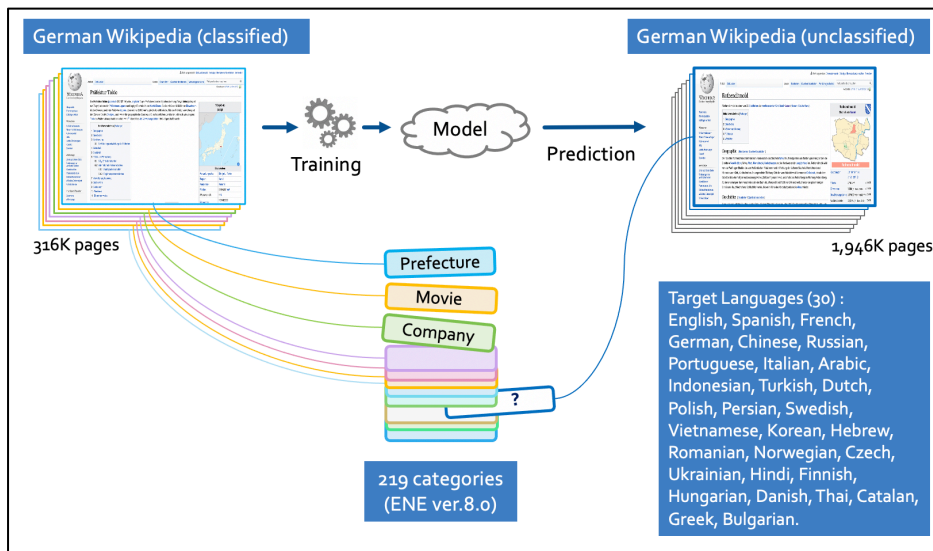


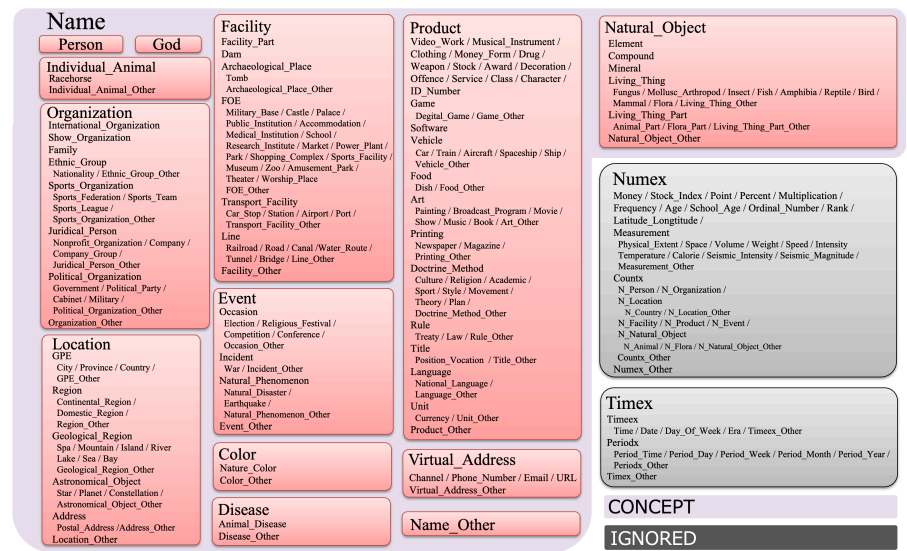
### SHINRA2020-ML: Classification task

- Task: to classify Wikipedia pages into 219 ENE categories (multi-label classification)
- Target languages: Participants can choose one or more of the 30 languages
- Evaluation metrics: Micro-average F1 measure

### Overview



### Extended Named Entity



### Participants

# of groups	10 (7 active participants)
nationality	Japan(4), Vietnam (2), India (1), Taiwan (1), Australia(1), Finland(1), Portugal(1)
affiliation types	University (6), Company (4), Institute (1)
target languages	8: Arabic, French 7: Chinese, German, Hindi, Italian, Portuguese, Spanish, Thai, Turkish 6: Bulgarian, Czech, Dutch, English, Indonesian, Korean, Norwegian, Polish, Russian, Vietnamese 5: Catalan, Danish, Finnish, Hebrew, Hungarian, Persian, Romanian, Ukrainian 4: Greek, Swedish
# of target languages	30(4), 28(1), 15(1), 9(1), 6(1), 1(2)

