



RELATE - O platformă modernă pentru procesarea limbii române contemporane

Vasile Păiș

Research Institute for Artificial Intelligence, Romanian Academy
vasile@racai.ro

13.07.2023

Overview

- Introduction
 - Platform Architecture
 - Language resources and pre-trained models
 - Features
- Text Processing
- Speech Processing
- Linked Data
- Development

RELATE - A platform for Romanian language technologies and resources

Includes technologies developed at ICIA and by Partners in multiple projects:

CoRoLa, ReTeRom, ROBIN, Presidency, MARCELL, CURLICAT,
Enrich4All, ELE, USPDATRO

Follows ELG philosophy:

- web services, REST APIs, dockers when possible
- services can be distributed across different physical nodes
- services can be consumed directly from partners

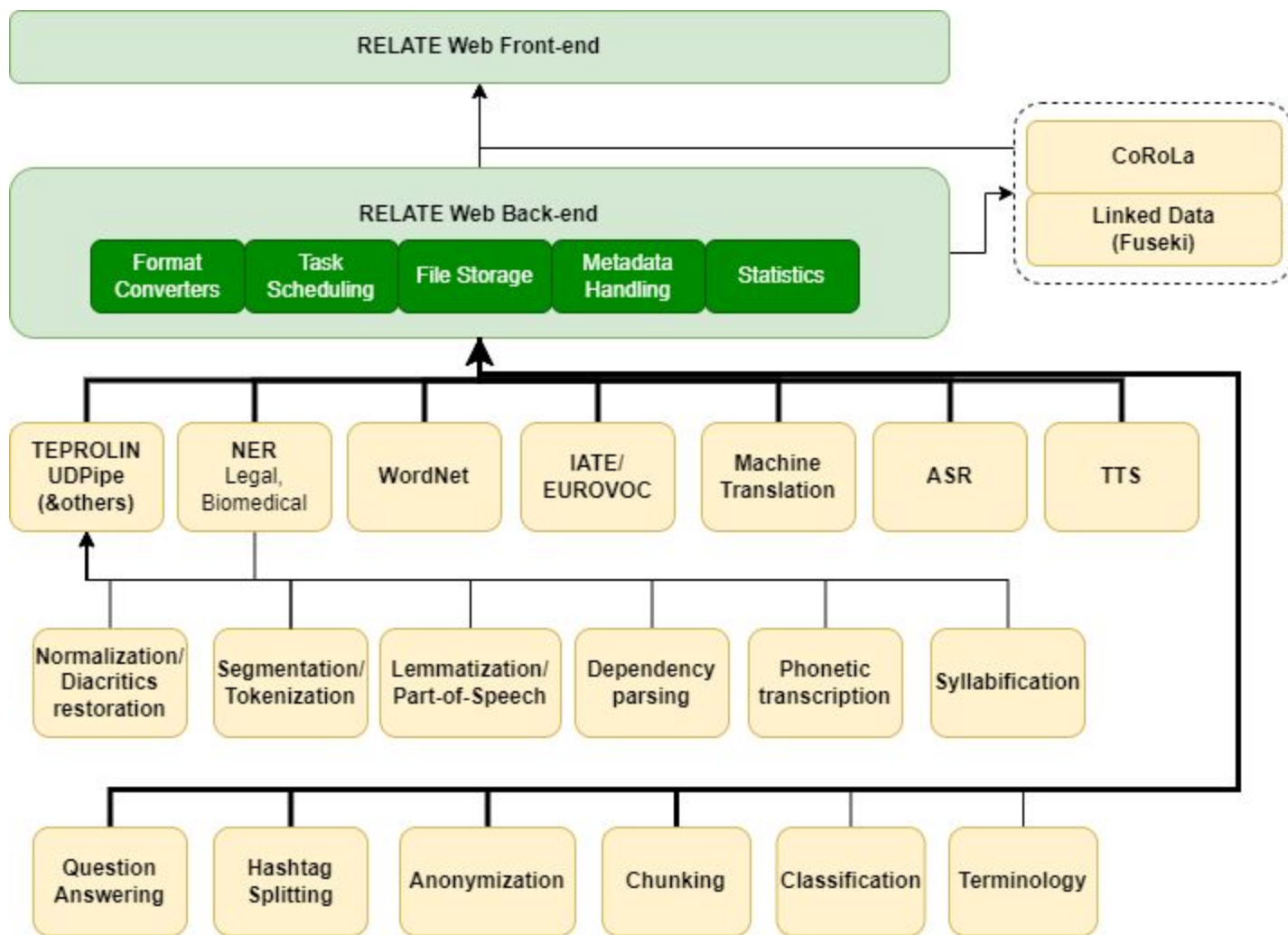
<https://relate.racai.ro>

RELATE - overview papers

Vasile Păiș, Radu Ion, and Dan Tufiș. “**A Processing Platform Relating Data and Tools for Romanian Language**”. English. In: Proceedings of the 1st International Workshop on Language Technology Platforms. Marseille, France: European Language Resources Association, 2020, pp. 81–88. URL: <https://www.aclweb.org/anthology/2020.iwltp-1.13>

Vasile Păiș. “**Multiple annotation pipelines inside the RELATE platform**”. In: The 15th International Conference on Linguistic Resources and Tools for Natural Language Processing. 2020, pp. 65–75. URL: <https://profs.info.uaic.ro/~consilr/wp-content/uploads/2021/03/volum-ConsILR-v-4-final-revizuit.pdf#page=73> .

Vasile Păiș, Dan Tufiș, and Radu Ion. “**Integration of Romanian NLP tools into the RELATE platform**”. In: International Conference on Linguistic Resources and Tools for Natural Language Processing. 2019, pp. 181–192. URL: https://profs.info.uaic.ro/~consilr/2019/wp-content/uploads/2020/01/ConsILR2019_final_BTT-60-ex-B5.pdf#page=189 .



Comparison of RO - BLARK

Table 2: Performance comparison of different models trained and evaluated against the RRT corpus for tokenization, sentence splitting, lemmatization, part-of-speech tagging.

	Tokens	Sentences	Lemma	UPOS	XPOS
	F1	F1	Acc	Acc	Acc
TreeTagger	99.72	98.37	89.62	95.58	92.97
TTL	99.66	96.19	95.84	95.54	95.18
UDPipe	99.88	97.39	95.91	97.15	96.24
NLP-Cube	99.86	98.62	51.39	97.73	96.94
RNNTagger	99.23	95.47	98.30	97.82	97.19
Stanza	99.94	98.42	98.90	97.64	97.07

Table 6: Training duration and running speed considering tokenization and complete annotation for RRT-Test and SiMoNEro

	RRT Train		RRT-test		RRT All		SiMoNEro	
	Train time	Tokenize tok/s	Run tok/s	Run tok/s	Train time	Tokenize tok/s	Run tok/s	
TreeTagger	70s	32648	8162		21s	43201	24337	
TTL	58s	2332	583		63s	279	202	
UDPipe	3h3m	9775	2720		3h15m	22024	2585	
NLP-Cube	13h20m	191	60		14h54m	188	64	
RNNTagger	1d3h44m	42958	281		1d8h46m	73010	453	
Stanza	4d16h8m	1484	430		5d2h10m	186	44	

Păiș, Vasile and Ion, Radu and Avram, Andrei-Marius and Mitrofan, Maria and Tufiș, Dan (2021). *In-depth evaluation of Romanian natural language processing pipelines*. In Romanian Journal of Information Science and Technology (ROMJIST). vol. 24, no. 4, pp. 384--401, <https://www.romjist.ro/full-texts/paper700.pdf>

	UPOS Acc	XPOS Acc
BERT-base-ro	98.01	96.43
RoBERT-small	97.52	96.01
RoBERT-base	98.02	97.18
RoBERT-large	98.10	97.65

Language resources and pre-trained models

RELATE

- TEPROLIN Service >
- CoRoLa >
- RoWordNet >
- Machine Translation >
- Speech >
- EUROVOC Classification >
- CURLICAT Anonymization
- Named Entity Recognition
- Punctuation Restoration >
- Social Media >
- Question Answering >
- Resources and Models ▾
 - Language Models
 - Language Resources
 - Repository**

☰ Romanian Portal of Language Technologies

Romanian Language Resources Repository

<< Showing 161 - 170 out of 215 >>

PyEuroVoc

Author(s):

Avram, Andrei-Marius; Păiș, Vasile; Tufiș, Dan

Description:

Classification of legal documents using EuroVoc descriptors, based on BERT models, for 22 languages (Bulgarian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Spanish, Slovak, Slovene, Swedish). A GitHub repo with scripts and example usage is available.

[View resource](#)

ro_sts

Description:

The RO-STs (Romanian Semantic Textual Similarity) dataset contains 8628 pairs of sentences with their similarity score. It is a high-quality translation of the STS benchmark dataset.

