NLP researchers design new word representation models to address the shortcomings of traditional word embeddings. However, (1) there is no standard method to evaluate the quality of these new models, and (2) majority of existing techniques are developed only for English and limited to similarity measurements among words.

**Task Description**

- Implement a unified evaluation framework in Pytorch (Inspector Gadget and his Gadgets)
- Evaluate existing word representation models with your framework
- Analyze the correlation between your gadgets and downstream NLP tasks

**References**

Conneau et. al, What you can cram into a single \$!#* vector: Probing sentence embeddings for linguistic properties, ACL (1) 2018: 2126-2136