

Named Entity Translation Website

INTRODUCTION

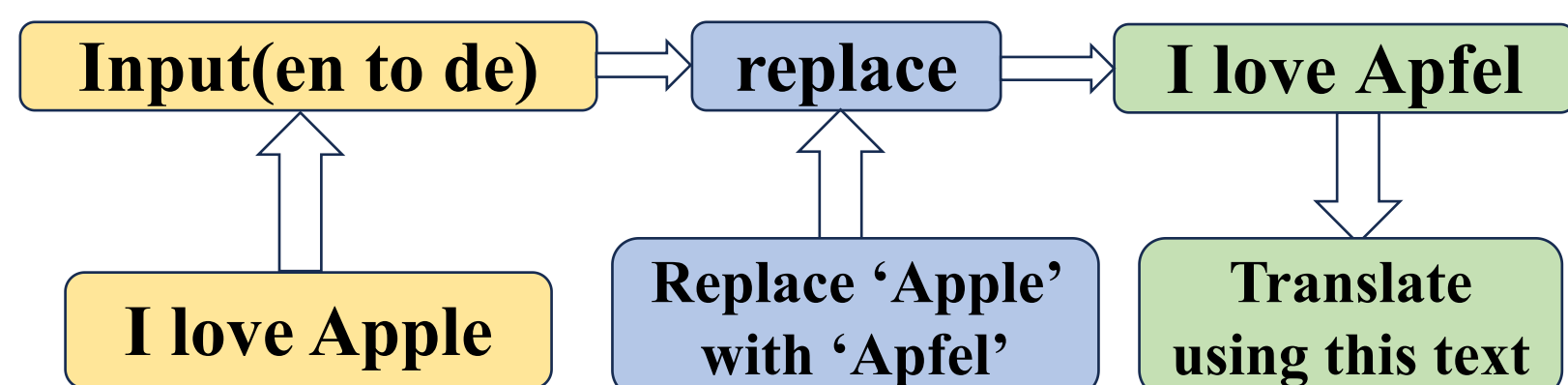
Named Entity

- A named entity is an entity identified by its name. Simply put, if we hear a name, we can know which specific thing it is, and that thing is a named entity.
- For example, Bird, RMB, Ipad.



Named Entity Translation(NET)

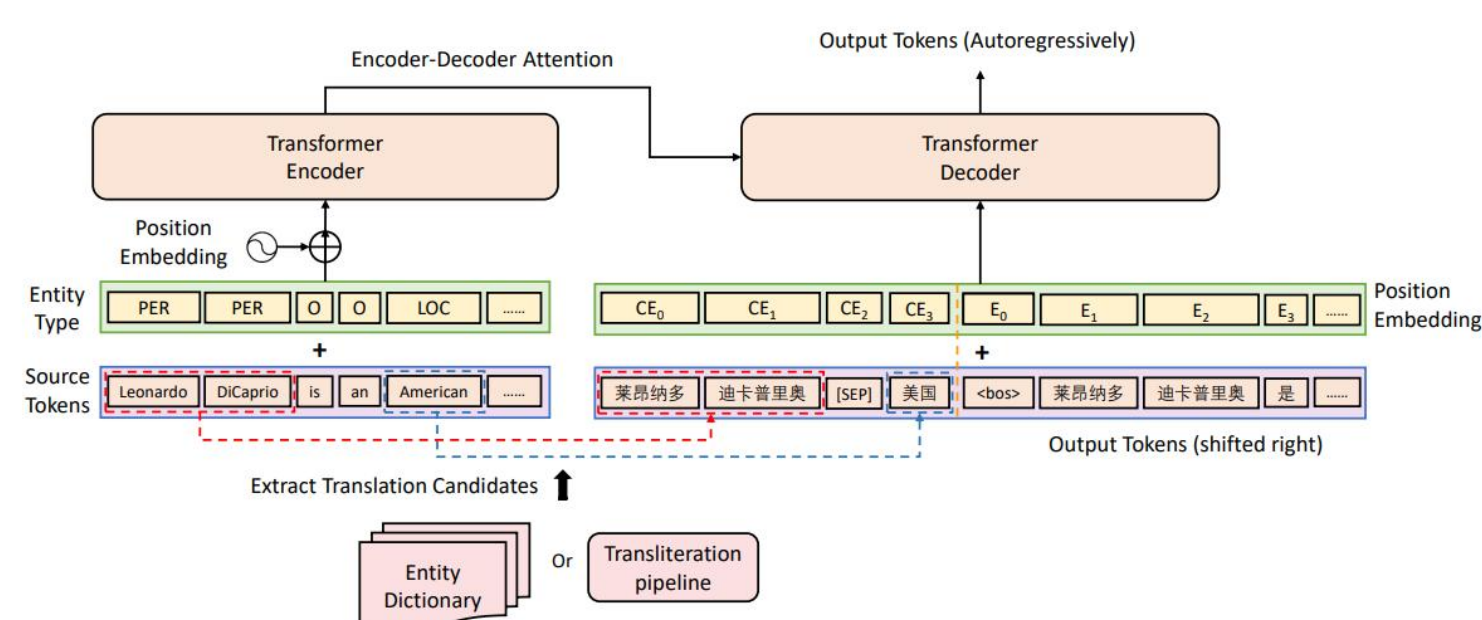
- NET is a technique for translating named entities from one language to another.
- Named Entity Translation plays an important role in the interaction between humans and machines linked by sound messages, which often contain named entities such as names of people, places, , and organizations.



