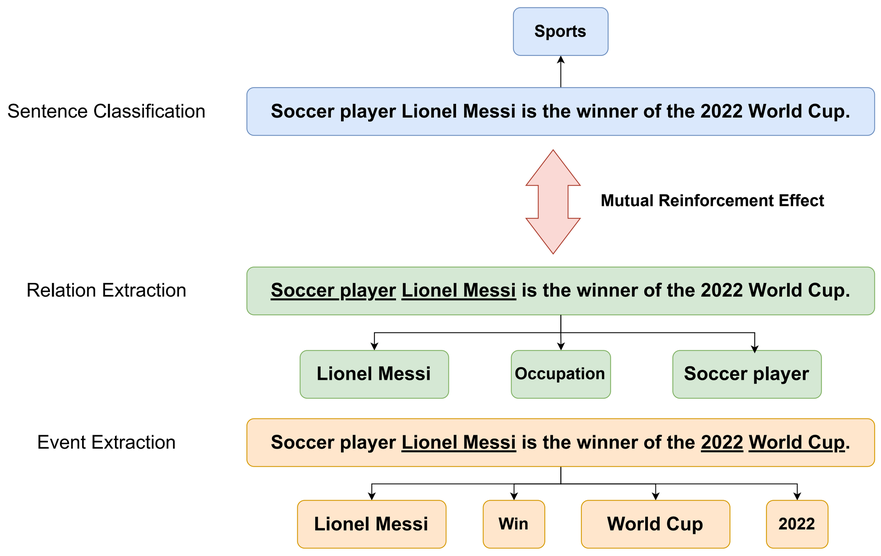
GIELLM: Japanese General Information Extraction Large Language Model Utilizing Mutual Reinforcement Effect

Chengguang Gan[1] Qinghao Zhang[2] Tatsunori Mori[1]

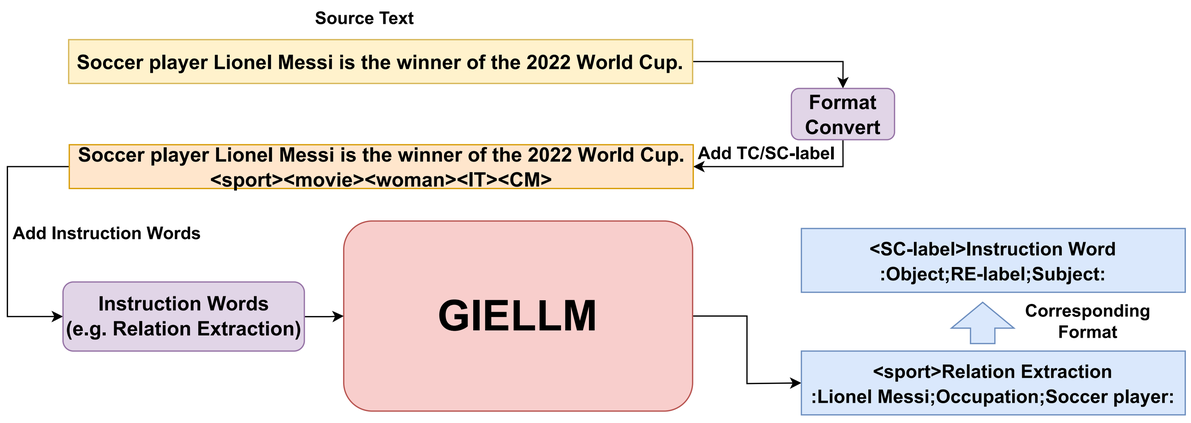
[1]Yokohama National University, Japan [2]Pusan National University, South Korea

Introduction

Novel Approach: Introduces a Generalized Large Language Model for Information Extraction (IE) that handles various IE sub-tasks.
Mutual Reinforcement Effect (MRE): Utilizes a mixed dataset for training and testing, allowing dual-level classification on the same textual data.
Contribution: Includes a new Format Converter, the Japanese TCREE dataset, and a fine-tuned Japanese GIELLM using multiple mix datasets.
Illustration of MRE: Shows the reinforcement between word-level and text-level classifications.

