A presentation of the Research Institute for Artificial Intelligence ”Mihai Drăgănescu”

Welcome to HE ambassador of Republic of India, Mr. Rahul Shrivastava

The plan of the meeting

- Part I

- 14:10-14:30 Acad. Dan Tufis: Presentation of the Institute:
  - Domains of interest and activities,
  - Main achievements
  - International positioning of the institute
- 14:30-14:40 Dr. Verginica Barbu Mititelu: Innovation in language resources development
  - Major resources developed
  - Standardization (open linked data format)
  - Access to the resources
- 14:40-14:50 Acad. Dan Tufis: Open portal for technologies for the Romanian language
  - Main characteristics of the portal, services and access
  - Availability

- PART II

- 9:50 - Question answering, Open discussions

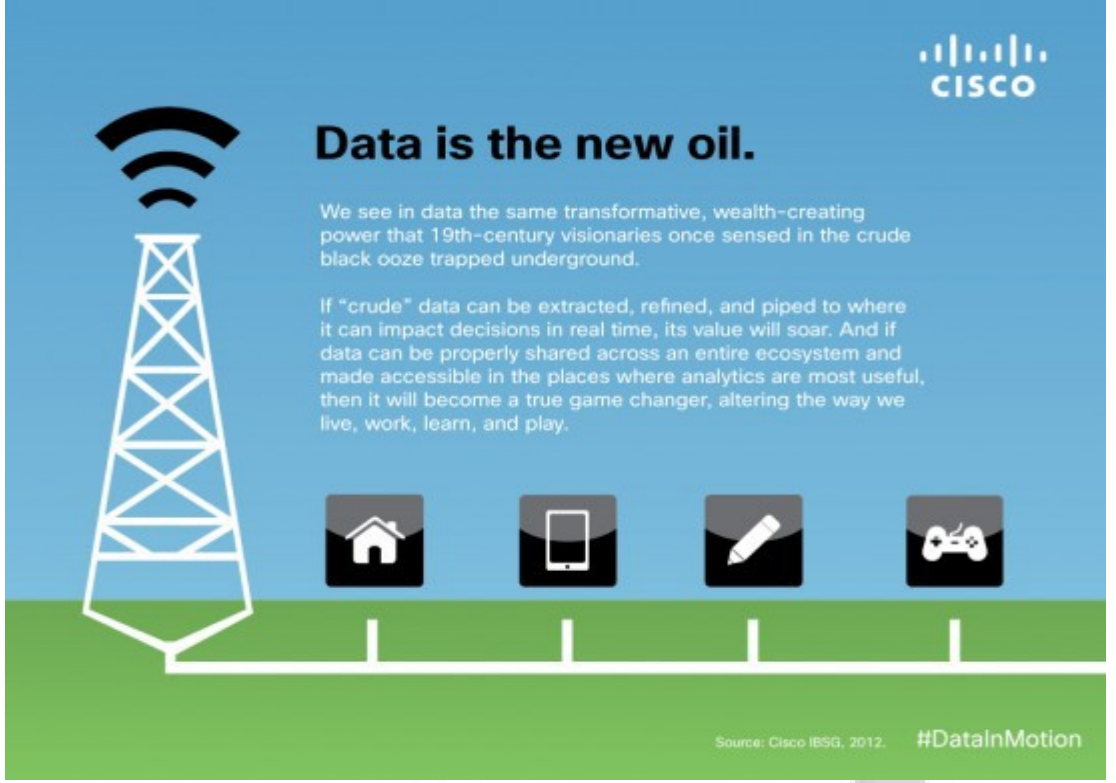
Who are we (ICIA)?

