

Fiche de projet tutoré / Project form

Thé ou café ? Étude de la relation question-réponse.

Encadrement / Supervisors

encadrant principal / main supervisor

Maxime Amblard Sémagramme -- Loria maxime.amblard@univ-lorraine.fr

- autres encadrants / other supervisors

Maria Boritchev Sémagramme -- Loria maria.boritchev@loria.fr

Michel Musiol

michel.Musiol@univ-lorraine.fr

Description / Description

1. projet global/global project

L'étude des interactions dialogiques est un sujet majeur dans le traitement automatique des langues, situé à l'interface entre la sémantique et la pragmatique du langage naturel et ayant plus généralement des liens avec la logique et les sciences de la cognition.

Ce projet s'inscrit dans le cadre plus général d'une étude d'interactions dialogiques conduite par l'équipe SÉMAGRAMME (Loria, Inria Nancy Grand Est). La tâche principale s'articulera autour d'une étude systématique de dialogues provenant de trois corpus : TCOF (Traitement des Corpus Oraux en Français), SLAM (Schizophrénie et Langage : Analyse et Modélisation) et DinG (Dialogues in Games). On s'attachera tout particulièrement aux relations de question-réponse afin de proposer des schémas d'annotation d'interactions dialogiques.

2. biblio. UE 705 (semestre 7)

Après une première phase bibliographique, on commencera par créer un schéma d'annotation de questions en adaptant les travaux de [1], [6] et [10] (dans la continuité duquel s'inscrit ce projet).

3. réalisation. UE 805 (semestre 8)

Afin d'établir un corpus de référence, on annotera manuellement un sous-corpus composé de plusieurs dialogues (nombre à déterminer). Nous travaillerons ensuite sur les témoins lexicaux (tels que la ponctuation) et sémantiques (tels que l'usage de lexiques spécifiques) pour pouvoir déterminer automatiquement les limites de réponses aux questions, en vue d'une automatisation de l'annotation de la relation question-réponse. Enfin, nous envisagerons l'implémentation et la comparaison de plusieurs algorithmes d'apprentissage.

**Informations diverses : matériel nécessaire, contexte de réalisation /
Various information: material, context of realization**

Le corpus TCOF est disponible au téléchargement sur la plateforme Ortholang. Des extraits choisis du corpus SLAM seront fournis aux étudiants. Le corpus DinG est en cours de constitution et sera disponible au téléchargement/fourni aux étudiants, qui pourront participer à sa constitution.

Les étapes suivantes pourront être envisagées :

- familiarisation avec la création et l'annotation de corpus
- sélection de phénomènes intéressants du point de vue de la relation question-réponse
- création d'un schéma d'annotation
- annotation de fragments de corpus présentant ces phénomènes
- comparaison des résultats obtenus après formalisations différentes (DRT, AMR, MRS, Inquisitive Semantics)
- proposition et implémentation d'un algorithme de recherche de paires de questions-réponses
- implémentation de solutions apprentissage machine
- expérimentations et comparaison des résultats

Livrables et échéancier / Deliverable and schedule

- Octobre - Janvier : bibliographie, participation à l'élaboration de DinG, définition du schéma d'annotation
- Février : annotation manuelle des corpus
- Mars : phases de test
- Avril - Mai : analyse des résultats et écriture de rapport

Bibliographie /References (max. 4-5)

[il ne s'agit pas de la bibliographie complète qui sera fournie aux étudiants au début du projet mais d'une bibliographie indicative pour aider à cerner le sujet]

1. Alex Boulton, « Alors la question c'est... ? Questions pragmatiques et annotation pédagogique des corpus », Cahiers de l'Association for French Language Studies, Vol. 13.2., pp. 31-60, 2007
2. Jonathan Ginzburg and Ivan Sag. Interrogative investigations. Stanford: CSLI publications, 2000.
3. María Andrea Cruz Blandón, Gosse Minnema, Aria Nourbakhsh. What's The Answer: Dialogue Annotation, May 2018, Supervised Project Report.
4. Emmanuelle Canut, Virginie André, Bertrand Gaiffe. Mise à disposition de corpus oraux interactifs : le projet TCOF (Traitement des Corpus Oraux en Français). Pratiques : théorie, pratique, pédagogie, 2010, Interactions et Corpus Oraux, pp.147-148