UniDive WG3

Subgroup - Multilingual Tool and Resource Documentation

Co-leaders:
A. Seza Doğruöz (Ghent University)
Maria Giagkou (Athena RC)
Teresa Lynn (Mohamed bin Zayed University of Artificial Intelligence)

UniDive

Task Overview

Task 3.1.1

- Assess the "discoverability" of NLP tools and resources
- Who can participate?
  - Everyone

Task 3.1.2

- Analyse the NLP tool availability in the ELG catalogue
- Who can participate?
  - Excel or Tableau enthusiasts
  - Those with skills in data visualisation

UniDive

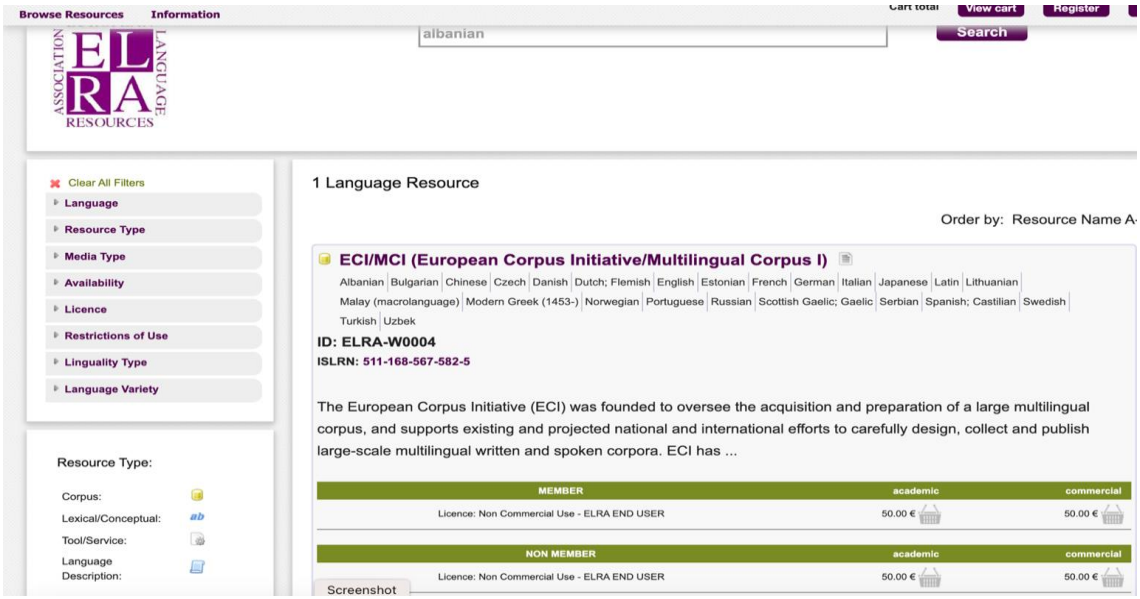
Task 3.1.1: Assessing the “discoverability” of NLP tools

- Choose your language(s) and NLP task(s) of interest
- Search for the relevant tools across a number of platforms
- Report on the discoverability of desired tool (Could you find easily it or not? What challenges?)
- Report on the metadata information available (was it sufficient and accurate?)
- What metadata do you recommend should be provided for a similar search?
- Is there a tool/ resource you are aware of that you can’t find on these platforms?