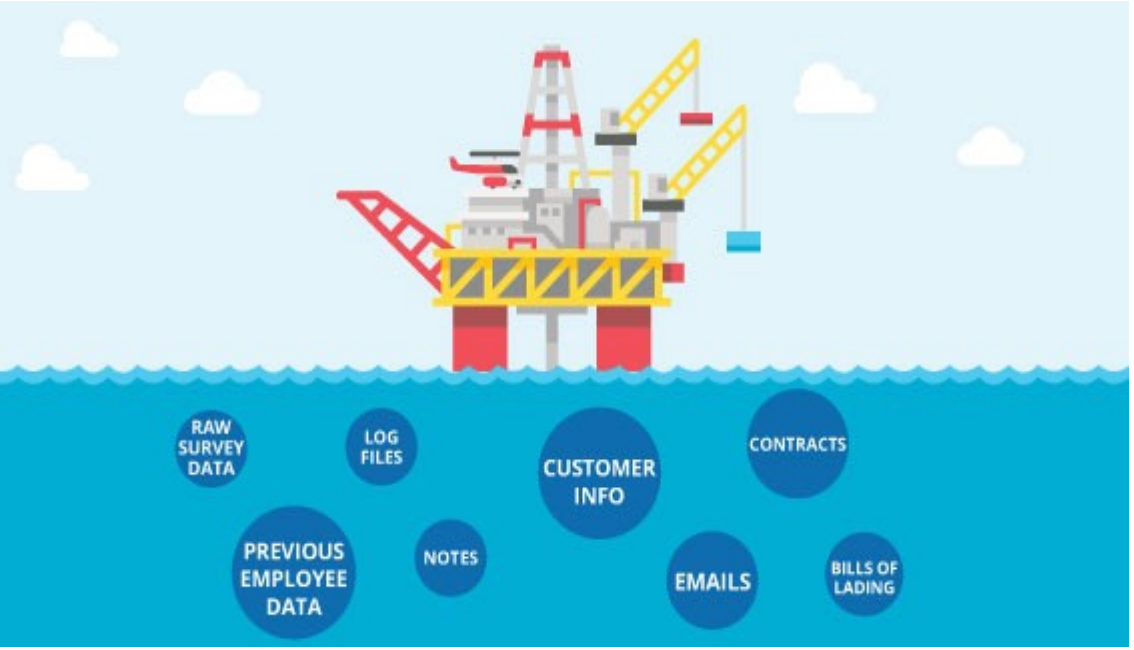
- A small research institute, established in 2002 on the premises of a smaller Center for Advanced Research established in 1994
- Permanent personnel: research positions 18, administrative: 6
- A larger number of internships (more than 25 every year)
- PhD programs, small number of PhD students, only 3 PhD advisors
- Organizer/co-organizer of various international scientific &educational events (15 editions of the EUROLAN intl. Training school, 15 editions of CONSILR conference, 11 editions of SPeD, 3 ELRC workshops, etc)
- Solid publications record (more than 1070 publications)
- Large scientific parnerships (more than 50 intl projects and 34 natl. Projects)
- ISP for most of the institutes in the House of the Academy
- A decent research infrastructure(10 servers both CPU and GPU equiped, individual desktops and laptops/tablets, smart router, several switches, etc)

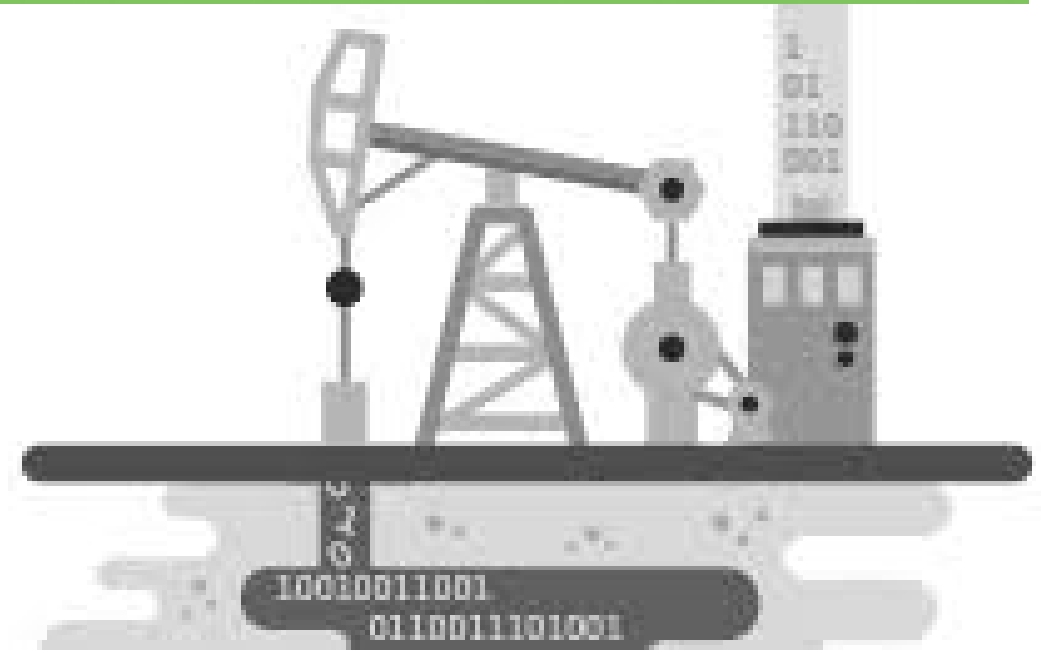
A few remarkable results of ICIA researchers

