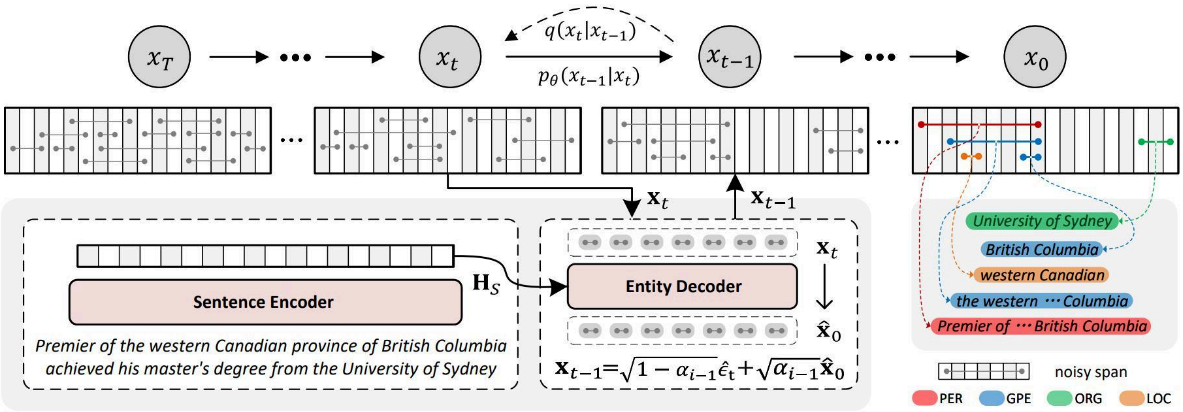
Russian Nested Named Entities

Name: Yazan Kbaili
Email: y.kbaili@innopolis.univeristy
Github link: [yazankb/NLP-Course (github.com)](https://github.com/yazankb/NLP-Course)
Codalab: Kbaili (I couldn’t make a submission because of technical issues)



DiffusionNER:

Diffusion models, typically used in generating high-quality images, operate by gradually adding noise to data and then learning to reverse this process to denoise or reconstruct the original data. DiffusionNER adapts this concept to NER, treating entity boundaries within text as data points that can be diffused and refined.

Methodology:

Forward Diffusion Process: This involves progressively adding Gaussian noise to the boundaries of known (gold) entities in the training data, gradually losing information until only noise remains.

Reverse Diffusion Process: During inference, the model generates entity boundaries from noisy data by iteratively denoising them. This reverse process uses a learned neural network model that predicts cleaner versions of the data in each step back towards the clean boundaries.

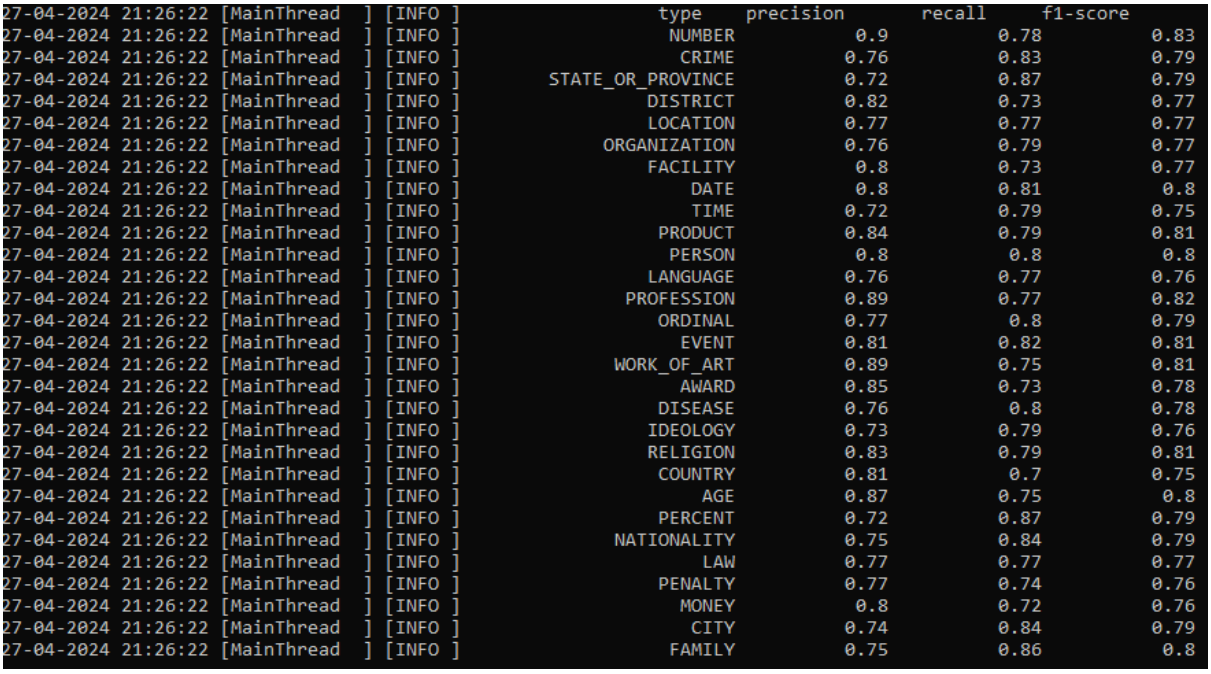
Training and Inference:

