Neural Translation for the European Union (NTEU) & Multilingual Anonymisation for Public Administrations (MAPA)

Manuel Herranz

Pangeanic

NTEU: Neural Translation for EU

MAPA: Multilingual Anonymization for Public Administrations

8.06.2022

Language matrix (24 x 23=552 engines)

	BG	CS	DA	DE	EL	ES	ET	FI	FR	GA	HR	HU	IT	LT	LV	MT	NL	PL	PT	RO	SL	SK	SV
BG		·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
CS	·		·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
DA	·	·		·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
DE	·	·	·		·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
EL	·	·	·	·		·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
ES	·	·	·	·	·		·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
ET	·	·	·	·	·	·		·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
FI	·	·	·	·	·	·	·		·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
FR	·	·	·	·	·	·	·	·		·	·	·	·	·	·	·	·	·	·	·	·	·	·
GA	·	·	·	·	·	·	·	·	·		·	·	·	·	·	·	·	·	·	·	·	·	·
HR	·	·	·	·	·	·	·	·	·	·		·	·	·	·	·	·	·	·	·	·	·	·
HU	·	·	·	·	·	·	·	·	·	·	·		·	·	·	·	·	·	·	·	·	·	·
IT	·	·	·	·	·	·	·	·	·	·	·	·		·	·	·	·	·	·	·	·	·	·
LT	·	·	·	·	·	·	·	·	·	·	·	·	·		·	·	·	·	·	·	·	·	·
LV	·	·	·	·	·	·	·	·	·	·	·	·	·	·		·	·	·	·	·	·	·	·
MT	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·		·	·	·	·	·	·	·
NL	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·		·	·	·	·	·	·
PL	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·		·	·	·	·	·
PT	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·		·	·	·	·
RO	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·		·	·	·
SL	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·		·	·
SK	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·		·
SV	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	