- The construction and maintenance of one of the world largest lexical resources: RoWordNet (more than 20 years)
- Preparing the Romanian part of the JRC-Aquis multilingual corpus (the most used multilingual resource in training MT systems)
- Aligning large comparable corpora and extracting parallel text segments (ACCURAT project)
- Implementing the first professional MT systems for Romanian in various technological epochs (dictionary-word based, grammar rule-based, statistics-based, deep NN-based)
- Winning a series of shared-tasks (intl. competions):
  - word alignment (ACL, Edmonton, 2003), word-sense disambiguation (ACL SenseEVAL III, Barcelona, 2004), word alignment (Ann Arbor (ACL, 2005), CLEF Q&A (2007, 2008, 2009, 2010), several other shared tasks in which our team was among the protagonists NER New Mexico (ACL, 2021), Punta Cana (ACL, 2021) etc.

Data is the new **oil** of the **21st century**





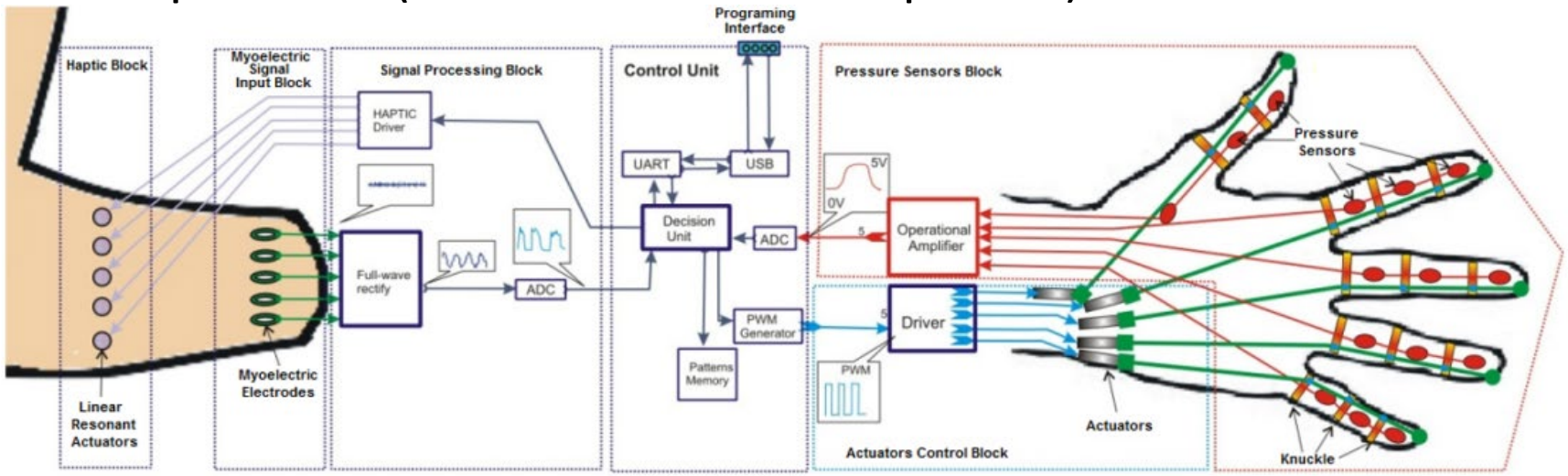


Research and Development Areas

- Large language Resources and development of SOTA Technologies
  - Language resources in a standardized reusable form
  - Instruments for processing large quantities of language data
  - Neural networks language models (BERT-like)
  - Applications mediated by language (Romanian) (QA, MT, chatbots, tutoring systems, etc.)
- Language Technology infrastructures
  - Trans-European Language Resources Infrastructure, European Network of Excellence in Human Language Technologies, Fostering Language Resources Network, Enhancing the European Linguistic Infrastructure,): European network for Web-centred linguistic data science (LLOD), European Language Grid, European Language Equality , European Language Research Coordination, Portal of Romanian Language Technology (tools and data)

