

Séminaire e-Juris

Lyon 23/24

Gildas TAGNY NGOMPÉ, Jacky MONTMAIN, Sébastien HARISPE, Guillaume ZAMBRANO

La masse disponible de décisions de justice exige des solutions automatiques pour aider les acteurs du droit, mais la spécificité des textes est importante et doit être prise en compte lors de l'application de méthodes d'extraction d'information.

La première phase de notre projet porte sur l'extraction d'information des décisions dans l'objectif de construire une base de connaissances jurisprudentielles structurant et organisant les décisions. Une telle base facilite l'analyse descriptive et prédictive de corpus de décisions. Cette présentation exposera une application des modèles probabilistes pour la segmentation des décisions et la reconnaissance d'entités dans leur contenu (lieu, date, participants, règles de loi...).

Plan principal de la présentation

- Introduction au projet CASSANDRA et réflexions sur les méthodes de la justice prédictive [Guillaume - **20min**] cf- mail doc joint au dossier de travail.
- **Positionnement général sur la partie technique [5min]**
 - Du corpus de décision à la base de connaissances (pivot permettant d'envisager l'analyse descriptive/prédictive). Introduction de la notion de base de connaissances et des perspectives qu'elles fournissent en terme d'inférence de connaissances (aspect formel). Présentation graphique au travers du schéma corpus - TAL -> base de connaissances Raisonnement /Modèle d'Inférence / Analyse -> découverte de connaissances / capacité de prédiction. Présentation de la représentation d'une décision considérée.
 - Introduction des aspects techniques discutés.
 -
- **Extraction d'information à partir de décisions [5min]**: exemples portant sur la segmentation et l'extraction d'entités nommées. Présentation de l'intérêt pour le projet global. Présentation des méthodes testées et des résultats présentés à EGC 2017 / discussion.
- **Classification des décisions par utilisation d'une approche semi-supervisée [5min]** (one class classification) - présentation brève des aspects techniques / des

données (discussion sur la tâche d'annotation réalisée par Guillaume) / premiers résultats / discussion.

- **Travaux réalisés sur une tâche de recherche d'information COLLIE 2017 [5min].**
Présentation de la tâche telle qu'elle est définie par le challenge / effectuer le parallèle avec la problématique consistant à identifier des décisions similaires à un cas exprimé en langage naturel ou au travers d'un formulaire interfacé. Présentation de l'approche proposée (en soulignant les membres impliqués / et le fait que l'implication reste tout de même modeste en terme de temps de travail alloué - calendrier délicat pour cette année) / présentation des résultats obtenus (en avant première / souligner que la diffusion des résultats doit rester privée. Discussion associée à cette partie.
- **Récapitulatif des différents points abordés [5min]** : (1) projet global / objectifs, (2) tâches réalisées, (3) perspectives court/moyen terme dans le cadre du projet, (4) enseignements / observations générales suite aux travaux réalisés : problématique complexe qui nécessite afin d'être abordée des compétences techniques diverses (en recherche d'information / représentation des connaissances / extraction d'information / apprentissage automatique - expertise dont dispose en partie le laboratoire mais intérêt évident à l'élargissement et aux collaborations techniques avec des informaticiens mais aussi juristes / annotations).