### Results

Group ID	FPTAI	LIAT	PriBL	PriBL	RH312	ousia	uomfj	uomfj	uomfj	FPTAI	HUKB	HUKB	HUKB	LIAT	
Method ID	BERT	ML-BERT	BERTOR U	BERTLIN CONCAT	RnnGnnXi	RobERTa +wikiZvec +wikidata	jointrep	jointrepPo	jointrepPo	BERT	AB	ABC	AC	ML-BERT	
Late Submission										Y	Y	Y	Y	Y	
ar	Arabic	73.25	63.16	76.27	75.45	-	70.52	64.55	64.55	73.25	30.98	30.98	13.51	-	
bg	Bulgarian	83.77	75.20	-	-	82.13	-	83.07	83.07	83.28	60.86	61.06	28.09	-	
ca	Catalan, Valencian	52.55	76.28	-	-	-	-	79.82	79.82	81.10	42.34	42.54	16.26	-	
cs	Czech	84.47	78.46	-	81.19	-	-	81.29	81.29	83.74	52.61	52.61	18.86	-	
da	Danish	82.30	74.80	-	-	-	-	80.56	80.56	81.74	49.01	49.01	13.99	-	
de	German	22.62	78.49	80.24	79.83	-	81.86	81.03	81.03	81.03	53.72	53.82	26.81	-	
el	Greek, Modern (1453-)	84.40	72.43	-	-	-	-	-	-	84.10	7.51	7.51	7.51	-	
en	English	82.23	78.56	81.27	80.12	-	-	82.73	82.57	82.68	45.11	45.11	11.92	-	
es	Spanish, Castilian	80.60	77.73	80.30	80.72	-	80.94	81.39	81.39	80.80	49.21	49.11	19.50	-	
fa	Persian	81.70	75.42	-	-	-	-	80.38	80.38	81.52	45.59	45.59	15.66	-	
fi	Finnish	83.62	79.13	-	-	-	-	80.91	80.91	80.91	53.15	53.45	17.06	-	
fr	French	21.59	76.88	77.93	78.52	80.31	81.01	78.21	78.21	80.68	43.84	43.74	11.23	-	
he	Hebrew	83.79	79.11	-	-	-	-	81.09	81.09	81.09	59.95	60.05	15.78	-	
hi	Hindi	76.43	16.49	-	-	71.70	69.75	66.67	66.67	75.65	39.70	39.51	22.02	-	
hu	Hungarian	85.46	78.93	-	-	-	-	85.02	85.02	84.78	69.15	69.44	26.09	-	
id	Indonesian	81.93	72.45	-	-	77.56	-	78.51	78.51	81.65	44.07	44.47	16.28	-	
it	Italian	26.55	81.36	81.92	81.89	-	81.21	82.02	82.02	82.81	45.55	45.55	12.06	-	
ko	Korean	83.67	80.38	81.51	81.04	-	-	82.51	82.51	83.77	63.68	63.98	13.95	-	
nl	Dutch, Flemish	83.29	79.86	80.95	81.26	-	-	81.64	81.64	83.17	42.36	42.45	17.12	-	
no	Norwegian	80.53	76.50	-	78.39	-	-	78.79	78.79	80.17	34.58	34.58	11.33	-	
pl	Polish	84.53	80.60	82.73	83.46	-	-	84.52	84.52	84.07	62.72	63.51	32.55	-	
pt	Portuguese	83.23	78.49	82.36	81.88	-	81.40	80.87	80.87	82.70	42.32	42.82	16.10	-	
ro	Romanian, Moldavian, Moldovan	84.60	78.17	-	-	-	-	80.83	80.83	84.60	57.60	57.70	28.50	-	
ru	Russian	84.08	79.09	82.60	83.07	-	-	82.90	82.90	83.44	42.04	42.24	11.30	-	
sv	Swedish	83.18	71.63	-	-	-	-	-	-	83.44	50.32	50.62	21.98	79.58	
th	Thai	81.26	49.58	-	-	76.77	78.36	65.02	65.02	81.16	39.98	40.38	24.05	-	
tr	Turkish	86.50	77.19	84.36	83.23	83.28	-	84.85	84.85	86.03	61.88	62.48	16.73	-	
uk	Ukrainian	83.12	78.71	-	-	-	-	81.61	81.61	82.51	60.29	60.19	22.51	-	
vi	Vietnamese	80.34	75.24	-	-	-	-	77.06	77.06	77.06	80.42	60.38	60.48	22.14	-
zh	Chinese	81.25	77.97	78.38	79.37	-	79.76	78.58	78.58	80.60	21.22	21.42	17.57	-	

### Plan for SHINRA2021-ML

- Continue the same categorization task in 30 languages
- Provide the results of 2020 outputs to encourage research topics, such as Unsupervised Ensemble learning, Unsupervised active learning
- Independent from NTCIR-16
- We also plan to do SHINRA2021-JP, (Crowdsourcing and Entity linking tasks)

<http://shinra-project.info/shinra2020ml/?lang=en>  
[shinra2020ml-info@googlegroups.com](mailto:shinra2020ml-info@googlegroups.com)