Research and Development Areas

- National Center for the Study of the Brain (newly established), coordonated by 2 members of the Romanian Academy and a distinguished Professor of neuro sciences
- New Electronic Architectures and applications
  - Myolectric prostesis (artificial arms for amputees)

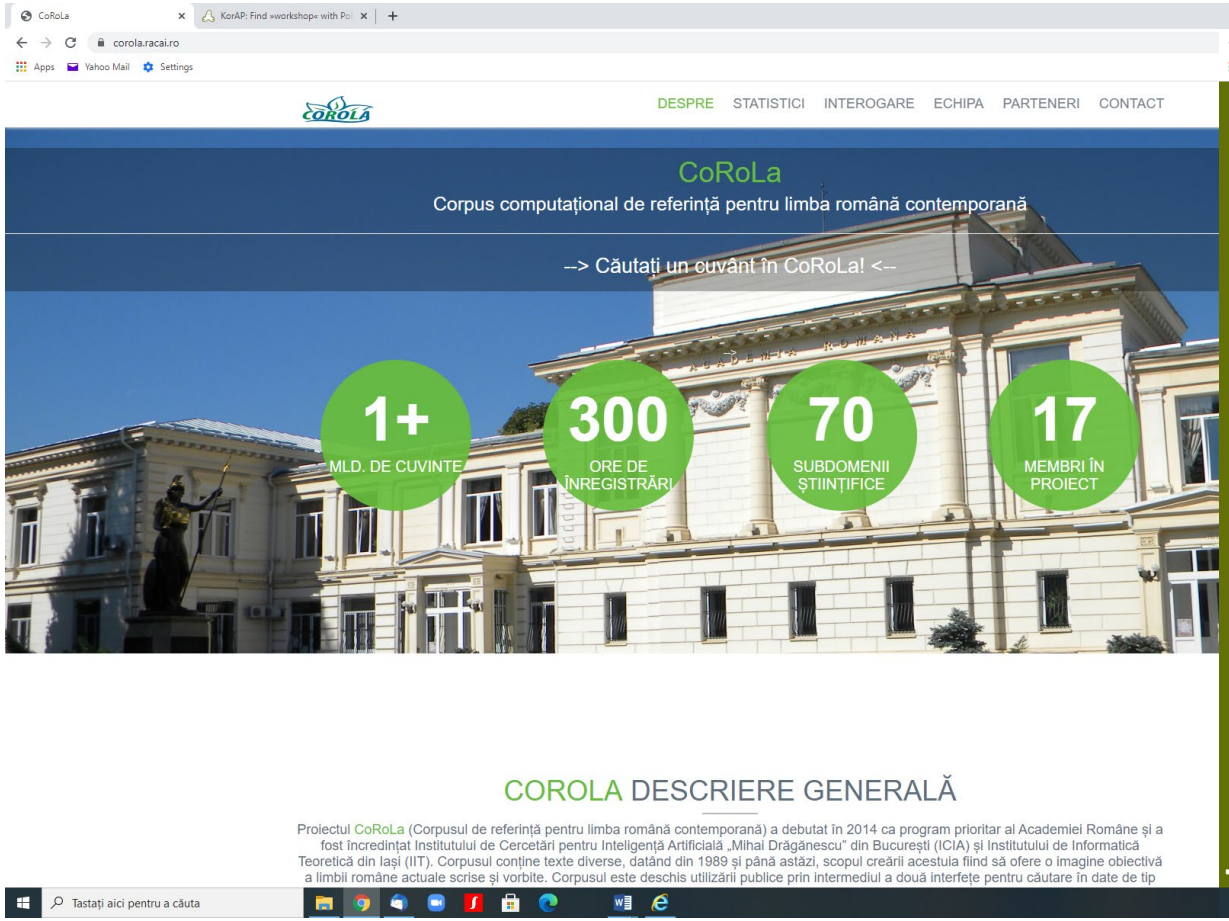


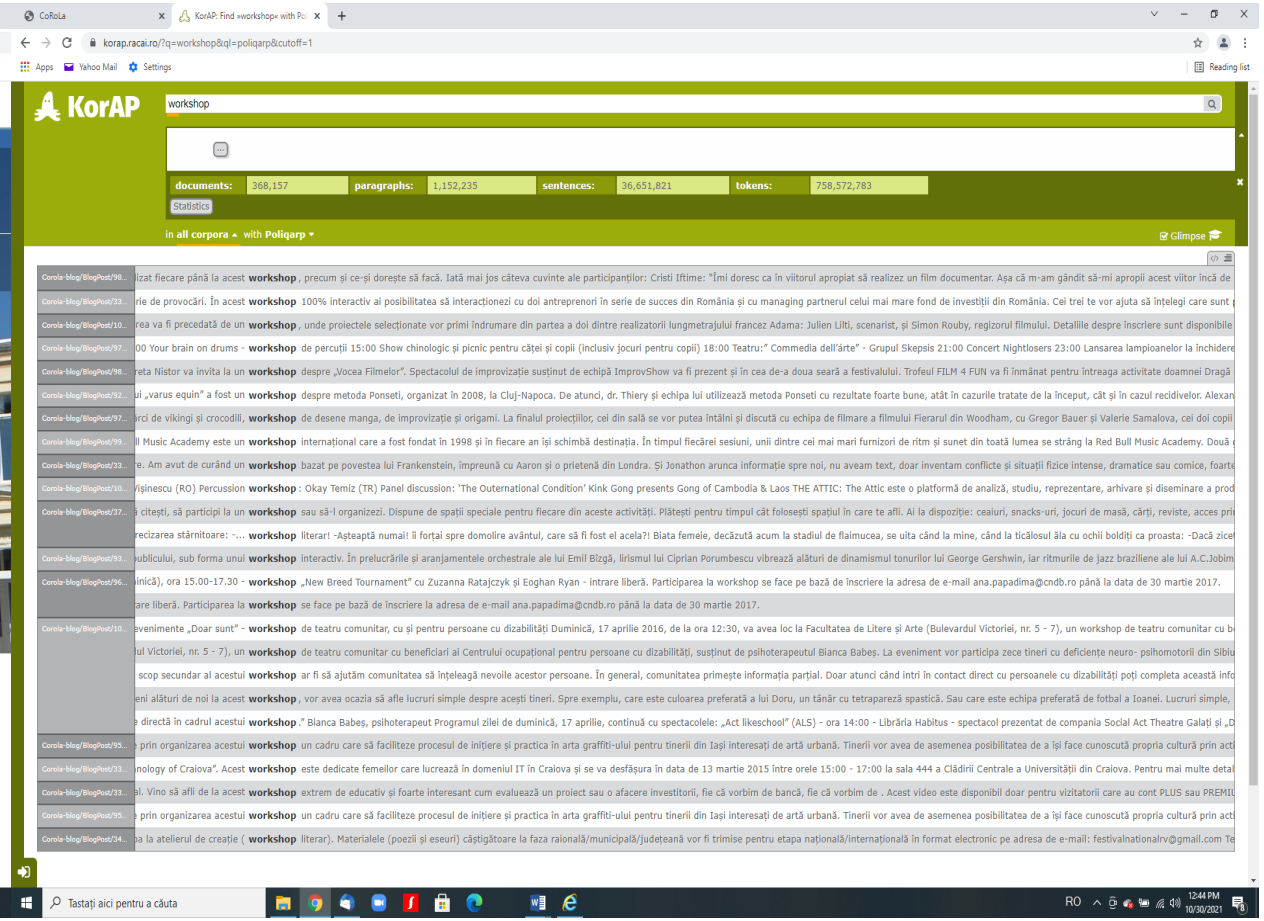
  - Study of the babies cry (Dunstan language – universal?)
- Study and combatting the Fake News (a recent Horizon-IA project proposal led by the governmental *Department of Relations with Moldovan Republic*)

Major results shared with everybody (I)

