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

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: Catalán, 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)

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)

Results

Extended Named Entity

http://shinra-project.info/shinra2020ml/?lang=en
shinra2020ml-info@googlegroups.com

# of target languages
 relación

http://shinra-project.info/shinra2020ml/?lang=en
shinra2020ml-info@googlegroups.com