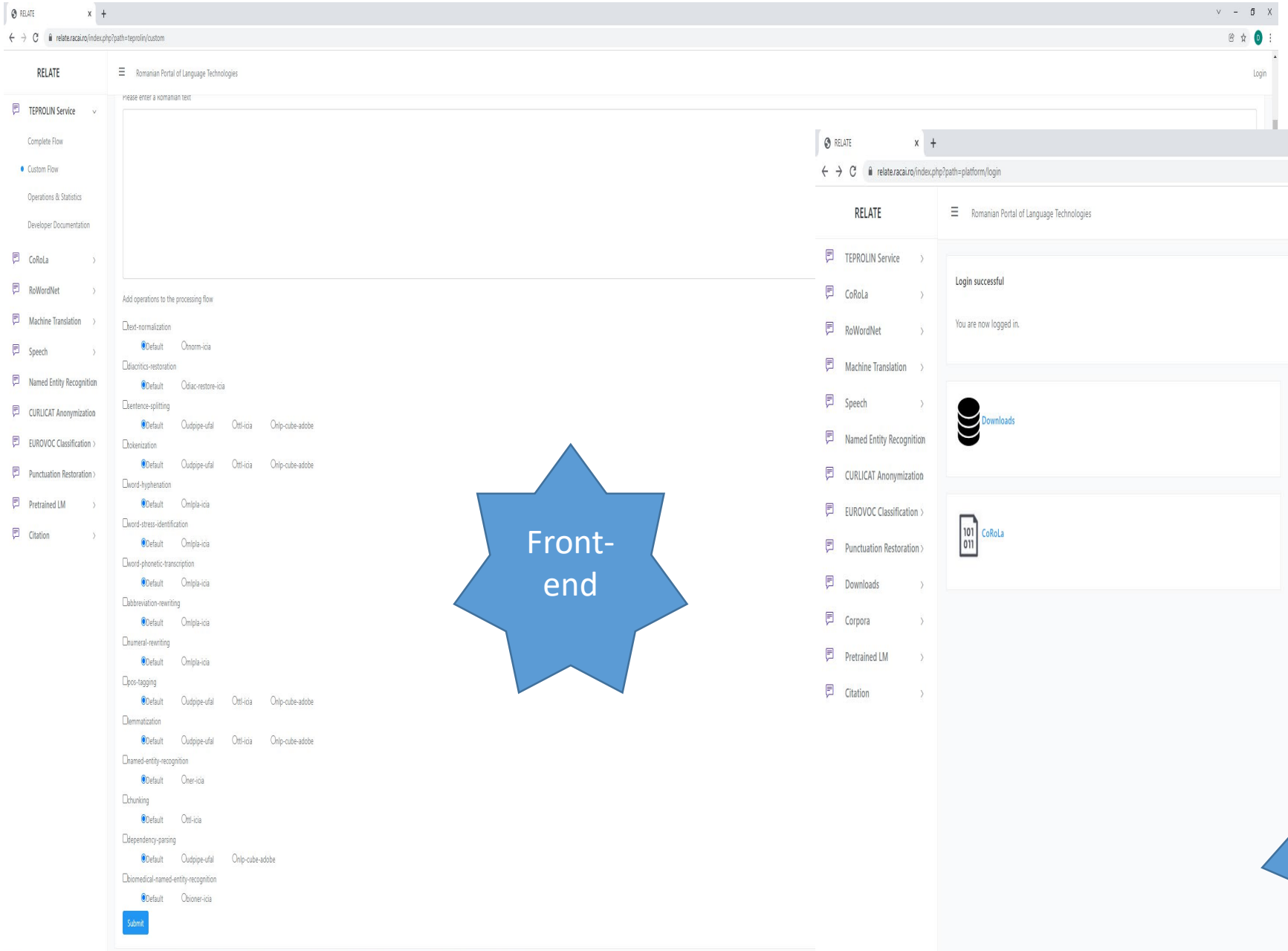
- offers full set of 18 (for now) processing modules for text and speech data and various visualization modes *for mass collections of documents*