[View resource](#)

ro_sts_parallel

Description:

The RO-STs-Parallel (a Parallel Romanian English dataset - translation of the Semantic Textual Similarity) contains 17256 sentences in Romanian and English. It is a high-quality translation of the English STS benchmark dataset into

Search expression:

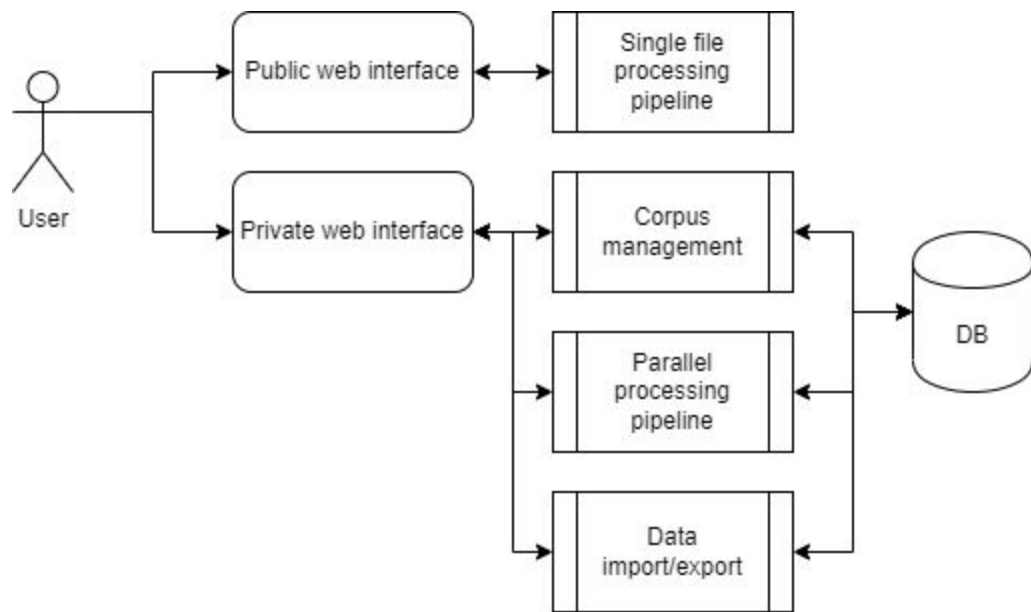
Resource type:

- Language Resource
- Language Model

Media type:

- Text
- Speech
- Image

Filter



RELATE - public view

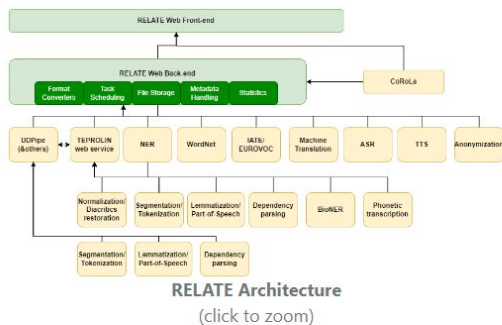
RELATE

☰ Romanian Portal of Language Technologies

Login

- TEPROLIN Service >
- CoRoLa >
- RoWordNet >
- Machine Translation >
- Speech >
- CURLICAT Anonymization >
- Social Media >
- Punctuation Restoration >
- Named Entity Recognition >
- EUROVOC Classification >
- Question Answering >
- Resources and Models >
- Citation >

RELATE - Romanian Portal of Language Technologies



RELATE is a Romanian language technology platform integrating different state-of-the-art tools, algorithms, models and language resources for processing the Romanian language. The modules are developed either in-house or by our partners in different research projects. Please check each page for appropriate references. The modular architecture (see the diagram) allows chaining the available modules into custom pipelines providing advanced language processing capabilities.

The platform allows direct interaction with Romanian language tools for annotating and processing data. For small data sizes it is possible to directly invoke the modules from the web interface in an interactive way. For larger data volumes, the internal platform components allow creating corpora of any size and execute parallel processing pipelines. The platform is open to the public for research purposes (including the internal part, following an account request). The platform is developed for research purposes and may not be suitable for any commercial or production use.

Platform development takes place at GitHub: <https://github.com/racai-ai/relate>

Featured components

TEPROLIN is a web service providing lemmatization, part-of-speech tagging, dependency parsing. The processing flow can be [customized](#) if needed.

RELATE - authenticated view

RELATE

☰ Romanian Portal of Language Technologies

Welcome  Vasile

- TEPROLIN Service >
- CoRoLa >
- RoWordNet >
- Machine Translation >
- Speech >
- CURLICAT Anonymization
- Social Media >
- Punctuation Restoration >
- Named Entity Recognition
- Downloads >
- EUROVOC Classification >
- Corpora >

• List

Corpora

Corpora list					
+ Add					
	Name	Lang	User	Description	Creation Date
1	coref_test	ro	pvf		2022-11-13
2	termeni nature lematizare	ro	elena@racai.ro		2022-10-09
3	enrich4all covid segmentare	ro	elena@racai.ro		2022-09-08
4	test_marcell_annoate_andrei	ro	andrei.avram		2022-08-10
5	test_marcell_annotate_andrei_v2	ro	andrei.avram		2022-08-10
6	TWITTER_ANONYMIZED_20220725	ro	pvf	Versiunea anonimata (finala)	2022-07-25
7	TWITTER_MERGED_20220712	ro	pvf		2022-07-12
8	TWITTER Corpus 20220602	ro	pvf		2022-06-02
9	LegalNERo_zenodo_v3	ro	pvf	Textul LegalNERo pentru adnotare c...	2022-03-29
10	BioRo_Extended	ro	maria@racai.ro		2022-02-14
11	corpus_punct	ro	pvf		2022-01-28
12	corpus_wav	ro	pvf		2022-01-28

Page 1 of 4 20 1 to 20 of 67

RELATE - Large corpora

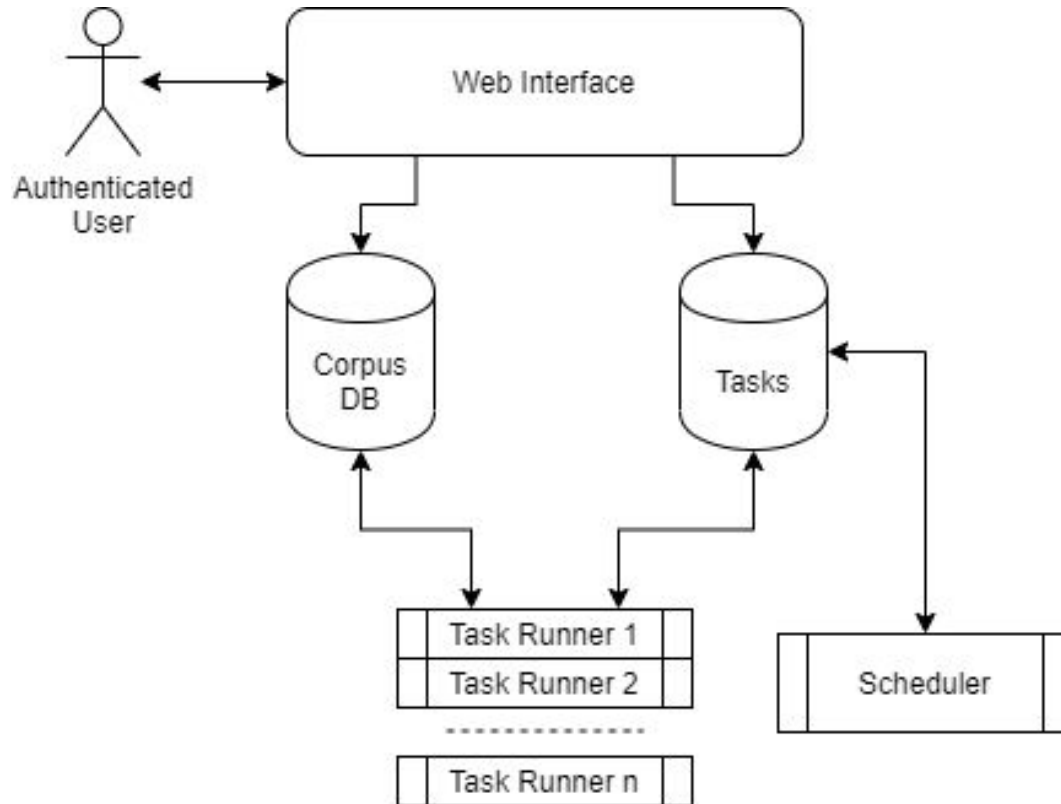
Efficiently stores and handles large corpora:

- text only
- text + audio
- text + PDF
 - + metadata
 - in the future => other types of bimodal/multimodal data

Parallel processing

- task scheduling
- services from multiple nodes

RELATE Task-based processing



RELATE Features

- **Corpora management:** create, upload, download, archive, annotate, statistics, visualize
- **Creation of gold corpora:** integrates BRAT for NER and other tasks, speech recorder for speech-text aligned corpora
- **Annotation:** lemma, part-of-speech, dependency parsing, syllabification, phonetic transcription, diacritics restoration, NER (Legal/BioNER), IATE, EuroVoc, generic terminology
- **Machine translation:** text-to-text, speech-to-speech
- **Speech:** ASR, TTS, speech-to-speech translation
- **WordNet interface:** RoWordNet, aligned query with EnWordNet
- **CoRoLa interface:** Korap, speech component, word embeddings

RELATE Features

- **Anonymization:** from the CURLICAT project
 - **Legal NER:** Person, Organization, Location, Time, Legal references (from a gold annotated sub-corpus derived from the MARCELL corpus)
 - **Pre-trained language models:** available for download
 - **Standard formats:** CoNLL-U, CoNLL-U Plus, XML, JSON, RDF
-
- **A modular architecture which continues to expand with new projects**

Text processing

- Gold corpora creation
- Task-based processing
- Standoff Metadata
- Import/Export

Gold corpora creation - NER - MicroBloggingNERo

RELATE

TEPROLIN Service >

CoRoLa >

RoWordNet >

Machine Translation >

Speech >

EUROVOC Classification >

Downloads >

Punctuation Restoration >

Romanian Portal of Language Technologies

Welcome Vasile

Corpus: TWITTER_2 View Corpora

Back Download Download ANN Save To GOLD PREVIOUS NEXT

Language Sentiment Hate Language Type Save

To see annotations from CONLLUP files make sure to run the CONLLUP2BRAT task.