Training: The model learns by reverse diffusion, starting from fully noisy spans and attempting to reconstruct the entity boundaries. This is achieved using a sequence of denoising steps guided by the network trained on the diffusion steps.

Inference: Begins with a set of noisy spans sampled from a Gaussian distribution. These spans are then refined using the reverse diffusion process to predict the most likely entity boundaries and their type.

Results:

After splitting the data 90% training 10% validation. Here are the results for each of the entity classes for the validation sub-dataset. The results reflect a comprehensive performance and F1-scores suggest a consistent and reliable identification of entities across various categories.



```
27-04-2024 21:26:22 [MainThread  ] [INFO ]                  type    precision       recall     f1-score
27-04-2024 21:26:22 [MainThread  ] [INFO ]                   NUMBER           0.9          0.78          0.83
27-04-2024 21:26:22 [MainThread  ] [INFO ]                    CRIME          0.76          0.83          0.79
27-04-2024 21:26:22 [MainThread  ] [INFO ]        STATE_OR_PROVINCE          0.72          0.87          0.79
27-04-2024 21:26:22 [MainThread  ] [INFO ]                 DISTRICT          0.82          0.73          0.77
27-04-2024 21:26:22 [MainThread  ] [INFO ]                 LOCATION          0.77          0.77          0.77
27-04-2024 21:26:22 [MainThread  ] [INFO ]             ORGANIZATION          0.76          0.79          0.77
27-04-2024 21:26:22 [MainThread  ] [INFO ]                 FACILITY           0.8          0.73          0.77
27-04-2024 21:26:22 [MainThread  ] [INFO ]                     DATE           0.8          0.81           0.8
27-04-2024 21:26:22 [MainThread  ] [INFO ]                     TIME          0.72          0.79          0.75
27-04-2024 21:26:22 [MainThread  ] [INFO ]                  PRODUCT          0.84          0.79          0.81
27-04-2024 21:26:22 [MainThread  ] [INFO ]                   PERSON           0.8           0.8           0.8
27-04-2024 21:26:22 [MainThread  ] [INFO ]                 LANGUAGE          0.76          0.77          0.76
27-04-2024 21:26:22 [MainThread  ] [INFO ]               PROFESSION          0.89          0.77          0.82
27-04-2024 21:26:22 [MainThread  ] [INFO ]                  ORDINAL          0.77           0.8          0.79
27-04-2024 21:26:22 [MainThread  ] [INFO ]                    EVENT          0.81          0.82          0.81
27-04-2024 21:26:22 [MainThread  ] [INFO ]              WORK_OF_ART          0.89          0.75          0.81
27-04-2024 21:26:22 [MainThread  ] [INFO ]                    AWARD          0.85          0.73          0.78
27-04-2024 21:26:22 [MainThread  ] [INFO ]                  DISEASE          0.76           0.8          0.78
27-04-2024 21:26:22 [MainThread  ] [INFO ]                 IDEOLOGY          0.73          0.79          0.76
27-04-2024 21:26:22 [MainThread  ] [INFO ]                 RELIGION          0.83          0.79          0.81
27-04-2024 21:26:22 [MainThread  ] [INFO ]                  COUNTRY          0.81           0.7          0.75
27-04-2024 21:26:22 [MainThread  ] [INFO ]                      AGE          0.87          0.75           0.8
27-04-2024 21:26:22 [MainThread  ] [INFO ]                  PERCENT          0.72          0.87          0.79
27-04-2024 21:26:22 [MainThread  ] [INFO ]              NATIONALITY          0.75          0.84          0.79
27-04-2024 21:26:22 [MainThread  ] [INFO ]                      LAW          0.77          0.77          0.77
27-04-2024 21:26:22 [MainThread  ] [INFO ]                  PENALTY          0.77          0.74          0.76
27-04-2024 21:26:22 [MainThread  ] [INFO ]                    MONEY           0.8          0.72          0.76
27-04-2024 21:26:22 [MainThread  ] [INFO ]                     CITY          0.74          0.84          0.79
27-04-2024 21:26:22 [MainThread  ] [INFO ]                   FAMILY          0.75          0.86           0.8
```