TULAP

An Accessible and Sustainable Platform for Turkish NLP Resources

Susan Üsküdarlı, Muhammet Şen, Furkan Akkurt, Merve Gürbüz
Onur Güngör, Arzucan Özgür, and Tunga Güngör

Department of Computer Engineering, Boğaziçi University, Istanbul, Turkey
tulap.cmpe.boun.edu.tr – tulap@boun.edu.tr

EACL 2023

Objectives

- To develop an online research output platform (ROP) to archive, demonstrate, and provide usable NLP resources created by a research group (deployed for TABILab)
- Offer the platform itself as an open-source resource

Motivation

- Keeping track of research outputs created by groups is challenging due to:
  - Lack of conventions and documentation
  - Potentially multiple versions
- Well-documented best versions of resources are essential for:
  - transfer of knowledge (within a lab as well as the scientific community)
  - continuity of research
- True accessibility of research outcomes calls for demonstrations as well as usability

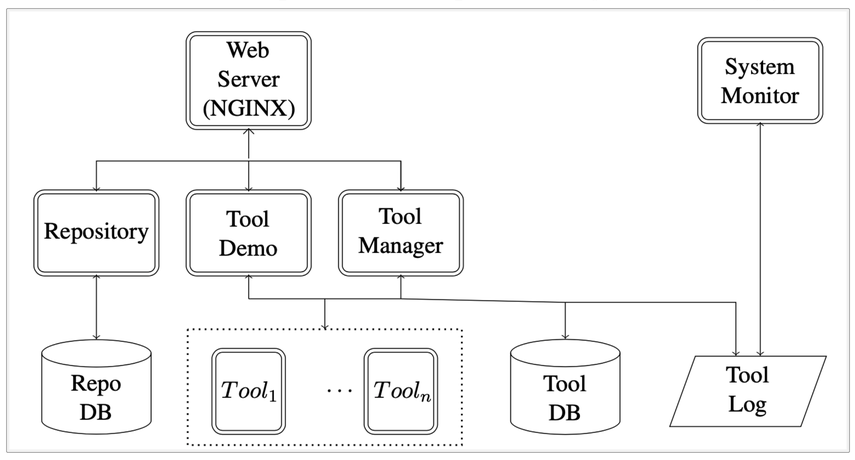
Main Contributions

- **R**esearch **O**utput **P**latform
  - Archival process to support continuous contributions
  - Systematic documentation of resources (datasets and tools)
  - APIs and containerized tool versions for accessibility
  - Monitoring system to identify problems and track resource usage
- NLP deployment (TULAP) [a]
  - Provides Turkish NLP resources developed at Boğaziçi University[b] including **13 datasets** and **17 tools** (April 2023)
  - Actively being used by our research group as a reference as well as for the addition of new datasets, tools and demos

[a] Turkish Language Processing Platform
[b] https://tulap.cmpe.boun.edu.tr/

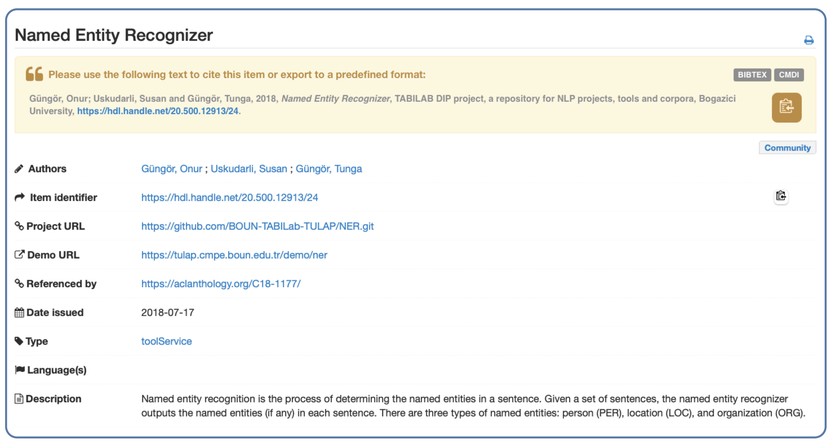
Design

Designed as containerized components for portability, scalability, and usability.