/TWITTER_2/147179008022244868.txt brat

1 Scaunul lui **PER** Boris Johnson se clatină. Conservatorii pierd o circumscripție pe care o controlau de două secole **TIME** - <url> <url>

- Păiș, Vasile and Mitrofan, Maria and Barbu-Mititelu, Verginica and Irimia, Elena and Gasan, Carol Luca and Micu, Roxana and Marin, Laura and Dicusar, Maria and Florea, Bianca and Badila, Ana (2022). *Romanian micro-blogging named entity recognition (MicroBloggingNERo)*. Zenodo, <https://doi.org/10.5281/zenodo.6905235>
- Păiș, Vasile and Barbu Mititelu, Verginica and Irimia, Elena and Mitrofan, Maria and Gasan, Carol Luca and Micu, Roxana (2022). *Romanian micro-blogging named entity recognition including health-related entities*. In Proceedings of The Seventh Workshop on Social Media Mining for Health Applications, pp. 190--196, <https://aclanthology.org/2022.smm4h-1.49>

Gold corpora creation - NER - LegalNERo

RoWordNet >

Machine Translation >

Speech >

Downloads >

EUROVOC Classification >

CURLICAT Anonymization >

Corpora >

Pretrained LM >

Citation >

MY >

Back Download Download ANN Save To GOLD

To see annotations from CONLLUP files make sure to run the CONLLUP2BRAT task.

/MARCELL_NER/mj_00000G1TVLRU8DK2LRS0OL3XTVLNBM9Q.txt brat

1 ORDIN nr. 533 din 30 octombrie 2019

2 privind înființarea Comisiei de concesionare a terenurilor pe care sunt amplasate amenajările piscicole, precum și a altor terenuri aferente acestora din domeniul public și privat al statului și de aprobare a Normelor referitoare la procedura privind concesionarea terenurilor pe care sunt amplasate amenajările piscicole și a altor terenuri aferente acestora din domeniul public și privat al statului

3 Publicat în

5 MONITORUL OFICIAL nr. 886 din 4 noiembrie 2019

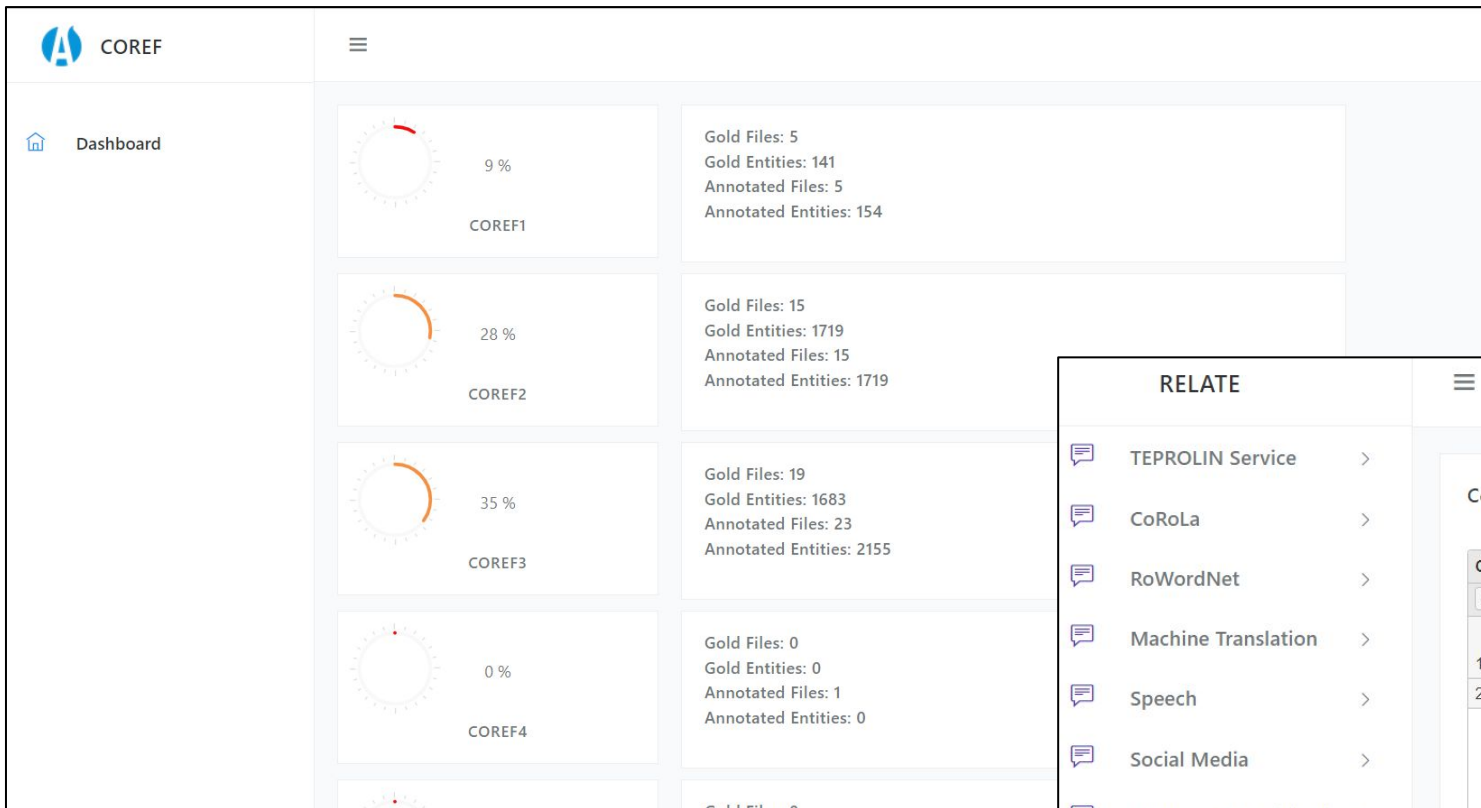
7 Având în vedere Referatul de aprobare nr. 10.064 din 30 septembrie 2019 al Agenției Naționale pentru Pescuit și Acvacultură,

8 -

9 art. 4 alin. (3) lit. b)-g) și alin. (5) din Ordonanța de urgență a Guvernului nr. 23 / 2008 privind pescuitul și acvacultura, aprobată cu modificări și completări prin Legea nr. 317 / 2009, cu modificările și completările ulterioare;

- Păiș, Vasile and Mitrofan, Maria and Gasan, Carol Luca and Ianov, Alexandru and Ghiță, Corvin and Coneschi, Vlad Silviu and Onuț, Andrei (2021). *Romanian Named Entity Recognition in the Legal domain (LegalNERo)*. Zenodo, <https://doi.org/10.5281/zenodo.4772094>
- Păiș, Vasile and Mitrofan, Maria. (2021) *Towards a named entity recognition system in the Romanian legal domain using a linked open data corpus*. In Workshop on Deep Learning and Neural Approaches for Linguistic Data, pp. 16--17.
- Păiș, Vasile and Mitrofan, Maria and Gasan, Carol Luca and Coneschi, Vlad and Ianov, Alexandru (2021). *Named Entity Recognition in the Romanian Legal Domain*. In Proceedings of the Natural Legal Language Processing Workshop, pp. 9--18. <https://aclanthology.org/2021.nllp-1.2>

Gold corpora creation - Annotator Instances + Dashboard



The RELATE interface includes a list of services and a Corpora list table:

- TEPROLIN Service >
- CoRoLa >
- RoWordNet >
- Machine Translation >
- Speech >
- Social Media >
- EUROVOC Classification >

Corpora

Corpora list


+ Add

	Name	Lang
1	COREF1	ro
2	playground	ro

Multiple NER models

RELATE

☰ Romanian Portal of Language Technologies

Welcome  Vasile

- TEPROLIN Service >
- CoRoLa >
- RoWordNet >
- Machine Translation >
- Speech >
- CURLICAT Anonymization
- Social Media >
- Punctuation Restoration >
- Named Entity Recognition

Named Entity Recognition

Selectați modelul NER dorit și introduceți un document în limba română pentru recunoașterea entităților. Modelele bazate pe corpusul LegalNERo presupun un text de tip legislație. Modelele bazate pe corpusul SiMoNERo presupun un text din domeniul biomedical. Utilizarea de texte din alte domenii reduce calitatea rezultatelor. Pentru a putea realiza ancorarea corectă a rezultatelor în textul introdus, nu sunt realizate curățări ale datelor. Textul trebuie să conțină doar caractere în limba română, cu virgula și punctul separate de cuvântul anterior.