EN

EN

pangeanic

Spanish, Portuguese, Italian, Dutch, Maltese, Polish, Czech, **French**

Romanian, German, English, Bulgarian, Hungarian, Slovene, Greek, English, Irish

KantanMT.com

Latvian, Estonian, Lithuanian, Finnish, Swedish, Danish, Croatian, Slovak

TILDE

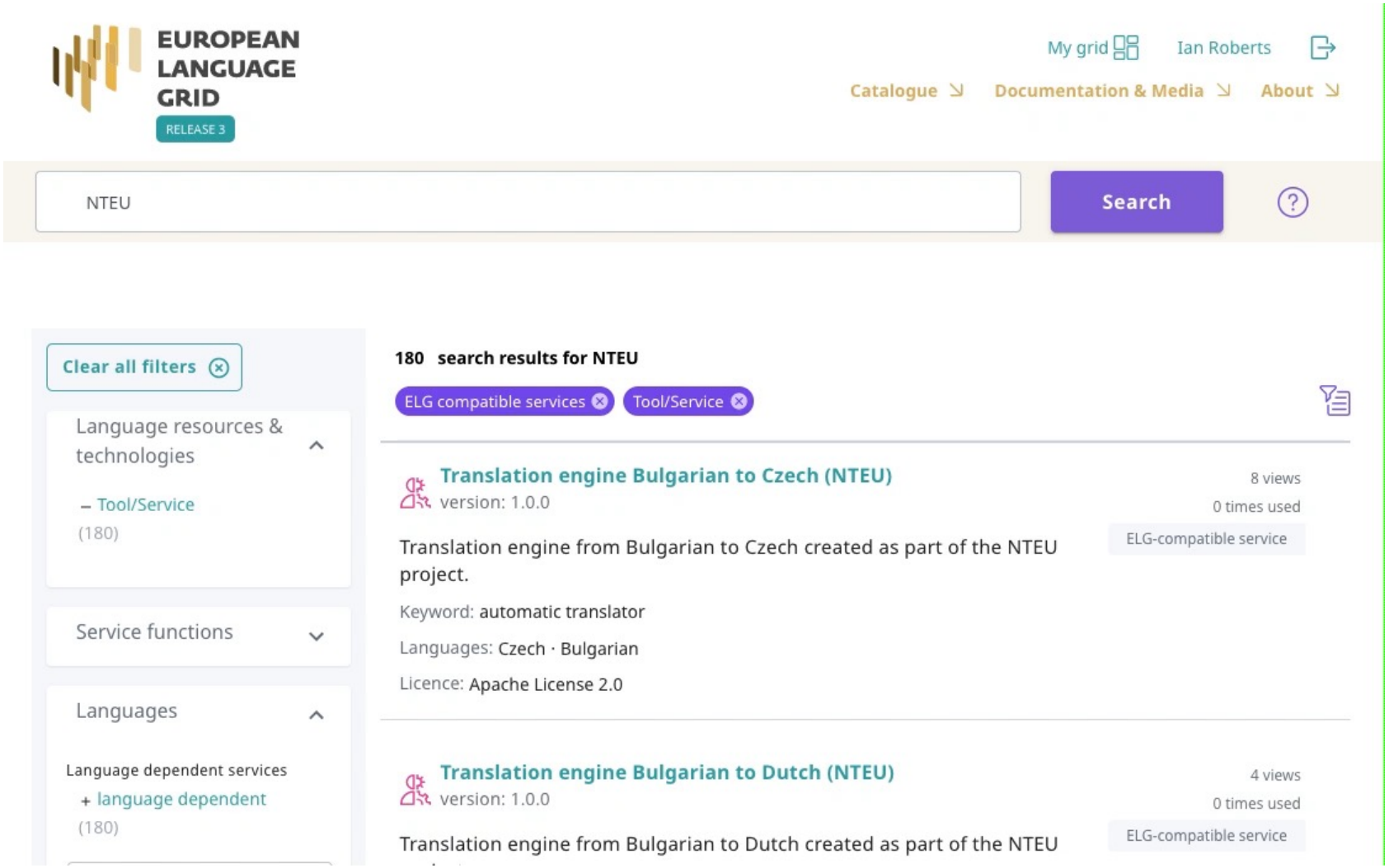
Evaluation in NTEU: Human+Automatic

- Domain: **administrative** language from the DGT.
- **Same test dataset** for all languages.
- Real documents, translated by humans into the 24 languages.
- Whole documents, not randomly extracted sentences

NTEU in ELG

- So far **180** of the 552 models are available on ELG
- Remainder are work-in-progress, expected by July

NTEU in ELG

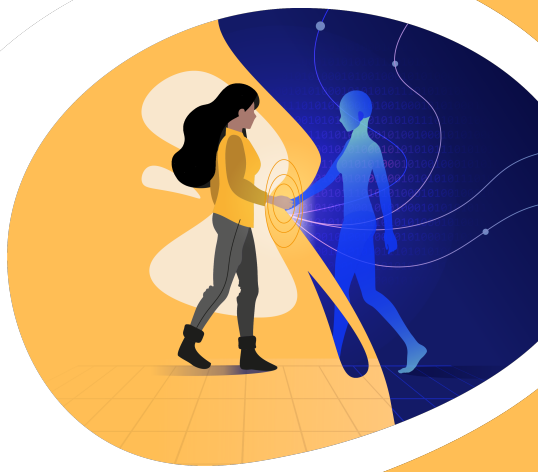

EUROPEAN LANGUAGE GRID
RELEASE 3
My grid
Ian Roberts
Catalogue
Documentation & Media
About
NTEU
Search
Clear all filters
180 search results for NTEU
ELG compatible services
Tool/Service
Language resources & technologies
– Tool/Service
(180)
Service functions
Languages
Language dependent services
+ language dependent
(180)
Translation engine Bulgarian to Czech (NTEU)
version: 1.0.0
8 views
0 times used
ELG-compatible service
Translation engine from Bulgarian to Czech created as part of the NTEU project.
Keyword: automatic translator
Languages: Czech · Bulgarian
Licence: Apache License 2.0
Translation engine Bulgarian to Dutch (NTEU)
version: 1.0.0
4 views
0 times used
ELG-compatible service
Translation engine from Bulgarian to Dutch created as part of the NTEU

MAPA

Multilingual **A**nonymization for **P**ublic **A**dministrations

MAPA

- Fine-tune BERT to detect Named Entities **in the 24 official languages of EU**
- Provide basic GPRD-compliance to PA's
- Docker version **available soon at ELG!** With custom models for legal in several languages
- https://mapa-project.eu
- Successful deployments: Spanish Ministry of Justice, DG Justice (EU Ombudsman x 27 countries)



MAPA Services in ELG

EUROPEAN LANGUAGE GRID
RELEASE 3

My grid | Ian Roberts

Catalogue | Documentation & Media | About

Go to catalogue

MAPA Anonymisation Service: Multilingual

Version: 1.0.8 (07/10/2021)