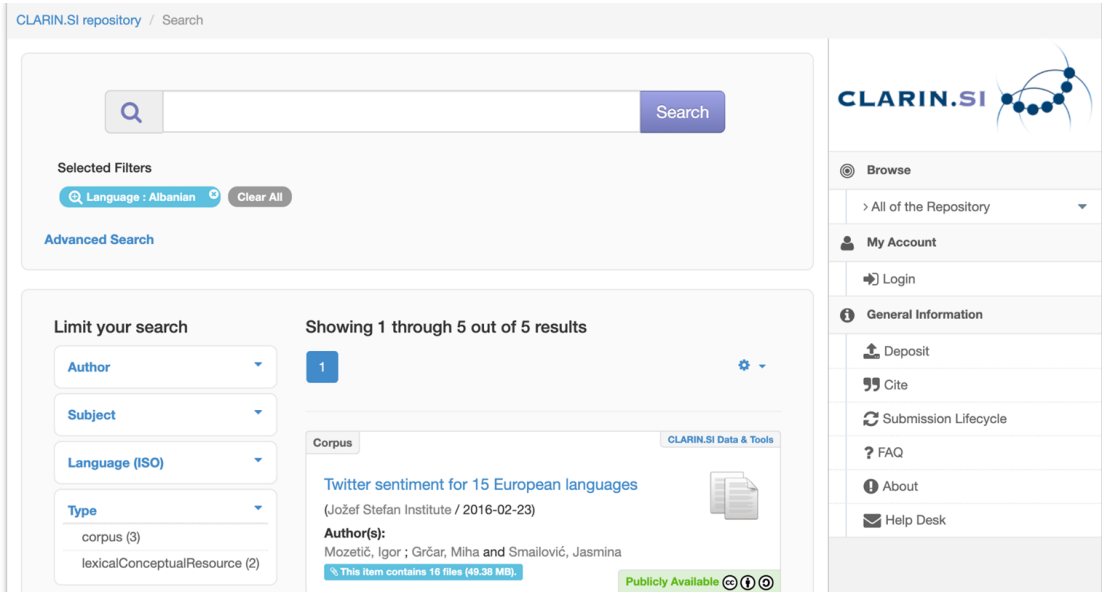
**(Activity will kick-off in Naples)

UniDive

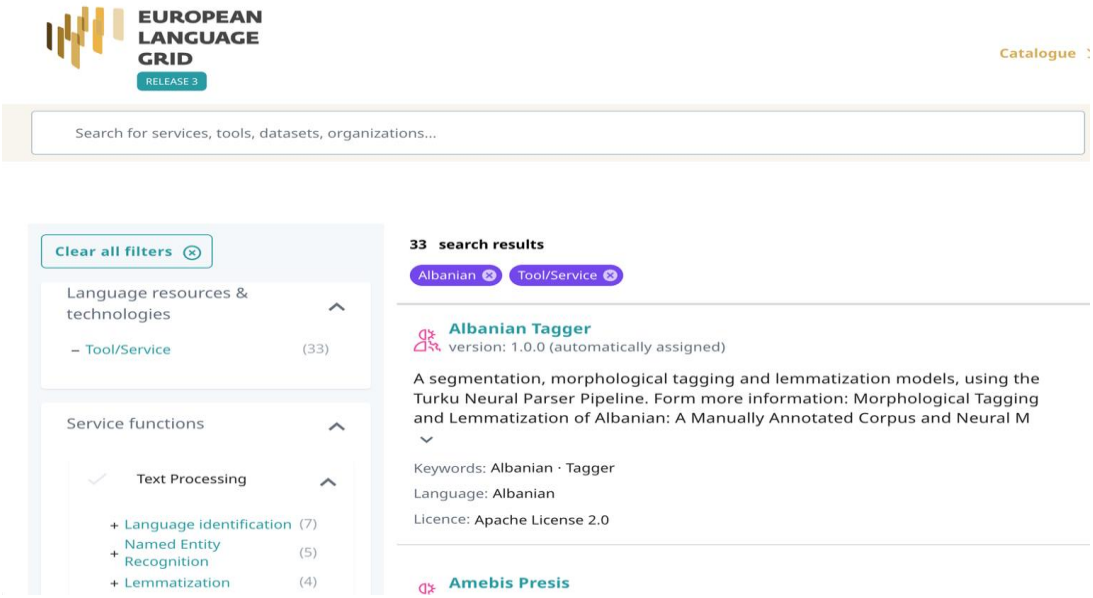
E.g Search for Albanian Tools - ELRA Catalogue