LegalNERo LEGAL, PER, LOC, ORG, TIME with Gazetteer ▾
LegalNERo LEGAL, PER, LOC, ORG, TIME with Gazetteer
LegalNERo PER, LOC, ORG, TIME with Gazetteer
LegalNERo LEGAL, PER, LOC, ORG, TIME
LegalNERo PER, LOC, ORG, TIME
SiMoNERo ANAT, CHEM, DISO, PROC

- Păiș, Vasile (2022). RACAI at SemEval-2022 Task 11: Complex named entity recognition using a lateral inhibition mechanism. In Proceedings SemEval, pp. 1562--1569, <https://aclanthology.org/2022.semeval-1.215>
- Mitrofan, Maria and Păiș, Vasile (2022). Improving Romanian BioNER Using a Biologically Inspired System. In Proceedings of the 21st Workshop on Biomedical Language Processing, pp. 316--322, <https://aclanthology.org/2022.bionlp-1.30>

Task-based processing

Corpus: Marcell

View Corpora

Files

Standoff

Tasks

Annotated

Statistics

Archives

Files list

+ Add TEXT + Add CSV/TSV + Add PDF + Add ZIP + Access Last File

	Name ▲	Type	Description	User	Creation Date
1	mj_00000G0001I4Q9GISD13TF7D...	text		pvf	2019-10-23 15:04:43
2	mj_00000G0001L9DUTVXXJ37F4F...	text		pvf	2019-10-23 15:04:43
3	mj_00000G0002CA2MKRE8V1QE...	text		pvf	2019-10-23 15:04:43
4	mj_00000G0002I4DEC59G23QLY...	text		pvf	2019-10-23 15:04:43
5	mj_00000G0005HOC5GKLDY0DN...	text		pvf	2019-10-23 15:04:43
6	mj_00000G0007WC5AJK76I255P9...	text		pvf	2019-10-23 15:04:43
7	mj_00000G000827N7KZV7704W5...	text		pvf	2019-10-23 15:04:43
8	mj_00000G0008L3G0KSO062ZUJ...	text		pvf	2019-10-23 15:04:43
9	mj_00000G0008LFFWVAOBH37E...	text		pvf	2019-10-23 15:04:43
10	mj_00000G0008NVRHIFAVJ3R6X...	text		pvf	2019-10-23 15:04:43
11	mj_00000G000BRZ93RU63D3NY0...	text		pvf	2019-10-23 15:04:43
12	mj_00000G000D2PI9LA0VR3NB3T...	text		pvf	2019-10-23 15:04:43

Page 1 of 7207 20 1 to 20 of 144131

MARCELL-RO Corpus

- Part of the CEF Telecom project Multilingual Resources for CEF.AT in the legal domain (MARCELL)
- 7 legislative corpora containing the total body of national legislative documents in effect for 7 countries included in the consortium: Bulgaria, Croatia, Hungary, Poland, Romania, Slovakia and Slovenia
- Tokenized, lemmatized, morphologically annotated, dependency parsed, named entities, nominal phrases, term annotation with IATE and EuroVoc.
- annotated in the RELATE platform
- statistics computed in the platform (4.3 million sentences, 144k documents, 67 million tokens)
- exported as archives (text, conllu plus and xml) from the platform

Task-based processing

Corpus: Marcell

View Corpora

Files

Standoff

Tasks

Annotated

Statistics

Archives

Corpus tasks

Type	Status	Description	User	Creation Date
1	DONE		pvf	2021-03-12 19:03:48
2	DONE		pvf	2021-03-12 15:40:04
3	DONE		pvf	2021-03-12 11:09:05

- Tamás Váradi, Svetla Koeva, Martin Yamalov, Marko Tadić, Bálint Sass, Bartłomiej Nitoń, Maciej Ogrodniczuk, Piotr Pezik and Barbu Mititelu, Verginica and Radu Ion and Irimia, Elena and Mitrofan, Maria and Păiș, Vasile and Tufiș, Dan and Radovan Garabik, Simon Krek, Andraz Repar, Matjaž Rihtar and Janez Brank (2020). *The MARCELL Legislative Corpus*. In Proceedings of The 12th LREC, pp. 3754-3761. <https://www.aclweb.org/anthology/2020.lrec-1.464/>
- Tufiș, Dan and Mitrofan, Maria and Păiș, Vasile and Ion, Radu and Coman, Andrei (2020). *Collection and Annotation of the Romanian Legal Corpus*. In Proceedings of The 12th LREC, pp. 2766-2770. <https://www.aclweb.org/anthology/2020.lrec-1.337/>
- <https://elrc-share.eu/repository/search/?q=marcell>

Task-based processing

Files Standoff **Tasks** Annotated Statistics Archives

Corpus tasks

+ TEPROLIN	+ UDPipe	+ NERRegex	+ ...
+ Gold NE list	+ Statistics	+ ZIP Text	+ Z...
+ Export Marcell	+ NER-Legal	+ Translate	

	Type	Sta
1	ner_regex	DC
2	brat2conllup	DC
3	zip_standoff	DC
4	zip_basic_tagging	DC
5	zip_text	DC
6	terminology	DC
7	ner_regex	SU
8	brat2conllup	SU
9	ner_legalner	DC
10	ner_legalner	ER
11	statistics	DC

Page 1 of 3

Add task Terminology

Annotate text a custom terminology. The text must be already tokenized and lemmatized. At least one file with .terminology extension must be present in the corpus standoff folder.

Column: RELATE:TERM

Max Term Size: 10

Terminology: custom.terminology (selected), IATE.terminology, custom.terminology

Description:

Runners (optional):

Overwrite:

Add Cancel

NER_Old	+ TTLChunker	+ BRAT2CoNLLUP
Marcell	+ Export CURLICAT	
nymization	+ Create Empty Metadata	
Creation Date		
2022-11-01 09:56:34		
2022-10-31 13:37:07		
2022-10-30 17:12:46		
2022-10-30 17:12:26		
2022-10-30 17:12:08		
2022-10-30 15:26:24		
2022-10-30 10:46:38		
2022-10-30 10:46:06		
2022-09-27 05:30:59		
2022-09-26 11:27:43		
2022-09-26 11:18:59		

- Váradi, Tamás and Nyéki, Bence and Koeva, Svetla and Tadić, Marko and Štefanec, Vanja and Ogrodniczuk, Maciej and Nitoń, Bartłomiej and Pezik, Piotr and Barbu Mititelu, Verginica and Irimia, Elena and Mitrofan, Maria and Păiș, Vasile and Tufiș, Dan and Garabík, Radovan and Krek, Simon and Repar, Andraž (2022). **Introducing the CURLICAT Corpora: Seven-language Domain Specific Annotated Corpora from Curated Sources**. In Proceedings LREC, pp. 100--108. <https://aclanthology.org/2022.lrec-1.11/>
- <https://elrc-share.eu/repository/search/?q=curlicat>

CURLICAT-RO Corpus

- Part of the CEF Telecom project Curated Multilingual Resources for CEF.AT(CURLICAT)
 - an ongoing project (will finish this year)
 - 7 monolingual corpora containing texts from the following fields: culture, education, economy, health, nature, politics, science
 - Bulgaria, Croatia, Hungary, Poland, Romania, Slovakia and Slovenia
 - Anonymized, tokenized, lemmatized, morphologically annotated, dependency parsed, named entities, nominal phrases, term annotation with IATE and automatic term recognition.
-
- annotated in the RELATE platform
 - statistics computed in the platform (4.4 million sentences)
 - will be exported as archives (text, conllu plus) from the platform

Task-based processing - Statistics

- Files
- Standoff
- Tasks
- Annotated
- Statistics
- Archives

Statistics

+ Download Stats + View WordForm Stats + View Lemma Stats + View WordForm Doc Freq + View Letters Stats + View Lemma UPOS Stats + View IATE Stats + View IATE Doc Freq

+ View EUROVOC ID Stats + View EUROVOC ID Doc Freq + View EUROVOC MT Stats + View EUROVOC MT Doc Freq