ELG-compatible service

Actions

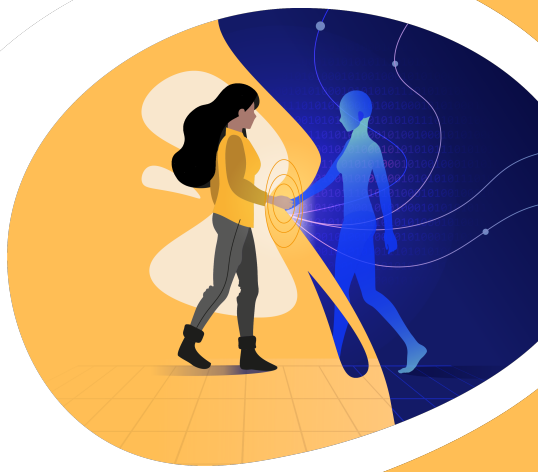
Keyword	Domain	Intended application
Anonymisation	Administration	Anonymization

Overview | Download/Run | Try out | Code samples

ELG-compatible version of the anonymisation web service from the MAPA project. The service detects mentions in the text that could identify real-world entities (people, organizations, l

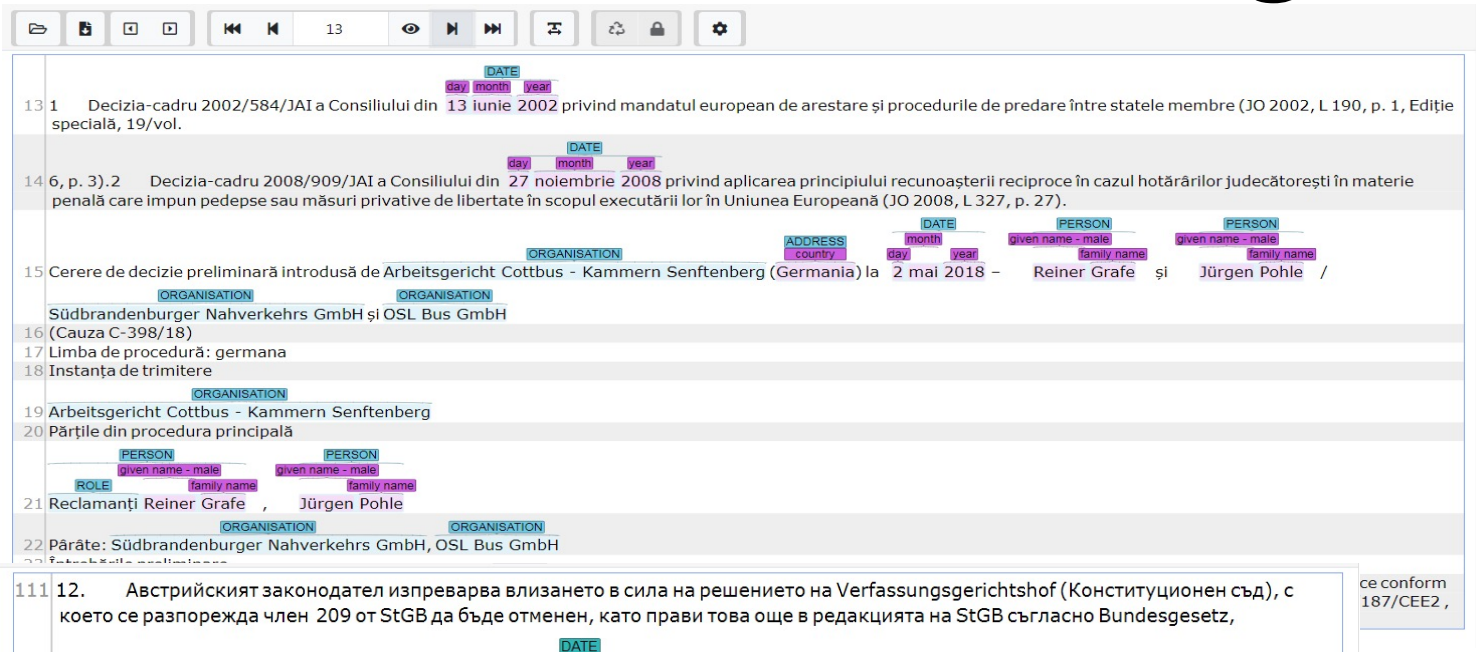
Read more

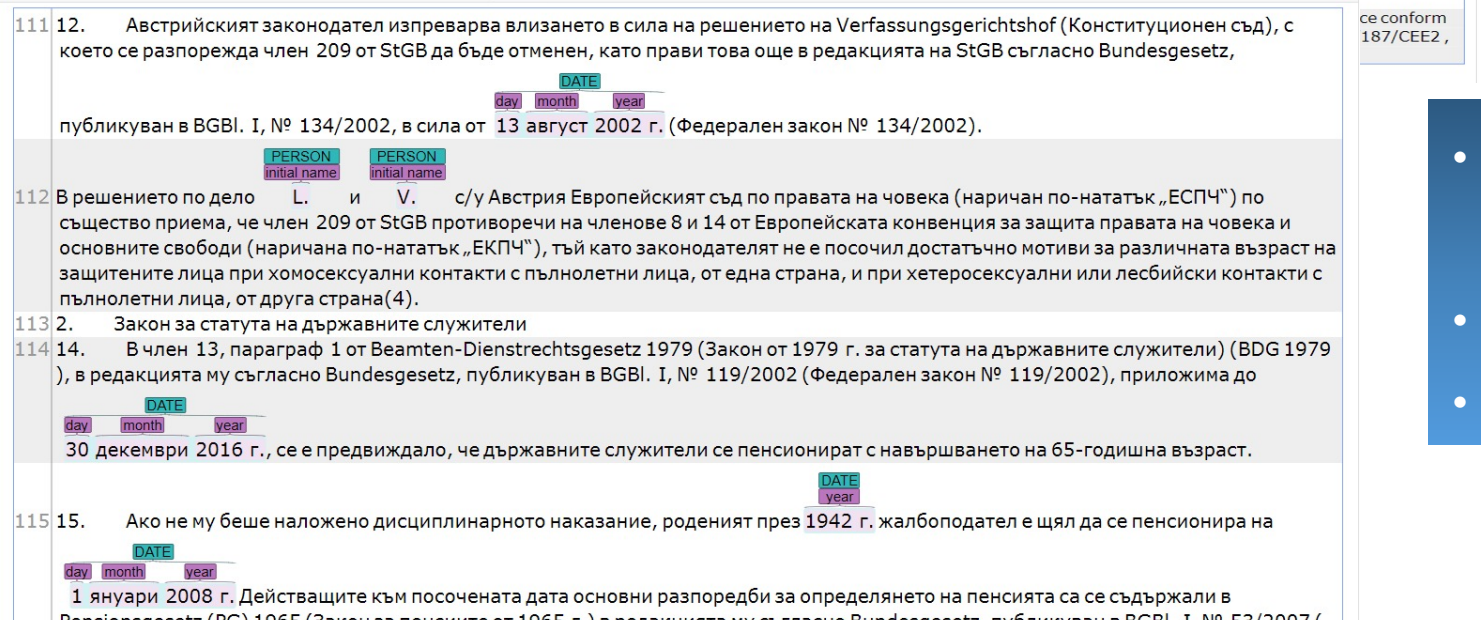
Views	Times used
0	0

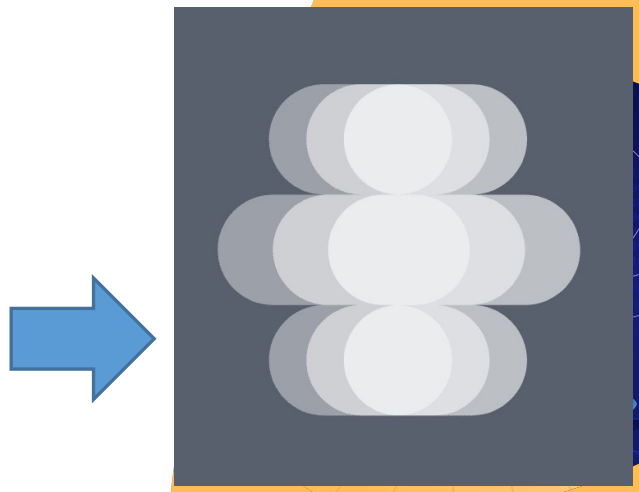


www.pangeanic.com

MAPA: Training Data Acquisition







- Manual data annotation
  in all EU official languages
- Repositories by data partners.
- Same annotation across languages

www.pangeanic.com

Manuel Herranz
CEO

pangeanic

THANKS!