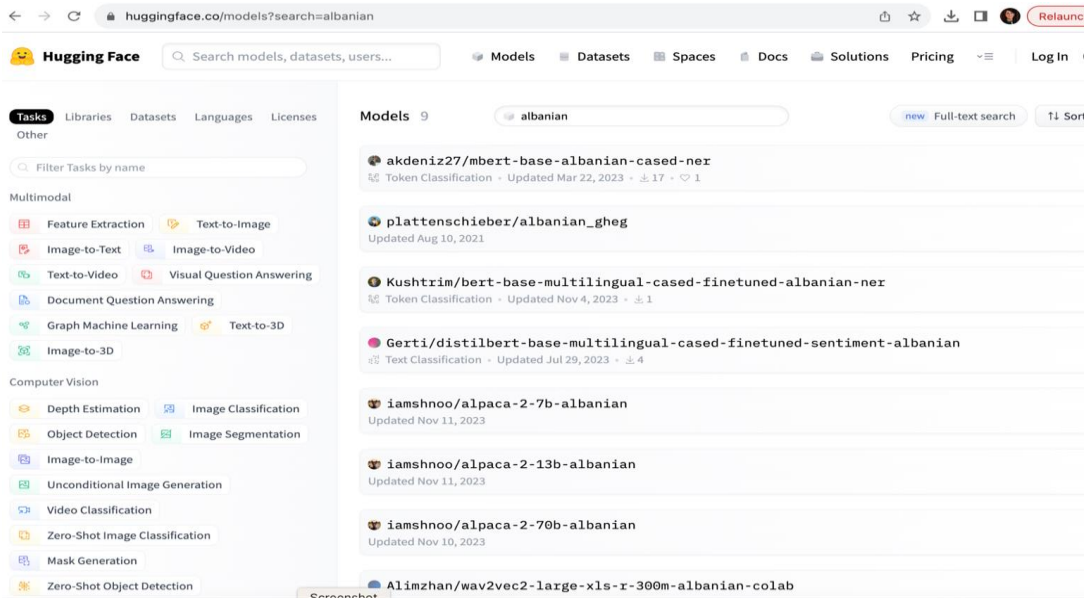
E.g Search for Albanian Tools - CLARIN-SI Catalogue



E.g Search for Albanian Tools - ELG Catalogue



E.g Search for Albanian Tools - Hugging Face



Task 3.1.1: Assessing the “discoverability” of NLP tools

Process

- A template will be provided with prompt questions
- Additional input also desired

Outcome

- An increased awareness and understanding of language technology platforms
- Insight into limitations of current schemas
- Honed research skills in searching for NLP tools/ resources
- Recommendations for improving discoverability of tools/ resources

UniDive

Task 3.1.2: Tool Availability Analysis

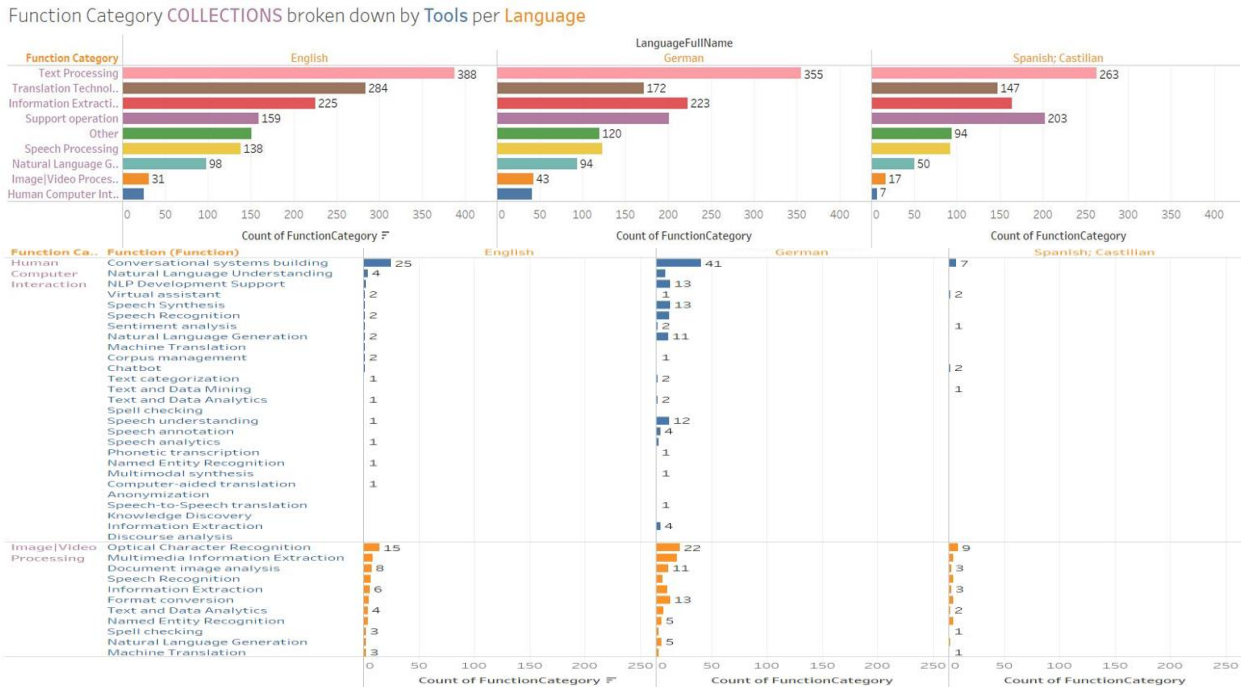
- Seeking volunteers with strong Excel/ Tableau skills
- Analysis required on ELG catalogue export - similar to Kristina's report as starting point
- Prompts below can be the start of investigation - let's see what else emerges:
  1. The tools that certain languages are missing (e.g. Irish doesn't have NER, Sentiment Analyser, etc)
  2. The multilingual tool types that are lacking across languages (e.g. NER is only available for X, Y, Z languages)
  3. Which languages tend to be left out of "multilingual" tools?