	Key	Value
	<input type="text"/>	
1	Basic.Dis Legomena	133276
2	Basic.Hapax Legomena	653298
3	Basic.Number of Annotated Documents	26477
4	Basic.Number of Characters	559682566
5	Basic.Number of EUROVOC IDs	0
6	Basic.Number of EUROVOC MTs	0
7	Basic.Number of IATE terms	18772350
8	Basic.Number of Lines	3390728
9	Basic.Number of Raw Documents	26477
10	Basic.Number of Romanian letters	426159374
11	Basic.Number of Sentences	3610766

Page 1 of 30 20 1 to 20 of 583

Standoff Metadata

Back

PREVIOUS

NEXT

Peștera Proserv SA / 205 3.3. Parodia literară și asumarea modern
Capitolul 4. Parodii românești ale secolului XX: de la practicile
Drumul spre postmodernism/ 249 4.2. Postmodernismul: un construct
Concluzii / 329
Bibliografie / 337 Abstract / 343 Apălimăriței / 347
Cuvânt înainte
Parodia literară: șapte rescrieri românești reprezintă, în peisaj
Autoarea încearcă să răspundă unei duble provocări. de ordin teor

1021_a_2529.xml

Save

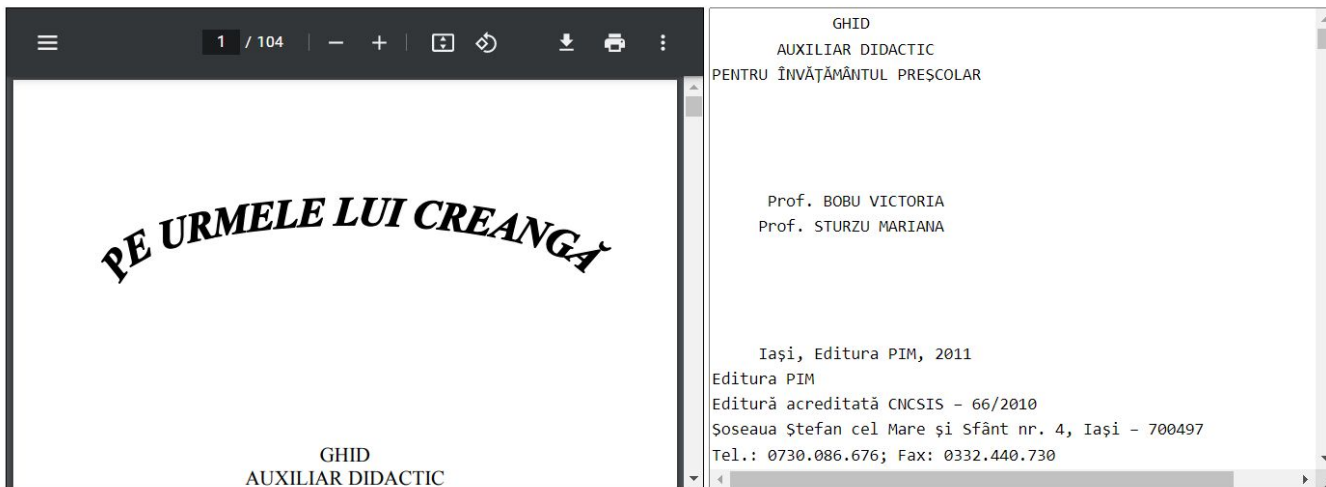
```
<?xml version="1.0" encoding="UTF-8"?>
  <Metadata>
    <DocumentTitle>Parodia literară. Șapte rescrieri românești</DocumentTitle>
    <ArticleTitle></ArticleTitle>
    <AuthorName>Livia Iacob</AuthorName>
    <PublicationDate>2011</PublicationDate>
    <Source>Publishing House</Source>
    <SourceName>Editura Europeană</SourceName>
    <TranslatorName></TranslatorName>
```

1021_a_2529.ann

Save

T1	LOC 125 132	România
T2	LOC 435 442	România
T3	ORG 857 873	Mahtab Impex SRL

Standoff Metadata - CoRoLa



Pe urmele lui Creanga FINAL.xml

Save

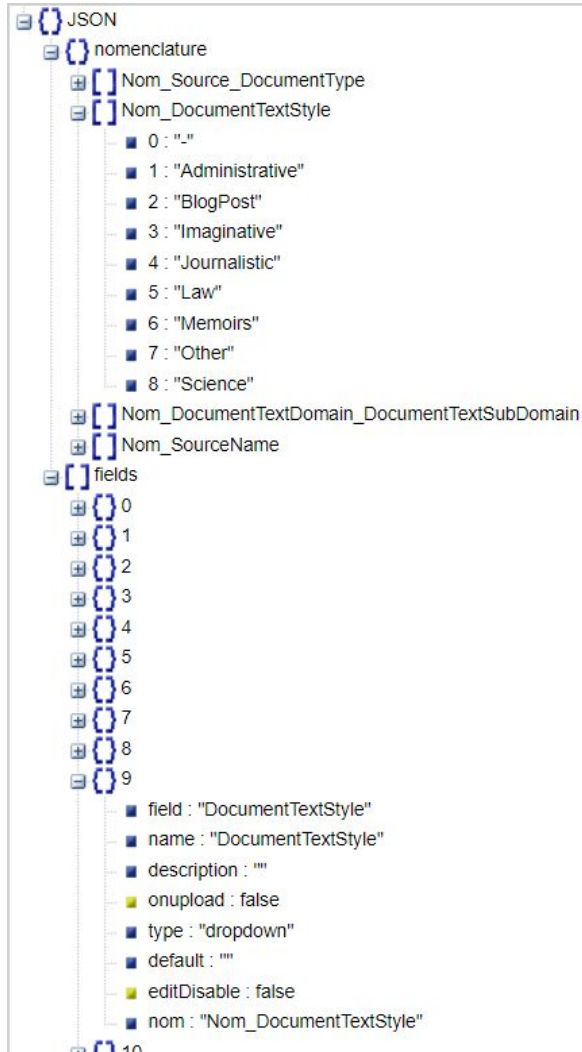
DocumentTitle	<input type="text" value="Pe urmele lui Creangă"/>	Titlul documentului. Pentru o colecție: 'Titlul cărții/[Volumul I, vol. I]:Titlul volumului'. Exemplu: 'Chirurgie generală/Volumul II, ediția a II-a' sau 'Moromeții/vol. I' (Se va trece 'Volumul' sau 'Vol.' după cum este în documentul electronic)
ArticleTitle	<input type="text" value="Pe urmele lui Creangă"/>	Pentru cărți introduceți valoarea din DocumentTitle
AuthorName	<input type="text" value="Bobu Victoria, Sturzu Mar"/>	Autorul sau autorii
TranslatorName	<input type="text" value="-"/>	Numele traducătorului sau '-'
PublicationDate	<input type="text" value="2011"/>	Anul de publicare
Source	<input type="text" value="Publishing House"/>	

Standoff Metadata - CoRoLa

Pe urmele lui Creanga FINAL.xml

Save

DocumentTitle	<input type="text" value="Pe urmele lui Creangă"/>	Titlul documentului. Pentru o colecție: 'I' ediția a II-a' sau 'Moromeții/vol. I' (Se va
ArticleTitle	<input type="text" value="Pe urmele lui Creangă"/>	Pentru cărți introduceți valoarea din Doc
AuthorName	<input type="text" value="Bobu Victoria, Sturzu Mar"/>	Autorul sau autorii
TranslatorName	<input type="text" value="-"/>	Numele traducătorului sau '-'
PublicationDate	<input type="text" value="2011"/>	Anul de publicare
Source	<input type="text" value="Publishing House"/>	
SourceName	<input type="text" value="Editura Pim"/>	
Medium	<input type="text" value="Written"/>	
DocumentType	<input type="text" value="Book"/>	
DocumentTextStyle	<input type="text" value="Other"/>	
DocumentTextDomain	<input type="text" value="Arts and Culture"/>	
DocumentTextSubDomain	<input type="text" value="Literature"/>	
SubjectLanguage	<input type="text" value="Romanian"/>	Dacă documentul este o traducere se tre
ISSN-ISBN	<input type="text" value="978-606-13-0496-7"/>	Dacă exista 2 ISBN, se prefera ISBN 13
CollectionDate	<input type="text" value="2023"/>	Anul în care se lucrează documentul



Import/Export



Files list

+ Add TEXT + Add CSV/TSV + Add PDF + Add ZIP + Access Last File

Name ▲	Type	D
--------	------	---

Import/Export

Corpus: test_COROLA2

View Corpora

Files

Add ZIP archive

Files will be distributed and processed according to their extension. Text from PDF files will be automatically extracted. XML, JSON files and unknown extensions will be considered standoff metadata.