- Offers large pre-trained language models

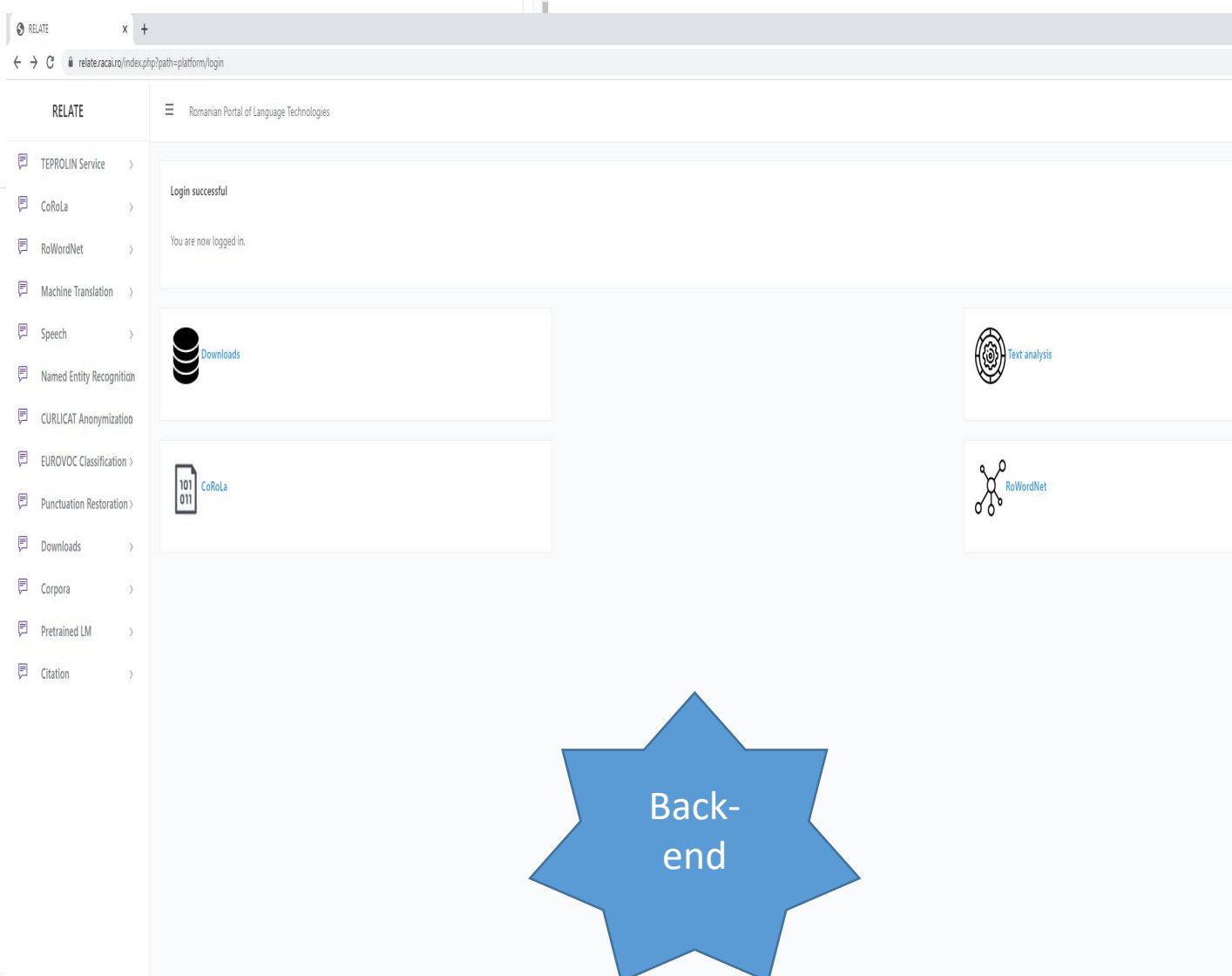
- Metadata management, statistics

- parallel processing (task scheduling, services from multiple nodes)





Front-end



Back-end

Large language models for Romanian available on RELATE

- Pre-Trained Language Models
  - RoBERT: There are two models available bert-base-romanian-cased-v1 and bert-base-romanian-uncased-v1 .
  - Romanian DistilBERT: Constructed based on the bert-base-romanian-cased-v1 model.
  - Word Embeddings from the CoRoLa project
- Annotation models for lemma, UPOS, XPOS and dependency parsing (where supported) trained on RRT UD 2.7.
- Classification models

Thank you!