The Reference Corpus of Contemporary Romanian Language

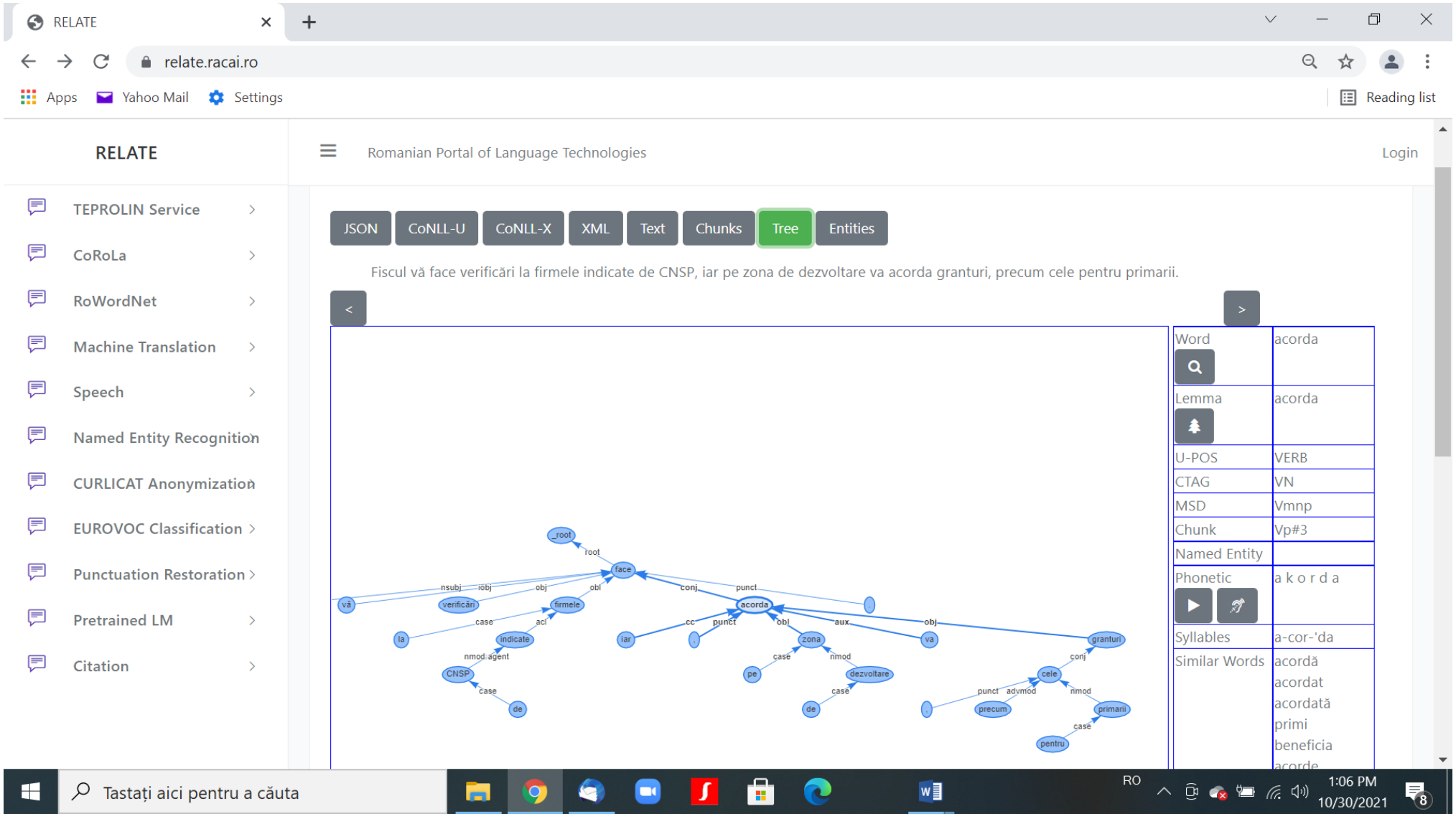