File: No file chosen

FileName (optional):

Description:

Source

SourceName

SubjectLanguage Dacă documentul este o traducere se trece limba autorului, în caz contrar este limba română

CollectionDate Anul în care se lucrează documentul

Files list

+ Add TEX

Name
1 9942-p
2 Pe urn
3 Surasu
4 Surasu
5 aforism
6 dans.b
7 periplu
8 teoria

Double click

Import/Export

Add File CSV/TSV ✕

File: No file chosen

FileName (optional):

Description:

Delimiter: (For TAB enter '\t')

Enclosure:

Escape:

Comment character:

Ignore rows: (headers)

Columns w text: (comma separated list, zero based)

- CSV/TSV for small texts
- each line is processed as an independent text document

Import/Export

- Files
- Standoff
- Tasks
- Annotated
- Statistics
- Archives
- Gold Standoff
- Gold Annotated

Corpus tasks

+ TEPROLIN	+ UDPipe	+ NERRegex	+ Classify EuroVoc	+ Terminology	+ IATE/EuroVoc	+ Cleanup	+ NER Baseline	+ CoNLLUP2BRAT	+ NER_Old	+ TTLChunker	+ BRAT2CoNLLUP
+ Gold NE list	+ Statistics	+ ZIP Text	+ ZIP Annotated	+ ZIP Standoff	+ ZIP Gold Standoff	+ ZIP Gold Annotated	+ ZIP Audio	+ Change Terms Marcell	+ Export CURLICAT		
+ Export Marcell	+ NER-Legal	+ Translate Text	+ ASR	+ Punctuation	+ TTS	+ Translate S2S	+ DeAnonymization	+ NER-SiMoNERo	+ Anonymization	+ Create Empty Metadata	

Type	Status	Description	User	Creation Date
------	--------	-------------	------	---------------

Import/Export

Corpus: **CURLICAT_Anonymized**

[View Corpora](#)

[Files](#)

[Standoff](#)

[Tasks](#)

[Annotated](#)

[Statistics](#)

[Archives](#)

Archives

	File	Size
	<input type="text"/>	
1	zip_text/CURLICAT.zip	266.18 Mb
2	zip_text/20220312_text_anonymized.zip	298.39 Mb
3	zip_text/meta.zip	17.57 Mb
4	zip_text/20221030_text.zip	242.9 Mb
5	zip_basic_tagging/20211227_annotated_before_chunking.zip	1.86 Gb
6	zip_basic_tagging/20211231_annotated_before_IATE.zip	2.09 Gb
7	zip_basic_tagging/annotated_fara_iate_20220321.zip	2.16 Gb
8	zip_basic_tagging/20220926_annotated.zip	1.8 Gb
9	zip_basic_tagging/20221030_annotated.zip	1.82 Gb
10	zip_standoff/20220114_standoff.zip	14 Mb
11	zip_standoff/metadate_curlicat.zip	14 Mb
12	zip_standoff/20221030_meta.zip	56.3 Mb
13	curlicat-out/ro-raw.zip	207 Mb
14	curlicat-out/ro-annotated.zip	1.23 Gb
15	curlicat-out/ro-xml.zip	1.79 Gb

Import/Export

- Files
- Standoff
- Tasks
- Annotated
- Statistics
- Archives

Corpus tasks ↻ 🔊

+ TEPROLIN	+ UDPipe	+ NERRegex	+ Classify EuroVoc	+ Terminology	+ IATE/EuroVoc	+ Cleanup	+ NER Baseline	+ CoNLLUP2BRAT	+ NER_Old	+ TTLChunker	+ BRAT2CoNLLUP
+ Gold NE list	+ Statistics	+ ZIP Text	+ ZIP Annotated	+ ZIP Standoff	+ ZIP Gold Standoff	+ ZIP Gold Annotated	+ ZIP Audio	+ Change Terms Marcell	+ Export CURLICAT		
+ Export Marcell	+ NER-Legal	+ Translate Text	+ ASR	+ Punctuation	+ TTS	+ Translate S2S	+ DeAnonymization	+ NER-SiMoNERo	+ Anonymization	+ Create Empty Metadata	

Type	Status	Description	User	Creation Date ▾
------	--------	-------------	------	-----------------

Import/Export

Corpus: **Marcell**

[View Corpora](#)

[Files](#)

[Standoff](#)

[Tasks](#)

[Annotated](#)

[Statistics](#)

[Archives](#)

Archives

	File	Size
1	zip_text/marcell_1990.zip	525.21 Mb
2	zip_basic_tagging/tagged_20191029.zip	556.54 Mb
3	zip_basic_tagging/tagged_20191221.zip	4.34 Gb
4	zip_basic_tagging/output2020.zip	3.14 Gb
5	zip_basic_tagging/marcell_annotated_clean_20200114.zip	1.25 Gb
6	zip_basic_tagging/marcell_annotated_clean_20200120.zip	1.26 Gb
7	zip_basic_tagging/marcell_annotated_clean_20200122.zip	3.64 Gb
8	zip_basic_tagging/ro-20200129-xml.zip	4.1 Gb
9	zip_basic_tagging/ro-20200129-annotated.zip	2.72 Gb
10	marcell-out/ro-raw.zip	512.28 Mb
11	marcell-out/ro-annotated.zip	2.55 Gb
12	marcell-out/ro-xml.zip	4.02 Gb

Speech Processing

- Corpora creation
- Modules
- Task-based processing

Speech - Corpora creation

Corpora

Corpora list

+ Add

	Name	Lang
1	coref_test	ro
2	termeni nature lematizare	ro
3	enrich4all covid segmentare	ro
4	test_marcell_annoate_andrei	ro
5	test_marcell_annotate_andrei_v2	ro
6	TWITTER_ANONYMIZED_20220725	ro
7	TWITTER_MERGED_20220712	ro
8	TWITTER Corpus 20220602	ro
9	LegalNERo_zenodo_v3	ro
10	BioRo_Extended	ro
11	corpus_punct	ro
12	corpus_wav	ro

Page 1 of 4 20

Add Corpus

Name:

Language:

Description:

Audio:

Gold Annotations:

BRAT Profiles:

Classification Profiles:

Corrected Text:

Add Cancel

Creation Date

2022-11-13
2022-10-09
2022-09-08
2022-08-10
2022-08-10
2022-07-25
2022-07-12
2022-06-02
are c... 2022-03-29
2022-02-14
2022-01-28
2022-01-28

Speech - Corpora creation

Corpus: **ROBIN Batch 4** File: **robin-batch-4.csv** [View Corpora](#)

[Back](#) [Download](#) [View as Text](#)

File View

C1	
1	Care e cel mai ieftin laptop ol viu, cu placă grafică dedicată Tesla ve o sută și șase gigabaiți RAM?
2	Care e cel mai scump laptop ol viu, cu placă grafică dedicată gi fors o mie șizeci ti ai și doi gigabaiți RAM?
3	Care e cel mai scump laptop huawei, cu placă grafică dedicată Tesla ve o sută și opt gigabaiți RAM?
4	Care e cel mai scump laptop Asus, cu placă grafică dedicată Radeon?
5	Care e cel mai scump laptop aisăr, cu placă grafică dedicată Tesla pe o sută și doi gigabaiți RAM?
6	Care e cel mai ieftin laptop acer, cu placă grafică dedicată Tesla ve o sută și șase gigabaiți RAM?
7	Care e cel mai scump laptop del, cu placă grafică dedicată gi fors o mie șizeci ti ai și opt gigabaiți RAM?
8	Care e cel mai scump laptop Xiaomi, cu placă grafică dedicată gi fors o mie optzeci și șase gigabaiți RAM?
9	Care e cel mai scump laptop Asus, cu placă grafică dedicată Tesla ve o sută și doi gigabaiți RAM?
10	Care e cel mai ieftin laptop Xiaomi, cu placă grafică dedicată gi fors o mie optzeci și șase gigabaiți RAM?
11	Mă poti ajuta?
12	Care e cel mai ieftin laptop aisăr, cu placă grafică dedicată Tesla ve o sută și șase gigabaiți RAM?
13	Ce laptop-uri biznis aveți?
14	Care e cel mai scump laptop huawei, cu placă grafică dedicată Tesla ve o sută și doi gigabaiți RAM?

