This internship will involve exploring alternative ways of evaluating terminological aspects of scientific documents and translation. Inspired by the use of question-based metrics to evaluate text generation tasks (Scialom et al., 2021), one possible direction is to explore how terminologies, relation extraction and information extraction can be used as a means of evaluation of terminological quality. For example, (i) can the same relations be found between a reference (human-produced) translation and an automatically produced one? (ii) can terms be matched in similar parts of the document? (iii) how coherent is the use of terms within a document? and (iv) can the same information be extracted from an MT output and the source or reference text? The internship will involve both analysis of existing data, for instance the biomedical translation task (Bawden et al., 2020; Yeganova et al., 2021) and the use and training of neural NLP models.

References


### Principales activités

The main activities will be to carry out research on the topic outlined by (i) studying the past literature, (ii) re-implementing previously proposed approaches and baselines, (iii) proposing improvements to these solutions or a novel approach, (iv) carrying out and writing up experiments, (v) communicating on those experiments with the group.

### Compétences

Candidates should be currently finishing a Master 2 or equivalent (e.g. engineering school) in computer science (speciality artificial intelligence, machine learning or natural language processing).

They should have a good level in programming (python), experience with neural networks and an interest in natural language processing. A good written and spoken level of English is required, and knowledge of French is preferred.

### Avantages

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

### Rémunération

This internship may be either compensated (3,90€/hour) or remunerated (SMIC = 1 678,95€) depending on the candidate's situation.