ICIA, UB, IIT, UAIC, IDS-Mannheim





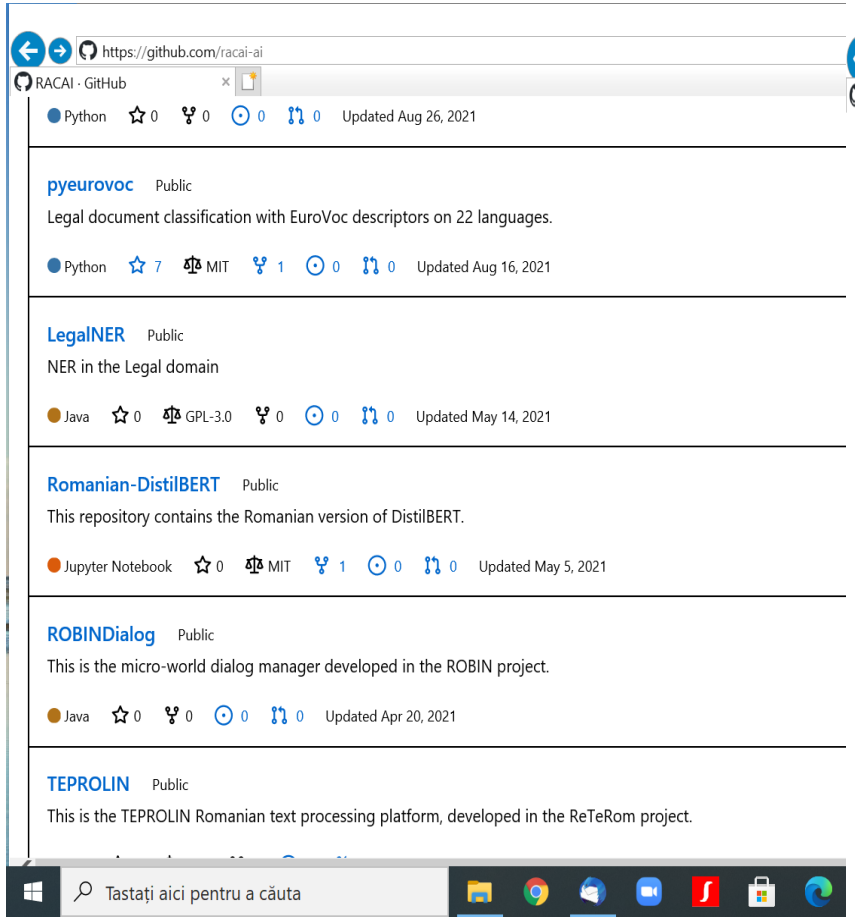
Major results shared with everybody (II)