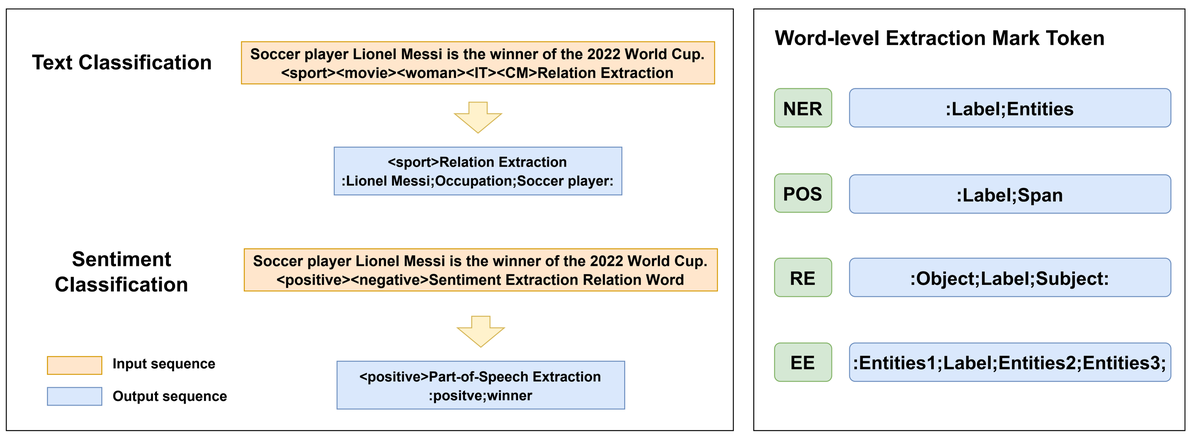


Model Workflow: Depicts how source text is processed and output to GIELLM.

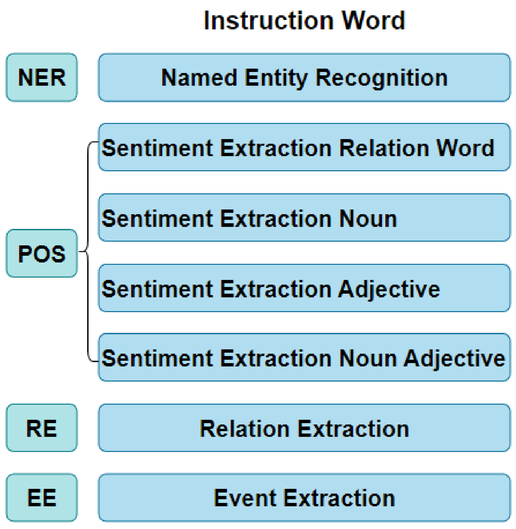


Japanese General Information Extraction Large Language Model

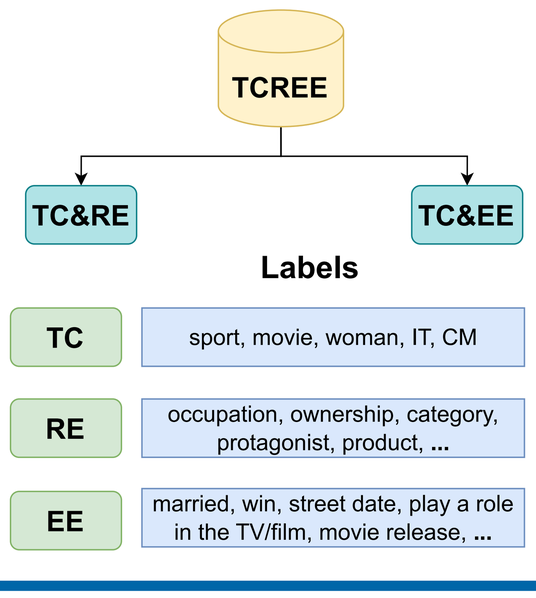
GIELLM Workflow: Processes source text through Format Converter and Instruction Words (IW) to create input sequence.
Unified Format: Reduces input text length and speeds up model inference, using specific marks for labels and entities.
Instruction Words (IW): Concise directives tailored for specific tasks, replacing the need for lengthy ICL samples and IL directives.
TCREE Dataset: Fills a gap in MRE mix dataset by providing datasets for IE domain subtasks.
Format Converter Examples: Demonstrates the Format Converter in different IE subtasks.



Instruction Words in IE Subtasks: Shows IWs in different IE subtasks.