METHOD

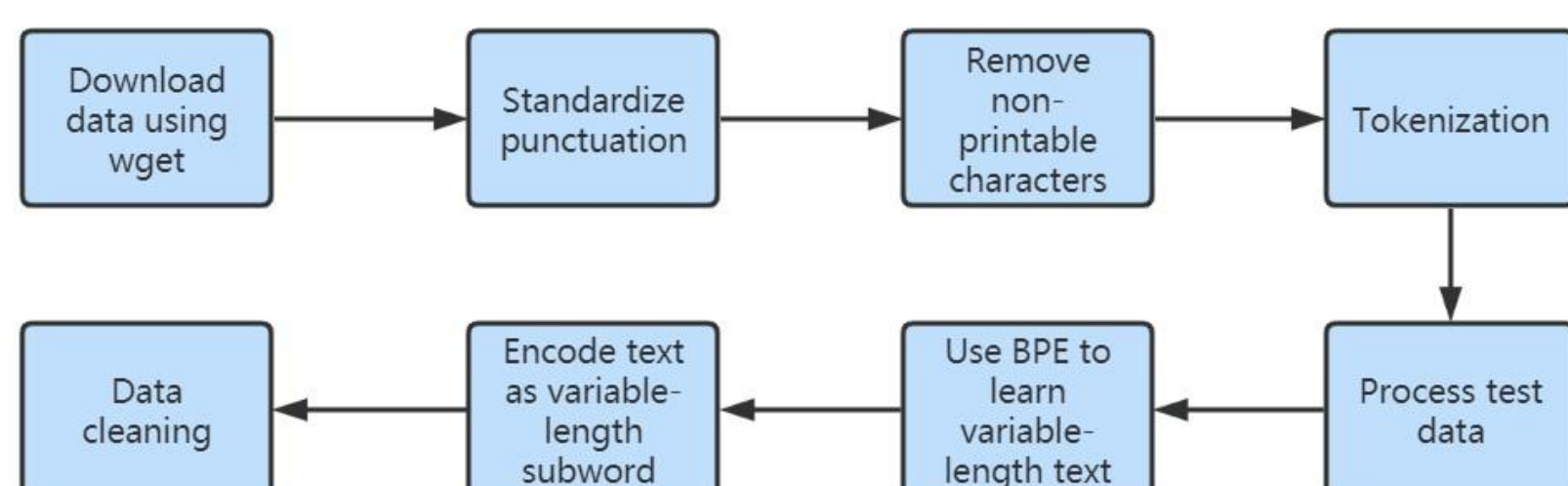
Extract-and-Attend

- Extract-and-Attend approach, where the translation candidates are extracted and added as a prefix of the decoder input.
- Entity type embeddings are added to the source input (e.g., 'PER' for person names, 'LOC' for locations and 'O' for other tokens other than entities).
- Independent position embeddings are used for the translation candidates and the shifted output tokens (i.e., 'CE' for translation candidates and 'E' for output tokens).



Data preprocessing

- Download training data.
- Process training data , standardize the data with Punctuation, remove non printing characters, and use tokenizer. perl for word segmentation.
- Using BPE to learn variable length text.
- Using script to encode a variable length sequence of sub word units.
- Run the perl script \$CLEAN for data cleaning.