UniDive

ELG Catalogue Export

O1 fx Licences

	B	G	J	K	O	R
1	Resource Name	Function	Input Media Types	Input Languages	Licences	Landing Page
379	CORDEX inflectional lookup data 1.0	undefined		sl	Creative Commons Attribution Non	http://hdl.handle.net/113
380	ANMOP	Text categorization|Text and Data	text	es		http://www.redilegra.com
381	Latvian grammar checker	Grammar checking	text	lv		https://www.tilde.lv/parei
382	COREA-coreferentieservice	Co-reference resolution		nl		http://hdl.handle.net/100
383	Collective Text to Speech	Text-to-Speech Synthesis|Speech	text	es|pt	GNU General Public License v2.0	https://pypi.org/project/c
384	Lengoo Termbase	Terminology	text	de		https://www.lengoo.com/
385	extraTerm	Term extraction	text	en|de		https://www.iailc.de/en/s
386	Korp, Kielipankki version	Concordance search	text	ru|es|fr|de|en|sv|fi|mdf|myv|sjd|swh		https://korp.csc.fi
387	Raudikko Analysis for Elasticsearch	Text and Data	text	fi	GNU Lesser General Public License	https://github.com/Evider
388	MIOPIA	Annotation|Sentiment analysis	text	es|en		https://miopia.grupolys.o
389	Across Translator Edition	Terminology	text	de		https://www.across.net/e
390	Norma	Summarization|NLP Development				http://simple4all.org/proc
391	StrokeAid	Speech Synthesis	text	hu		http://magyarbeszed.tmit
392	Recognizer	Speech understanding	audio	lt	Creative Commons Attribution 4.0	https://xn--ratija-
393	SpeCT - Speech Corpus Toolkit for Praat (v1.0.0)	Speech annotation|Text and Data			GNU Lesser General Public License	https://zenodo.org/record
394	voiceovermaker.io	Speech Synthesis	text	no|ko|ja|it|id|hu|hi|el|de|fr|fi|fil|en|		https://voiceovermaker.io
395	IRIS English-Irish Translation System	Machine Translation	text	ga|en		http://server1.nlp.insight-
396	Г„Гtni Company's Automatic Speech Recognition	Speech understanding|Multimedia	audio|t	fi|en|fi|en		https://www.aanicompany
397	iTranslate Offline Translation	Machine Translation	text	vi|tr|th|fa|ru|ko|ja|id|he|zh|bs|sq|ar		https://itranslate.com/lan