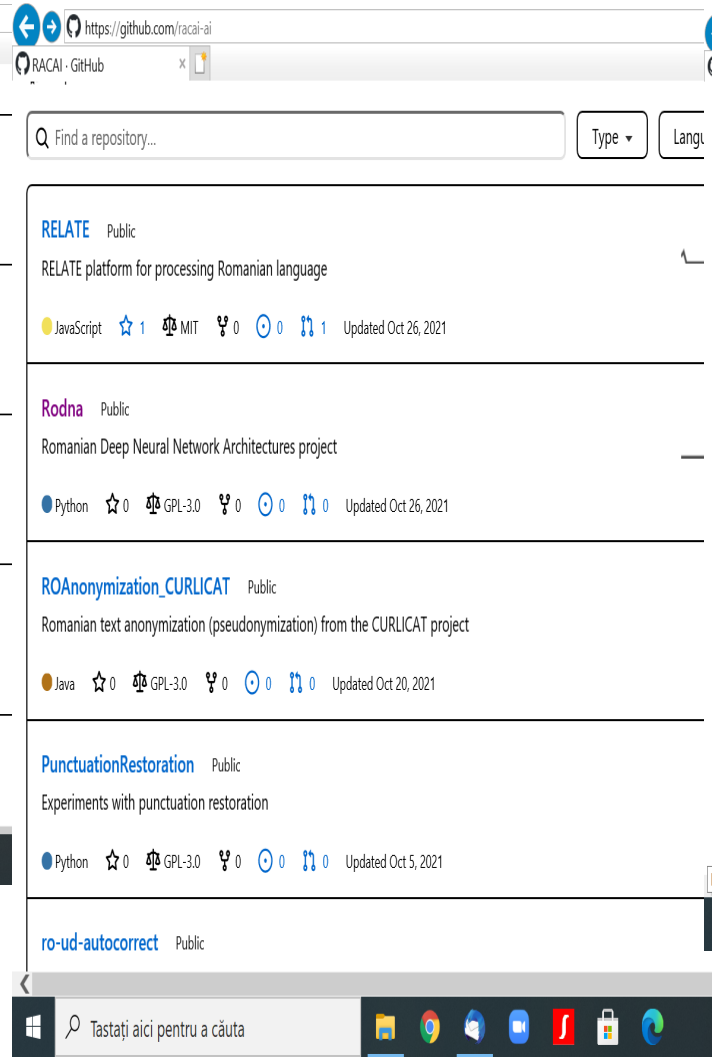
Portalul RELATE

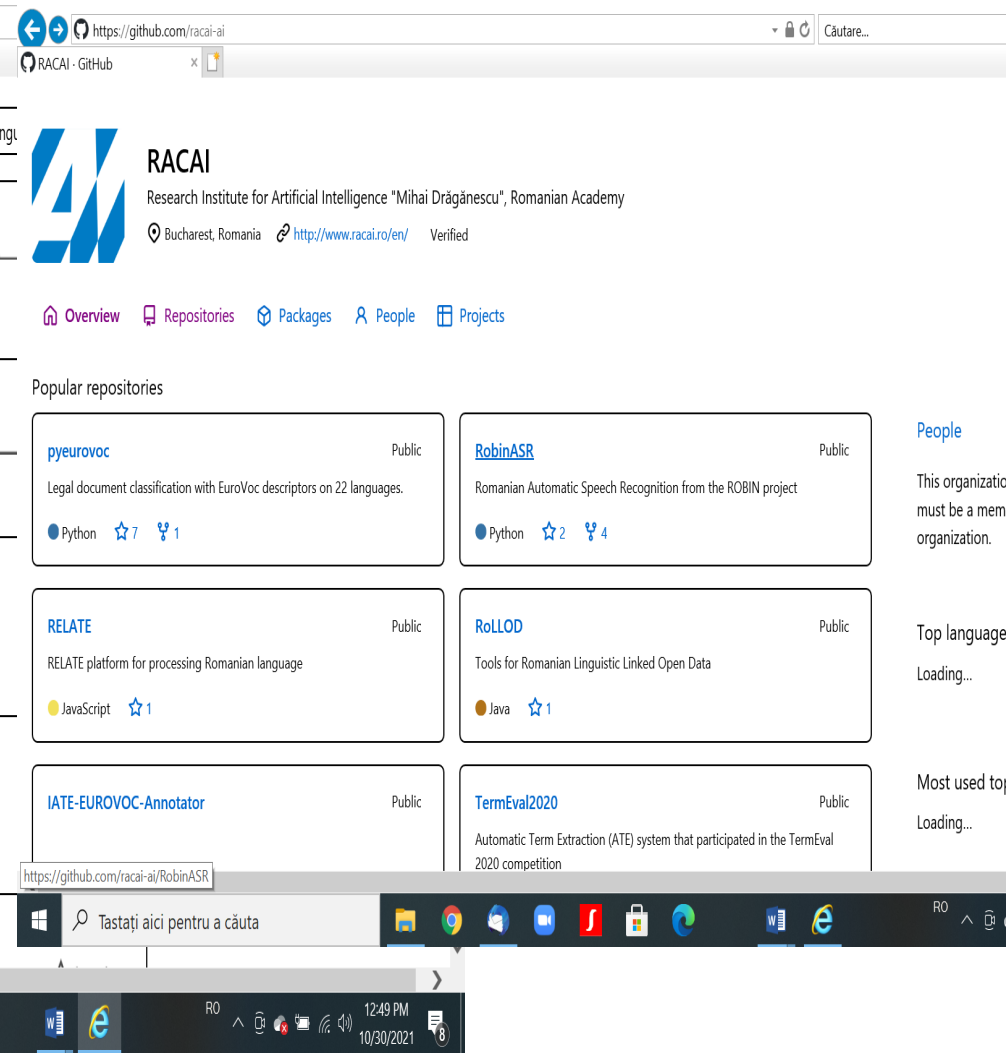


Major results shared with everybody (III)

ICIA github







Thank you!