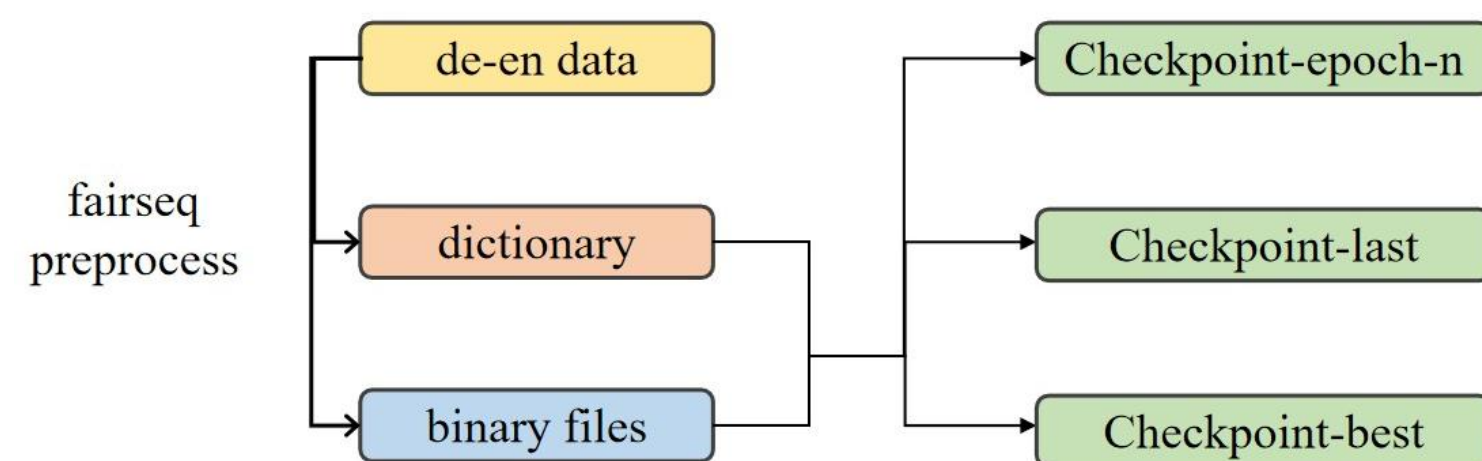


model training

- Specify the GPU to use, and then train the model by calling the fairseq train tool.
- Specify the model architecture as a variant of the transformer model: transformer_iwslt_de_en.
- Use Adam as the optimizer and calculate BLEU evaluation metrics during training, saving the model parameters after each -

model training

- round of training during the training process.
- Save the model with the highest evaluation index during the training process, mark it as best, and mark the model obtained from the last round of training as last.



RESULT DISPLAY

Comparative Experiment of Papers

- Comparative experiments have shown that the method we use can effectively improve the accuracy and overall translation quality of entity translation.

Model	$En \rightarrow Ru$		$Ru \rightarrow En$		$En \rightarrow Zh$		$Zh \rightarrow En$	
	BLEU	COMET	BLEU	COMET	BLEU	COMET	BLEU	COMET
Transformer	31.83	52.2	34.63	54.0	26.32	34.8	27.45	41.5
Transformer w/ Dictionary	31.85	53.6	34.67	56.1	26.36	38.1	27.49	43.2
Replacement	30.52	55.2	32.01	56.7	25.92	41.4	27.21	45.0
Placeholder	31.88	57.6	34.72	59.1	26.41	42.9	27.50	47.2
Annotation	31.91	59.4	34.84	60.5	26.44	45.8	27.73	48.0
Multi-task	31.88	57.8	34.76	60.3	26.38	45.0	27.64	47.4
Extract & Attend (ours)	32.68	62.2	35.41	63.5	26.79	48.6	27.98	50.1

Visualization interface

- We use HTML and CSS to create a visual web interface.



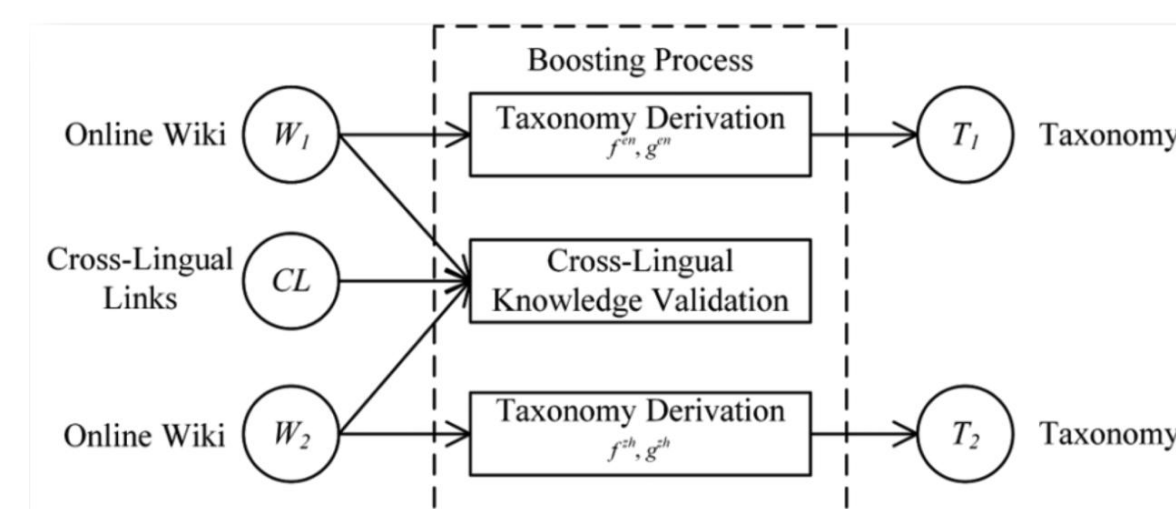
ADVANTAGE And APPLICATION

Extremely high accuracy

- The named entity translation system greatly improves translation accuracy and lays the foundation for future high-precision or high-level translation.

Convenient translation of high-level papers

- Our system can be used to quickly and conveniently translate high-level papers by replacing special vocabulary corpora.



Cross language information retrieval

- Named entity translation can help translate named entities in query words into the target language for accurate retrieval in target language text.

Cross language Knowledge graph construction

- The Knowledge graph is composed of various entities and relationships between entities, and named entity translation can help to correspond cross language entities, thus promoting the connection and sharing of different language Knowledge graph.