NOVA IMS MT Metrics Shared Task

To help you in this project I'll leave here a few links for readings that might help you develop a Machine Translation Metric.

Lexical metrics:
- NLP Metrics Made Simple: The BLEU Score: The author describes a simplified version of BLEU Metric.
- C5W3L06 Bleu Score (Optional): Video explanation of the original and complete BLEU Metric.
- The Ultimate Performance Metric in NLP: Detailed explanation of ROUGE, a widely used metric for summarisation (but applicable to Machine Translation also)

Embedding based Metrics:
- Word Mover's Distance for Text Similarity: Blogpost describing the idea behind word movers distance metric
- BERTScore: Evaluating Text Generation with BERT: Original paper that presents BERTScore

Trainable Metrics:
- COMET- the Crosslingual Optimised Metric for Evaluation of Translation: Blogpost explaining COMET a newly released MT metric that is trained to predict human quality assessments
- Evaluating Natural Language Generation with BLEURT: MT metric developed by google that is also trained to predict human quality assessments.

NOTE:
Be creative! These links are just pointers that you can use to get inspired and to know what other have been doing in this field.
There are other available metrics that are sometimes simpler and perform pretty well (e.g chrF Metric).

Important Notes:

- The delivery deadline for the project and report is May 28th, at 17:00.
- Students should not use code from each other!
- The project is to be developed in groups of up to three.