Page 1 of 6 20 1 to 20 of 101

Păiș, Vasile and Ion, Radu and Barbu Mititelu, Verginica and Irimia, Elena and Mitrofan, Maria and Avram, Andrei (2021). **ROBIN Technical Acquisition Speech Corpus**. Zenodo, <https://doi.org/10.5281/zenodo.4626539>

Păiș, Vasile and Ion, Radu and Avram, Andrei-Marius and Irimia, Elena and Mititelu, Verginica Barbu and Mitrofan, Maria (2021). **Human-Machine Interaction Speech Corpus from the ROBIN project**. In Proceedings SPED, pp. 91-96. <https://ieeexplore.ieee.org/document/9587355>

Speech - Corpora creation

Corpus: **ROBIN Batch 4**

View Corpora

Files

Standoff

Tasks

Annotated

Statistics

Archives

Audio

Recorder

Audio	
File	Size
8 audio/andrei_66.wav	1.12 Mb
9 audio/andrei_67.wav	1.16 Mb
10 audio/andrei_68.wav	624.04 Kb
11 audio/andrei_69.wav	1.18 Mb
12 audio/andrei_70.wav	1.12 Mb
13 audio/andrei_71.wav	840.04 Kb
14 audio/andrei_72.wav	1.38 Mb
15 audio/andrei_73.wav	720.04 Kb
16 audio/andrei_74.wav	392.04 Kb
17 audio/andrei_75.wav	680.04 Kb
18 audio/andrei_76.wav	1.13 Mb
19 audio/andrei_77.wav	1.29 Mb
20 audio/andrei_78.wav	1.1 Mb

Page 9 of 26 20 161 to 180 of 505

Speech - Corpora creation

Corpus: **test_audio2**

View Corpora

Back

Recording as: [vasile]

Care e cel mai scump laptop Xiaomi

START

STOP

Sentence 10 / 203

- This app was tested on the Chrome Browser.
- For better results use headphones with mic for recording and speak naturally.
- If you hear your own sound during recording, **TURN OFF YOUR COMPUTER SOUND.**

Speech - Modules

- ASR - DeepSpeech2

Avram, Andrei-Marius and Păiș, Vasile and Tufis, Dan (2020). *Romanian speech recognition experiments from the ROBIN project*. In Proceedings CONSILR, pp. 103--114,

<https://profs.info.uaic.ro/~consilr/2021/wp-content/uploads/2021/03/volum-ConsILR-v-4-final-revizuit.pdf#page=111>

Avram, Andrei-Marius and Păiș, Vasile and Tufis, Dan (2020). *Towards a Romanian end-to-end automatic speech recognition based on DeepSpeech2*. In Proceedings of the Romanian Academy Series A. vol. 21, pp. 395--402,

https://acad.ro/sectii2002/proceedings/doc2020-4/11-Avram_Tufis.pdf

- ASR - WAV2VEC2

Avram, Andrei-Marius and Păiș, Vasile and Tufiș, Dan (2022). *Self-Supervised Pre-Training in Speech Recognition Systems*. Chapter in Speech Recognition Technology and Applications (ed. Păiș, Vasile). Nova Science Publishers, pp. 27--56, <https://doi.org/10.52305/BKWM899>

Speech - Modules

RELATE

- TEPROLIN Service >
- CoRoLa >
- RoWordNet >
- Machine Translation >
- Speech ▾
 - Translate RO - EN
 - Translate EN - RO
 - ROBIN ASR
 - ROBIN ASR Development
 - ROBIN TTS
 - ASR WAV2VEC2
- CURLICAT Anonymization

☰ Romanian Portal of Language Technologies

Welcome  Vasile

WAV2VEC2 ASR (Automatic Speech Recognition) Development

File

Recording

Results

Selectati un fisier WAV asociat unui text in limba romana. Este indicat sa fie inregistrat cat mai clar.

This model is constructed based on fine-tuning a WAV2VEC2 model on a Romanian speech corpus (RTASC). We are working on an improved model making use of multiple resources.

Choose File No file chosen

Run ASR

Scientific papers

Avram, Andrei-Marius and Păis, Vasile and Tufiş, Dan. 2022. [Self-Supervised Pre-Training in Speech Recognition Systems](#). In *Speech Recognition Technology and Applications*, pages 27-56, Nova Science Publishers.

Speech - Modules

- Punctuation Restoration

Păiș, Vasile and Tufiș, Dan (2022). *Capitalization and punctuation restoration: a survey*. Artificial Intelligence Review 55(3), 1681–1722, <https://rdcu.be/cpOch>

Păiș, Vasile (2022). *Punctuation Recovery for Romanian Transcribed Documents*. Chapter in Speech Recognition Technology and Applications (ed. Păiș, Vasile). Nova Science Publishers, pp. 119--154, <https://doi.org/10.52305/BKWM899>

Punctuation Restoration

Selectați modelul dorit și introduceți un document în limba română, fără punctuație și cu litere mici. Modelele bazate pe corpul MARCELL presupun un text de tip legislație. Utilizarea de texte din alte domenii reduce calitatea rezultatelor. Nu sunt realizate curățări ale datelor, astfel încât textul trebuie să conțină doar caractere în limba română.

MARCELL Punctuation Restoration ▾

președintele a vizitat românia suedia și danemarca

Punctuation

Demo text

Speech - Modules

- TTS - RACAI SSLA

Zafiu, Adrian and Dumitrescu, Ștefan Daniel and Boroș, Tiberiu (2015). *Modular Language Processing framework for Lightweight Applications (MLPLA)*. In Proceedings of LTC 2015.

Boroș, Tiberiu (2013). *Contributions to the modeling and implementation of Text-To-Speech Synthesis System*. Case Study: Romanian Language. PhD thesis, Romanian Academy.

Boroș, Tiberiu and Ion, Radu and Dumitrescu, Ștefan Daniel (2013). *The RACAI text-to-speech synthesis system*. In Blizzard Challenge Workshop 2013.

- TTS - ROBIN

Tușiș, Dan and Barbu Mititelu, Verginica and Irimia, Elena and Mitrofan, Maria and Ion, Radu and George, Cioroiu (2019). *Making Pepper Understand and Respond in Romanian*. In the 22nd International Conference on Control Systems and Computer Science.

- bazat pe TTS-Cube : <https://tiberiu44.github.io/TTS-Cube/>

Speech - Modules

- TTS - RACAI SSLA

Zafiu, Adrian and Dumitrescu, Ștefan Daniel and Boroș, Tiberiu (2015). *Modular Language Processing framework for Lightweight Applications (MLPLA)*. In Proceedings of LTC 2015.

Boroș, Tiberiu (2013). *Contributions to the modeling and implementation of Text-To-Speech Synthesis System*. Case Study: Romanian Language. PhD thesis, Romanian Academy.

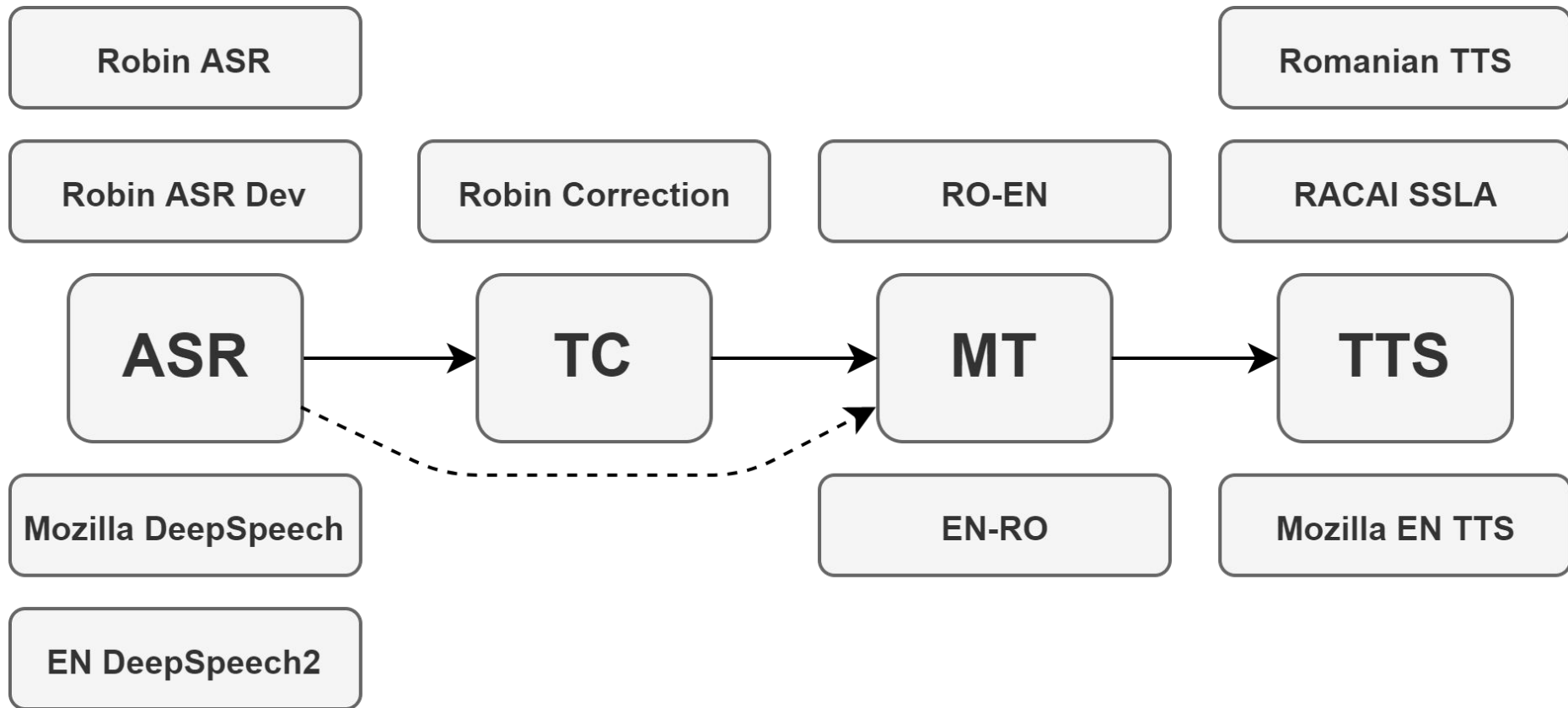
Boroș, Tiberiu and Ion, Radu and Dumitrescu, Ștefan Daniel (2013). *The RACAI text-to-speech synthesis system*. In Blizzard Challenge Workshop 2013.

