Findings of a Machine Translation Shared Task Focused on Covid-19 Related Documents

F. Casacuberta[1,2], A. Ceausu[3], K. Choukri[4], M. Deligiannis[5], M. Domingo[1,2], M. García-Martínez[6], M. Herranz[6], G. Jacquet[7], V. Papavassiliou[5], S. Piperidis[5], P. Prokopidis[5], D. Roussis[5], M. Hadj Salah[4]

[1] PRHLT Research Center, Universitat Politècnica de València, {fcn, midobal}@prhlt.upv.es
[2] ValgrAI - Valencian Graduate School and Research Network for Artificial Intelligence
[3] Court of Justice of the EU, Alexandru.Ceausu@curia.europa.eu
[4] Evaluations and Language resources Distribution Agency, {choukri, marwa}@elda.org
[5] Athena Research Center, {mdel, vpapa, spip, prokopis, dimitris}.roussis@athenarc.gr
[6] Pangeanic S. L. - PangeaMT, {m.garcia, m.herranz}@pangeanic.com
[7] European Commission, guillaume.jacquet@ec.europa.eu

Introduction

- In emergency situations, the public needs to aggregate and summarize different sources of information.
- They need to complement different pieces of information, resolving possible inconsistencies and preventing misinformation.
- This should happen across multiple languages, sources and levels of linguistic knowledge.
- As a response to the Covid-19 crisis, the Covid-19 MLIA initiative organized a community evaluation effort.
- This initiative consisted of three NLP tasks:
  1. Information extraction.
  2. Multilingual semantic search.
  3. **Machine translation**.

Task description

- **Goal**: benchmark MT systems focused on Covid-19 related documents for several language pairs.
- Two rounds over a two-year period.
- Categories:
  - **Constrained**: only data provided by the organizers.
  - **Unconstrained**: no data restriction.
- Language pairs:
  - English–German.
  - English–French.
  - English–Spanish.
  - English–Italian.
  - English–Modern Greek.
  - English–Swedish.
  - English–Arabic (round two only).

Examples

- *30% of children and adults infected with measles can develop complications.*
- *The first dose is given between 10 and 18 months of age in European countries.*
- *Note: The information contained in this factsheet is intended for the purpose of general information and should not be used as a substitute for the individual expertise and judgement of a healthcare professional.*

Data collection

1. Collected an updated version of the European Medicines Agency.
2. Applied new methods for text extraction from *pdf* files, sentence splitting, sentence alignment and parallel corpus filtering.
3. Collected and process medical-related multilingual collections by the Publications Office of EU.
4. Identification of several bilingual websites with Covid-19-related content.
5. Mined sentence alignments using LASER toolkit.
6. Filtering methods.

Corpora generation

1. Computed the ratio of data from each different source over the total data.
2. Computed the average number of words per segment over this set.
3. Constituted a subset [0.7 * average words per segment, 1.3 * average words per segment].
4. Sorted this subset (from best to worst) according to its alignment score.
5. Selected the best 8000 segments, making sure to maintain the ratio of data from the different sources from step 1.
6. Shuffled those segments and selected half of them for validation and the other half for test.

Evaluation

- Main:
  - BLEU.
- Round 1:
  - ChrF.
- Round 2:
  - TER, BEER.

Example of results

English–Swedish

	Rank	Team	Description	BLEU [↑]	chrF [↑]
Constrained	1	PROMT	multilingual	30.7	0.595
		Lingua Custodia	multilingual	30.4	0.589
		CUNI-MT	transfer2	30.1	0.590
	2	CUNI-MT	transfer	28.5	0.578
	-	Baseline	Transformer	27.8	0.566
	3	CUNI-MT	base	26.6	0.561
	4	CUNI-MTIR	r1	25.1	0.541
	-	Baseline	RNN	19.2	0.481
	5	TARJAMA-AI	base	11.2	0.443
Unco.	1	PROMT	transformer	41.3	0.671
	2	CUNI-MTIR	r1	24.0	0.514

Conclusions

- This work presents a community evaluation effort to improve the generation of MT systems as a response to a global problem.
- The initiative consisted of generating specialized corpus for a new and important topic: Covid-19.
- Round 1:
  - 6 language pairs.
  - 8 teams.
  - Transformer architecture.
  - Best approaches:
    - Multilingual MT.
    - Transfer learning.
- Round 2:
  - 7 language pairs.
  - 5 teams.
  - Data cleaning yielded great improvements.
  - Transformer and big Transformer architectures.
  - Beast approaches:
    - Multilingual MT (specially for low-resourced languages).

Acknowledgements

The Covid-19 MLIA @ Eval initiative has received support from the European Commission, the European Language Resources Coordination (ELRC), European Language Resources Association (ELRA), the European Research Infrastructure for Language Resources and Technology (CLARIN), the CLEF Initiative and the Joint Research Centre (JRC).