

## TP\_17/3

### Partie 1 | étude de cas CoNLL 2003 :

1. Quelle type de tâche propose CoNLL 2003 ?
2. Quel type de données y a-t-il dans CoNLL 2003 ?
3. A quel besoin répond CoNLL 2003 ?
4. Quels types de modèles ont été entraînés sur CoNLL 2003 ?
5. Est un corpus monolingue ou multilingue ?

### Réponses :

1. NER (Named Entity Recognition) , une tâche de Reconnaissance d'Entités Nommées
2. Le format IOB
3. CoNLL-2003 répond au besoin de reconnaissance d'entités nommées dans des textes.  
Comme l'analyse de documents, l'extraction des mot-clés, etc.
4. Les modèles statistiques comme HMM, ou les modèles à base de réseaux neuronaux comme BERT et RNN.
5. Il s'agit d'un corpus multilingue. Car il contient des données en anglais et en allemand, donc il est souvent utilisés dans un contexte monolingue.