

TASK 2: CONTEXT-ENRICHED MULTILINGUAL NAMED ENTITY RECOGNITION USING KNOWLEDGE BASES

Iker García-Ferrero¹, Jon Ander Campos¹, Oscar Sainz¹, Ander Salaberria¹, Dan Roth²
 {iker.garciaf,jonander.campos,oscar.sainz,ander.salaberria}@ehu.eus¹ danroth@seas.upenn.edu²

Mary Pickford and Margaret Thatcher are some of the famous guests who stayed at the hotel.

1. ENTITY BOUNDARY DETECTION

A XLM-RoBERTa sequence-labeller is fine-tuned to detect the boundaries of the candidate entities.

Mary Pickford and Margaret Thatcher are some of the famous guests who stayed at the hotel.

LANG	F1
EN	90.1
ZH	90.5
MULTI	90.5

F1-score of the Entity Boundary Detection step in the development split.

2. WIKIDATA ENTITY LINKING

mGENRE is used to link the WikiData ID for every span detected.

Q104109

Q104109

Mary Pickford and Margaret Thatcher are some of the famous guests who stayed at the hotel.

PAPER

CODE



3. ENTITY CATEGORY CLASSIFIER

A XLM-RoBERTa model uses the context, the Wikidata and Wikipedia information to categorize the entity.

INITIAL CONTEXT	WIKIDATA DESCRIPTION	WIKIDATA ARGUMENTS	WIKIPEDIA SUMMARY	EN	ZH
✓				79.1	81.1
✓	✓			82.2	84.3
✓		✓		88.1	88.2
✓			✓	86.0	86.1
✓	✓	✓		89.1	86.1
✓	✓	✓	✓	90.2	89.1

MARY PICKFORD - WIKIPEDIA

Gladys Marie Smith
 (April 8, 1892 - May 29, 1979),
 known professionally as [...]

Q104109 - WIKIDATA

Canadian actress and
 producer (1892–1979)
 - INSTANCEOF: Human
 - OCCUPATION: Film Producer

Ablation study on different external knowledge used by the model. F1 scores using gold entity boundaries.

ARTIST

Mary Pickford and Margaret Thatcher are some of the famous guests who stayed at the hotel.

OVERALL RESULTS

LANG	EN	ES	SV	UK	PT	FR	FA	DE	ZH	HI	BN	IT	MULTI
XLM-RoBERTa _{LARGE}	59.8	62.6	65.8	65.7	61.1	63.0	57.2	62.9	51.9	68.1	66.0	65.2	69.1
OUR SYSTEM	72.8	73.8	76.5	75.3	72.3	74.5	69.5	80.4	64.9	79.6	79.0	74.7	74.7
RANKING	7°	4°	3°	3°	5°	5°	5°	3°	8°	3°	4°	6°	4°

Results for all tracks compared to our baseline (XLM-RoBERTa_{large}). We also report the position of our system on each leaderboard.