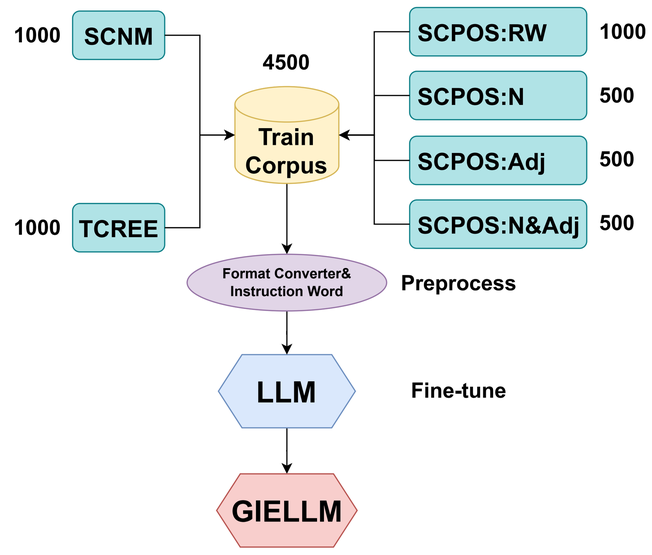


TCREE Dataset Labels: Labels of the TCREE dataset.



Experiment

Training Corpus: Consists of 4500 sequences, fine-tuning LLaMA2's 7B and 13B parameter models, and LLM-jp.
Training Details: Uses A800 80GB * 8, epoch of 3, learning rate of 1e-5, and per device batch size of 1.
Evaluation Metric: Accuracy at both word-level and text-level sequences.
Performance: GIELLM outperforms other models in 14 out of 18 accuracy metrics across six datasets.
Bilingual vs. Monolingual: GIELLM-13B-JP generally outperforms GIELLM-13B and other English-trained LLMs.
Training Corpus Composition: Composition of the training corpus.



Dataset Statistics: Statistical data of training and testing sets for six MRE-based Japanese datasets.

Datasets	train set	test set
SCNM	1000	4343
SCPOS:RW	1000	1000
SCPOS:N	500	187028
SCPOS:Adj	500	187028
SCPOS:N & Adj	500	187028
TCREE	1000	1000

Accuracy Results: Accuracy of Six Mix Datasets in different methods or models.

	SCNM			SCPOS: RW			SCPOS: Adj & N		
Accuracy	TL	WL	ALL	TL	WL	ALL	TL	WL	ALL
SLG Framework	**88.89**	**81.96**	**72.41**	88.21	55.57	17.28	87.30	26.22	1.60
USA-7B	-	-	-	**89.60**	56.32	18.10	90.20	60.09	3.97
GPT-3.5-Turbo	49.46	11.87	6.97	53.60	14.99	1.60	73.20	10.34	0.13
GIELLM-7B	85.70	63.16	54.29	86.31	66.90	25.37	92.27	54.23	3.23
GIELLM-13B	85.06	54.06	45.96	83.33	65.25	24.75	90.43	**63.98**	**6.27**
GIELLM-13B-JP	86.44	62.95	54.43	85.83	**67.62**	**26.15**	**93.2**	48.04	3.43

	SCPOS: N			SCPOS: Adj			TCREE		
Accuracy	TL	WL	ALL	TL	WL	ALL	TL	WL	ALL
SLG Framework	89.50	27.62	3.00	83.00	73.84	52.47	96.49	73.12	71.72
USA-7B	91.50	62.41	6.86	92.17	64.94	50.90	-	-	-
GPT-3.5-Turbo	73.83	10.44	0.23	78.83	15.45	9.87	80.02	7.33	6.73
GIELLM-7B	92.10	58.33	4.60	91.6	75.71	58.8	97.19	77.01	75.80
GIELLM-13B	90.13	**68.28**	**9.80**	91.9	77.58	60.70	94.58	74.90	73.29
GIELLM-13B-JP	**92.43**	49.63	4.43	**93.47**	**78.71**	**63.33**	**97.49**	**78.51**	**77.41**

Analysis MRE in TCREE task

MRE Analysis: Significant drop in accuracy for separate tasks compared to mixed dataset, confirming the MRE's impact on performance.
MRE Accuracy Comparison: Table showing accuracy comparison of TC and REE tasks in separate and mixed datasets.

	Accuracy		
Dataset	TL	WL	ALL
TCREE	**96.49**	**73.12**	**71.72**
Single TC	95.89	-	-
Single REE	-	58.73	-