

# PROMT Systems for WMT 2021 Terminology Translation Task

Alexander Molchanov, Vladislav Kovalenko & Fedor Bykov, PROMT LLC

- Two directions
  - 🇬🇧 EN-FR 🇫🇷
  - 🇬🇧 EN-RU 🇷🇺
- Four models
- Rank 1 in both directions
- Two approaches

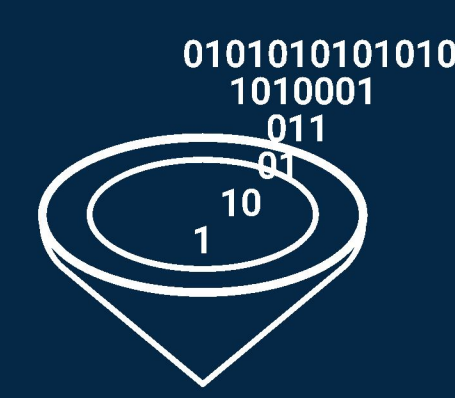
## Models

- Baseline transformers
- MarianNMT-based
- Joint BPE
- Shared vocabulary
- Iterative back-translation
- Fine-tuning using the provided terminology glossaries

## Data



- All Provided constrained data
- Back-translations (newscrawl)



- Data filtering
  - basic filtering
  - language identification
  - bicleaner
  - scoring with intermediate NMT models

### Final data

🇬🇧 EN-FR 🇫🇷 75.8M human sentence pairs

🇬🇧 EN-RU 🇷🇺 18.7M human sentence pairs

## Terminology translation

### PROMT SmartND

- Based on the PROMT RBMT technology
- No need for training/fine-tuning

### Soft-constrained

- Based on the paper "Training neural machine translation to apply terminology constraints." by Dinu et al. with modifications
- Requires specific baseline training

### Source

Find a COVID-19 testing center.

### Translation

Trouvez un **centre de test COVID-19**.

Smart ND module

### Preprocessed source

Find a <term> COVID-19 testing center <term\_end>  
centre de dépistage du COVID-19 <term\_translation> .

### Final Translation

Trouvez un **centre de dépistage du COVID-19**.

## Results

🇬🇧 EN-FR 🇫🇷

	BLEU	EM	WO (2)	WO (3)	COMET	Rank (EM)	Rank (COMET)
Soft-constrained	47.69	<b>0.974</b>	<b>0.359</b>	<b>0.352</b>	0.752	<b>1-3</b>	3
SmartND	47.89	0.966	0.357	0.348	0.746	<b>1-3</b>	4-5
Best Scores	49.60	0.974	0.359	0.352	0.781	-	-

🇬🇧 EN-RU 🇷🇺

	BLEU	EM	WO (2)	WO (3)	COMET	Rank (EM)	Rank (COMET)
Soft-constrained	31.06	<b>0.909</b>	<b>0.254</b>	<b>0.255</b>	0.631	<b>1</b>	<b>1-2</b>
SmartND	<b>31.92</b>	0.788	0.243	0.241	<b>0.634</b>	10	<b>1-2</b>
Best Scores	31.92	0.909	0.254	0.255	0.634	-	-