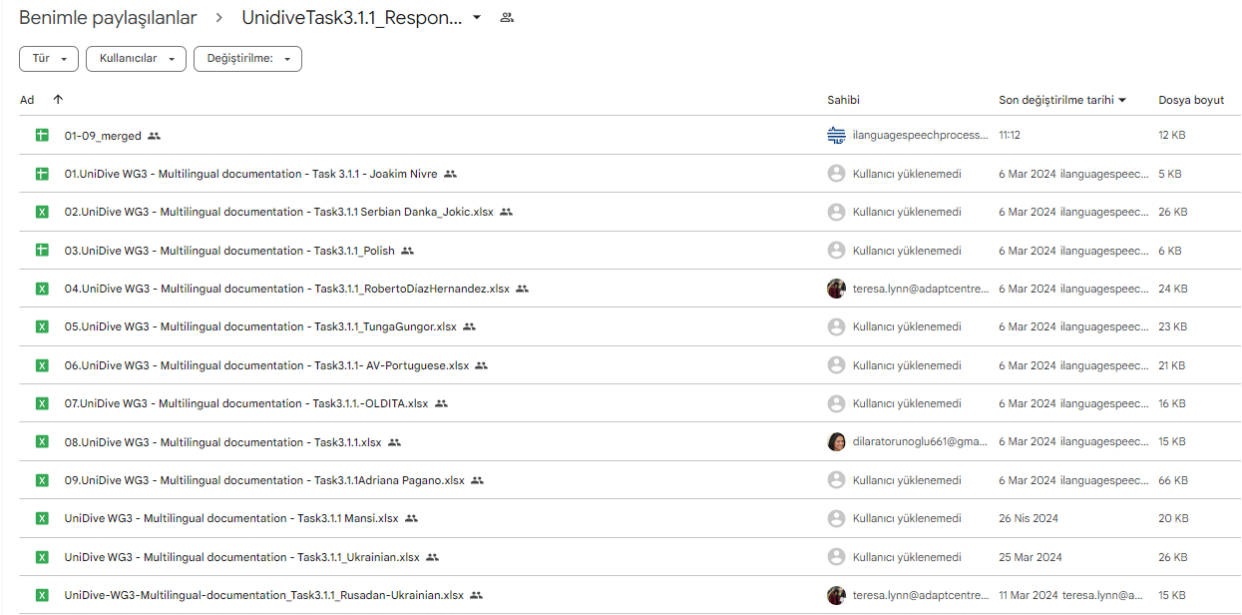


Task 3.1.2: Tool Availability Analysis

Expected Outcomes

- A better insight into current NLP tool availability
- A better insight into existing gaps and digital language inequality
- A basis for improved reporting on language support or tool availability status (visually/ written reports)

Current Participants



Benimle paylaşılanlar › UnidiveTask3.1.1_Respon...

Tür | Kullanıcılar | Değiştirilme:

Ad	Sahibi	Son değiştirilme tarihi	Dosya boyut
01-09_merged	ilanguagespeechprocess...	11:12	12 KB
01.UniDive WG3 - Multilingual documentation - Task 3.1.1 - Joakim Nivre	Kullanıcı yüklenemedi	6 Mar 2024 ilanguagespeec...	5 KB
02.UniDive WG3 - Multilingual documentation - Task3.1.1 Serbian Danka_Jokic.xlsx	Kullanıcı yüklenemedi	6 Mar 2024 ilanguagespeec...	26 KB
03.UniDive WG3 - Multilingual documentation - Task3.1.1_Polish	Kullanıcı yüklenemedi	6 Mar 2024 ilanguagespeec...	6 KB
04.UniDive WG3 - Multilingual documentation - Task3.1.1_RobertoDiazHernandez.xlsx	teresa.lynn@adaptcentre...	6 Mar 2024 ilanguagespeec...	24 KB
05.UniDive WG3 - Multilingual documentation - Task3.1.1_TungaGungor.xlsx	Kullanıcı yüklenemedi	6 Mar 2024 ilanguagespeec...	23 KB
06.UniDive WG3 - Multilingual documentation - Task3.1.1- AV-Portuguese.xlsx	Kullanıcı yüklenemedi	6 Mar 2024 ilanguagespeec...	21 KB
07.UniDive WG3 - Multilingual documentation - Task3.1.1.-OLDITA.xlsx	Kullanıcı yüklenemedi	6 Mar 2024 ilanguagespeec...	16 KB
08.UniDive WG3 - Multilingual documentation - Task3.1.1.xlsx	dilaratorunoglu661@gma...	6 Mar 2024 ilanguagespeec...	15 KB
09.UniDive WG3 - Multilingual documentation - Task3.1.1Adriana Pagano.xlsx	Kullanıcı yüklenemedi	6 Mar 2024 ilanguagespeec...	66 KB
UniDive WG3 - Multilingual documentation - Task3.1.1 Mansi.xlsx	Kullanıcı yüklenemedi	26 Nis 2024	20 KB
UniDive WG3 - Multilingual documentation - Task3.1.1_Ukrainian.xlsx	Kullanıcı yüklenemedi	25 Mar 2024	26 KB
UniDive-WG3-Multilingual-documentation_Task3.1.1_Rusadan-Ukrainian.xlsx	teresa.lynn@adaptcentre...	11 Mar 2024 teresa.lynn@a...	15 KB

Link to the Excel form

- https://docs.google.com/spreadsheets/d/17_5jhUWeYy7WD6OY79Kdyn3ngoB-3Y82YF0yaRpQSv8/edit#gid=0

Thank you for your attention!

Questions?