- TTS - ROBIN - bazat pe TTS-Cube : <https://tiberiu44.github.io/TTS-Cube/>

Tușiș, Dan and Barbu Mititelu, Verginica and Irimia, Elena and Mitrofan, Maria and Ion, Radu and George, Cioroiu (2019). *Making Pepper Understand and Respond in Romanian*. In the 22nd International Conference on Control Systems and Computer Science.

- TTS - RomanianTTS - <http://romaniantts.com>

Adriana Stan, Junichi Yamagishi, Simon King, Matthew Aylett (2011) "The Romanian speech synthesis (RSS) corpus: Building a high quality HMM-based speech synthesis system using a high sampling rate", In Speech Communication, vol. 53, no. 3, pp. 442-450.



A Modular Approach for Romanian-English Speech Translation, Avram, Andrei-Marius and Păiș, Vasile and Tufiș, Dan, NLDB 2021

Speech-to-Speech Translation

File

Recording

Results

Selectați un fișier WAV asociat unui text în limba română. Este indicat să fie înregistrat cât mai clar.

Choose File No file chosen

lanț de prelucrare:

Traducere

For this implementation we used the following:

- Romanian ASR from the ROBIN Project: Andrei-Marius Avram, Vasile Păiș, Dan Tufiș, 2020. Towards a Romanian end-to-end automatic speech recognition based on DeepSpeech2. Proc. Ro. Acad., Series A, Volume 21, No. 4, pp. 395-402.
- Romanian TTS from <http://romaniantts.com> : Adriana Stan, Junichi Yamagishi, Simon King, Matthew Aylett, "The Romanian speech synthesis (RSS) corpus: Building a high quality HMM-based speech synthesis system using a high sampling rate", In Speech Communication, vol. 53, no. 3, pp. 442-450, 2011.
- Romanian SSLA TTS: Tiberiu Boroș, Ștefan D. Dumitrescu, Vasile Păiș, "Tools and resources for Romanian text-to-speech and speech-to-text applications", CoRR, vol. abs/1802.05583, 2018. <https://arxiv.org/pdf/1802.05583.pdf>
- Translation using the EU Council Presidency Translator developed by Tilde with support from RACAI during the Romanian presidency.
- English DeepSpeech2 ASR from: Amodei et al. 2016. Deep Speech 2 : End-to-End Speech Recognition in English and Mandarin. In Proceedings of The 33rd International Conference on Machine Learning (PMLR), 48:173-182.
- English Mozilla DeepSpeech ASR: Hannun et al. 2016. Deep Speech: Scaling up end-to-end speech recognition. arXiv:1412.5567 [cs.CL]
- English Mozilla TTS: <https://github.com/mozilla/TTS>

Text Translation

The screenshot displays the RELATE Romanian Portal of Language Technologies interface. On the left is a navigation menu with the following items: TEPROLIN Service, CoRoLa, RoWordNet, Machine Translation (expanded), Text RO - EN (selected), Text EN - RO, Speech RO - EN, Speech EN - RO, Text Dev Documentation, Speech, and Downloads. The main content area is titled 'Machine translation' and contains the instruction 'Introduceți un text în limba ROMÂNĂ' above a large text input field. A blue 'Submit' button is located below the input field. The page header includes a hamburger menu icon and the text 'Romanian Portal of Language Technologies'.

CEF Automated Translation toolkit for the Rotating Presidency of the Council of the EU

Developed by Tilde with support from ICIA during the Romanian presidency

The API is integrated in RELATE

Speech - Task-based processing

Files Standoff **Tasks** Annotated Statistics Archives Audio

Corpus tasks

+ TEPROLIN + UDPipe + NERRegex + Classify EuroVoc + Terminology + IATE/EuroVoc + Cleanup + NER Baseline + CoNLLUP2BRAT + NER_Old + TTLChunker + BRAT2Co
+ Gold NE list + Statistics + ZIP Text + ZIP Annotated + ZIP Standoff + ZIP Gold Standoff + ZIP Gold Annotated + ZIP Audio + Change Terms Marcell + Export CURLICAT
+ Export Marcell + NER-Legal + Translate Text + ASR + Punctuation + TTS + Translate S2S + DeAnonymization + NER-SiMoNERo + Anonymization + Create Empty Metadata

Type Creation Date

Add task translate S2S

Perform speech to speech translation.

ASR: RO DeepSpeech2

Punctuation: MARCELL

Translate: RO-EN

TTS: EN Mozilla TTS

Description:

Runners (optional):

Overwrite:

Add Cancel

- The correct sequence for C between tasks.
- The correct sequence for M and the results checked be

ould be added one at a time and the results of

Export Marcell. Tasks should be added one at

Linked Data

<https://www.racai.ro/p/lod/>

- Corpora converted to linked data (RDF)
- Apache Jena Fuseki available in RELATE
 - several Romanian corpora loaded in Fuseki
- RELATE supporting RDF in certain places

Barbu Mititelu, Verginica and Irimia, Elena and Păiș, Vasile and Avram, Andrei-Marius and Mitrofan, Maria (2022). *Use case: Romanian language resources in the LOD paradigm*. In Proceedings of the Linked Data in Linguistics Workshop, pp. 35--44, <http://www.lrec-conf.org/proceedings/lrec2022/workshops/LDL/2022.ldl2022-1.0.pdf#page=43>

Barbu Mititelu, Verginica and Irimia, Elena and Păiș, Vasile and Mitrofan, Maria and Avram, Andrei-Marius and Curea, Eric (2021). *Linked open data resources for Romanian*. In Proceedings CONSILR. pp. 7--19, https://profs.info.uaic.ro/~consilr/2022/wp-content/uploads/2022/04/consilr2021_14_03_2022_P.pdf#page=15

Linked Data

Back

[View as RDF](#) [View as JSON](#)

Synset: RO30-09728285-n român

(Noun) Persoană care aparține populației de bază a României sau care este

- hypernym [RO30-09686536-n european](#)
- member_holonym [RO30-08813978-n România](#)

Back

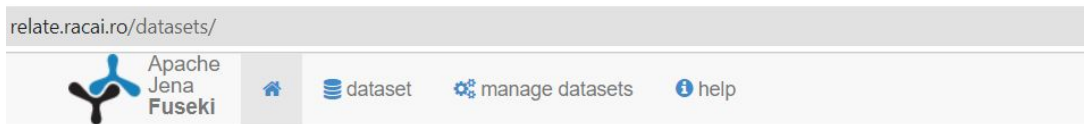
```
@prefix dc: <http://purl.org/dc/terms/> .
@prefix ili: <http://ili.globalwordnet.org/ili/> .
@prefix lime: <http://www.w3.org/ns/lemon/lime#> .
@prefix ontolex: <http://www.w3.org/ns/lemon/ontolex#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfa: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix schema: <http://schema.org/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix synsem: <http://www.w3.org/ns/lemon/synsem#> .
@prefix wn: <http://wordnet-rdf.princeton.edu/ontology#> .
@prefix pwnlemma: <http://wordnet-rdf.princeton.edu/rdf/lemma/> .
@prefix pwnid: <http://wordnet-rdf.princeton.edu/id/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
```

```
<#român-n>
  ontolex:canonicalForm [
    ontolex:writtenRep "român"@ro
  ] ;
  ontolex:sense <#român-09728285-n> ;
  wn:partOfSpeech wn:noun ;
  a ontolex:LexicalEntry .

<#român-09728285-n>
  ontolex:isLexicalizedSenseOf pwnid:09728285-n ;
  a ontolex:LexicalSense .

pwnid:09728285-n
  wn:partOfSpeech wn:noun ;
  wn:definition [ rdf:value "Persoană care aparține populației de bază a României" ] ;
  wn:09686536-n pwnid:hypernym ;
  wn:08813978-n pwnid:member_holonym ;
  a ontolex:LexicalConcept .
```

Linked Data



<https://relate.racai.ro/datasets/>

Allows interrogation via
SPARQL queries

Datasets on this server

dataset name	actions
/legalnero	query add data info
/parseme	query add data info
/rolex	query add data info
/rown	query add data info
/rtasc	query add data info
/simonero	query add data info

Păiș, Vasile and Barbu-Mititelu, Verginica (2022). *Linguistic Linked Open Data for Speech Processing*. Chapter in Speech Recognition Technology and Applications (ed. Păiș, Vasile). Nova Science Publishers, pp. 155--188, <https://doi.org/10.52305/BKWM899>

RELATE development

- open source

<https://github.com/racai-ai/RELATE>

- component-based
 - each functionality developed as a separate component
 - usually no direct interactions between components