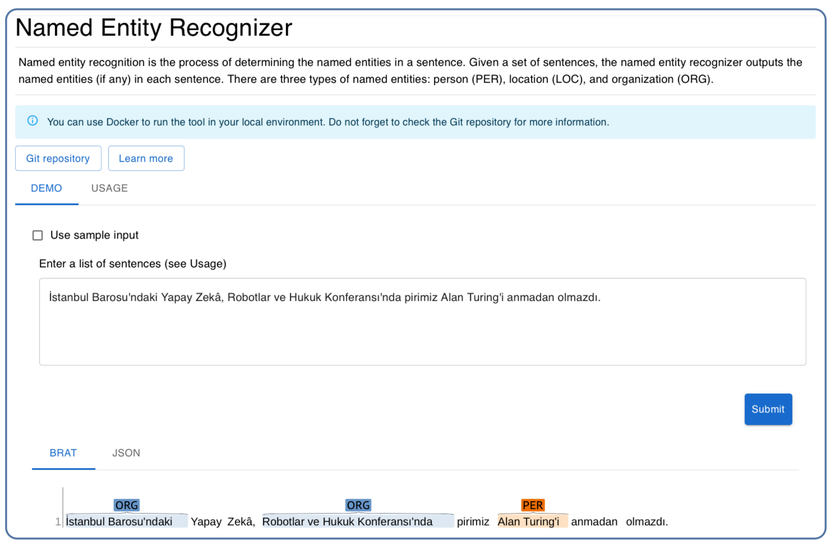


Use Case for Named Entity Recognizer (NER)

1. Discovery – Find and view info about some NER research



2. Demo – NER tool (seen above)



3. Executing tool – local deployment and API call (same as above)

```shell
$ git clone https://github.com/BOUN-TABILab-TULAP/NER.git
$ docker build -t ner .
$ docker run -d -p 8080:8080 ner
$ curl -X POST http://localhost:8080/ner/predict/ -H 'Content-Type:
application/json' -d '{"text":"İstanbul Barosu\'ndaki Yapay Zekâ,
Robotlar ve Hukuk Konferansı\'nda pirimiz Alan Turing\'i anmadan
olmazdı."}'

{'tagger_output': {'0': ['İstanbul', 'B-ORG'], '1': ["Barosu'ndaki",
'I-ORG'], '2': ['Yapay', 'O'], '3': ['Zekâ', 'O'], ...}
```

4. New tool demos can be added by specifying general information, demo input-output specifications, and a user guide.
5. Monitoring demos: *System Monitor* to analyze usage metrics

Requirements of platform

- Resources can easily be added
- Resources can be searched, browsed, and utilized
- An interactive demo to showcase tool functionality
- Open-source tools are easily deployable to support accessibility
- Robust building, monitoring, and restarting of platform via container-based architecture
- Seamless addition of new tools due to containers and systematic access via APIs
- A monitoring tool to track the state and use of the system

Implementation

- All tools provide an API
- All tools and system components are containerized using Docker
- DSpace for systematic and standardized documentation of resources

Future Work

- Improving present tools
- Addition of new tools
- Improved user and tool interaction

Acknowledgements

This work was supported by Boğaziçi University Research Fund Grant Number 16909. We are grateful to all resource providers and the TABILab alumni who provided valuable feedback.

Resources

[1] github.com/ufal/clarin-dspace
[2] wiki.lyrasis.org/display/DSPACE
[3] https://github.com/BOUN-TABILab-TULAP/tabi-rop
[4] tabilab.cmpe.boun.edu.tr
[5] github.com/BOUN-TABILab-TULAP