

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

Objec	tives	Design	Requirements of platform
• To develop an on output platform ((ROP) to	Designed as containerized components for portability, scalability, and usability.	Resources can easily be addedResources can be searched, browsed,

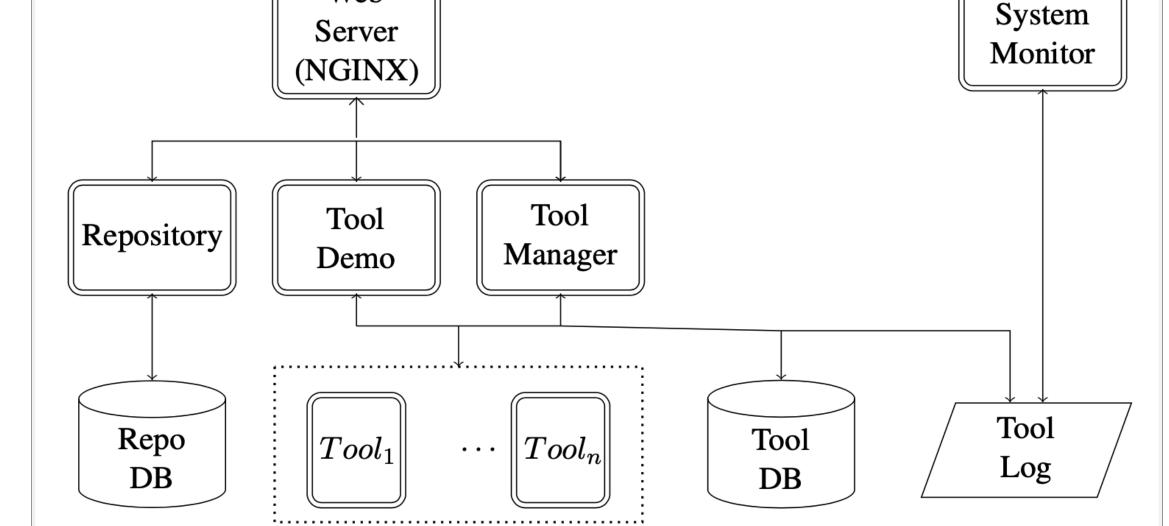


archive, demonstrate, and provide usable NLP resources created by a research group (deployed for TABILab)

• 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 documentationPotentially 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



Use Case for Named Entity Recognizer (NER)

1 Discovery – Find and view info about some NER research

Named Entity	Recognizer	Ð
	following text to cite this item or export to a predefined format: darli, Susan and Güngör, Tunga, 2018, Named Entity Recognizer, TABILAB DIP project, a repository for NLP projects, tools and corpora, Bogazici	
	I.handle.net/20.500.12913/24.	
Authors	Güngör, Onur ; Uskudarli, Susan ; Güngör, Tunga	Community
Item identifier	https://hdl.handle.net/20.500.12913/24	Ē
% Project URL	https://github.com/BOUN-TABILab-TULAP/NER.git	
♂ Demo URL	https://tulap.cmpe.boun.edu.tr/demo/ner	

- 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 APIAll tools and system components are

outcomes calls for demonstrations as well as usability

Main Contributions

- Research Output Platform
- 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)

⊗ Referenced by	https://aclanthology.org/C18-1177/		
mate issued	2018-07-17		
🗣 Туре	toolService		
🗖 Language(s)			
Description	Named entity recognition is the process of determining the named entities in a sentence. Given a set of sentences, the named entity recogniz outputs the named entities (if any) in each sentence. There are three types of named entities: person (PER), location (LOC), and organization (

2 Demo – NER tool (seen above)

Named Entity Recognizer

Named entity recognition is the process of determining the named entities in a sentence. Given a set of sentences, the named entity recognizer outputs the named entities (if any) in each sentence. There are three types of named entities: person (PER), location (LOC), and organization (ORG).

O You can use Docker to run the tool in your local environment. Do not forget to check the Git repository for more information.						
Git repository Learn more						
DEMO USAGE						
Use sample input						
Enter a list of sentences (see Usage)						
İstanbul Barosu'ndaki Yapay Zekâ, Robotlar ve Hukuk Konferansı'nda pirimiz Alan Turing'i anmadan olmazdı.						
	Submit					
BRAT JSON						
ORG ORG PER						

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

Istanbul Barosu'ndaki Yapay Zeka, Robotlar ve Hukuk Konferansi'nda pirimiz Alan Luring'i

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.

 Actively being used by our research group as a reference as well as for the addition of new datasets, tools and demos

^aTurkish Language Processing Platform ^bhttps://tulap.cmpe.boun.edu.tr/ 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', '0'], '3': ['Zekâ', '0'], ...}

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

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