





Présentation du Laboratoire Commun TALia

Journées Partenaires Entreprises | Jeudi 10 mars 2022

Matthieu Labeau

Vue d'ensemble

- **TALia**, laboratoire commun entre Telecom Paris et **Onepoint**, entreprise française de conseil spécialisée dans la **transformation numérique** des entreprises et organisations (2300 salariés, 200M€ de chiffre d'affaires)



- **Sujet** : le traitement automatique du langage (TAL)
- **Enseignants-Chercheurs impliqués** :
Tiphaine Viard, Thomas Bonald, Jean-Louis Dessalles, Matthieu Labeau

Traitement automatique du langage

- Techniques ayant pour objectif de modéliser le langage, écrit ou parlé : focalisation sur l'**analyse**, la **compréhension** et la **génération de texte**
- Exemples d'application :
 - Chatbots, Moteurs de recherche, Traduction automatique, Analyse des opinions, Résumé automatique...
- Domaine **multi-disciplinaire** en pleine effervescence !
- Techniques au cœur de l'interaction homme-machine

Enjeux

- **Scientifique** : sujets fondamentaux, comme
 - La compréhension des données,
 - L'interprétabilité des modèles.
- **De société** :
 - Applications très concrètes : détection de la haine en ligne, de fake news..
 - Monopole de certains acteurs,
 - Prédominance de l'anglais.
- **D'écologie** : énorme coût environnemental des modèles états-de l'art.

Quelle recherche à TALia ?

- TALia a pour mission fondamentale la **formation** :
 - 5 thèses CIFRE en TAL prévues ; stages, ateliers ;
 - Séminaires scientifiques hebdomadaires.
- Garder un **équilibre** entre exploration des questions **fondamentales** et études **empiriques**
- Ancrer notre recherche dans des applications **concrètes**
- Le TAL est très concurrentiel : on cherchera à **se distinguer** en développant une expertise spécifique, un centre de gravité pour l'équipe

Approches suivies

- **Utilisation de graphes** : exploitation directe des relations entre..
 - Mots et contextes, mots d'une phrase,
 - Plusieurs documents,
 - Parties d'un texte et d'autres modalités.
- **Bases de connaissances** :
 - Exploiter les bases de connaissances,
 - Exploiter le sens commun,
 - Les intégrer dans les modèles de manière efficace.
- **Le raisonnement** :
 - Introduire de la logique,
 - Exploiter le contexte, les contrastes.