



sujet du stage: Traitement de l'Implicite dans le Domaine Médical par la Synthèse d'Information Contextuelle

Description du sujet. Les récentes recherches en TAL, dont notamment celles dans le domaine de la fouille d'argument, ont largement contribué à la systématisation des procédés sous-tendant la structuration du langage. Les modèles de langue pré-entraînés comptent aujourd'hui parmi les réalisations les plus abouties et sont au cœur d'un grand nombre de technologies participant à la compréhension et la manipulation du langage naturel, notamment dans le contexte médical. Cependant, ces modèles ne saisissent pas toujours la totalité des connaissances linguistiques nécessaires pour une compréhension contextuelle fine et des inférences pertinentes. Cela est particulièrement vrai en raison de la nature hautement spécialisée et souvent abstraite des données médicales, qui sont destinées à des professionnels de la santé. En conséquence, ces données contiennent un degré substantiel d'implicite, fondé sur des connaissances présupposées. De plus, cela est renforcé par le format de diffusion (cas cliniques, notes de progression, rapports, etc.), l'auditoire auquel elles sont destinées, ainsi que le contexte énonciatif lui-même, tous ces éléments influant sur le niveau d'abstraction. Dans le but d'optimiser le traitement de ces données par des systèmes d'IA, l'incorporation de l'analyse pragmatique a pour objectif d'adresser les situations implicites en facilitant la déduction des informations induites par le contexte - dans son sens large (connaissances dites circonstancielles, épistémiques, linguistiques). Cependant, il convient de noter que cette dimension d'analyse linguistique représente un domaine à haut risque scientifique, jusqu'à présent peu exploré, en raison de la complexité de sa formalisation et de son opérationnalisation. Dans le cadre de notre recherche et du financement d'un(e) stagiaire, notre objectif consiste à tirer parti des aspects multimodaux de la communication afin de révéler les éléments implicites dans les données médicales. Nous envisageons d'adopter des approches basées sur la génération de texte, notamment celles liées à l'utilisation de techniques de prompting avancées, pour élaborer des "cadres" contextuels. Ces cadres contextuels nous permettront de rendre compte d'éléments tels que la situation d'énonciation, les implicatures implicites, ainsi que d'autres facteurs qui influencent la manière dont les arguments sont formulés et interprétés dans des contextes

médicaux. À travers cette approche novatrice, notre objectif est de développer des techniques de fouille d'argumentation qui non seulement captent les mécanismes fondamentaux sous-tendant le langage médical, mais qui sont également robustes face aux nuances et aux subtilités des communications observées dans ce milieu. Cette approche ouvre la voie à une compréhension plus complète et riche des processus argumentatifs et de persuasion, pouvant aller au-delà du traitement des données médicales.

Mots-clés: traitement automatique des langues, apprentissage machine, fouille d'arguments, médecine numérique

Candidatures. Les candidatures (CV, notes et lettre de motivation) doivent être adressées **avant le 15 janvier 2024** à Anaïs Ollagnier (ollagnier@i3s.unice.fr).

Profil attendu. Étudiant de 5ème année DS4H avec une expérience en Traitement Automatique des Langues et Machine Learning. Un intérêt pour l'analyse de données médicales serait un plus.

Encadrement du stage. Le stage se déroulera au laboratoire I3S de Sophia-Antipolis. Il sera co-encadré par :

Anaïs Ollagnier & Elena Cabrio - Équipe Wimmics

Durée du stage : maximum 6 mois

Financement du stage. Le montant de gratification est celui en vigueur à UniCA, soit approximativement de ~580 €/mois

Références.

Lawrence, J., & Reed, C. (2020). Argument mining: A survey. *Computational Linguistics*, 45(4), 765-818.

Mayer, T., Cabrio, E., Lippi, M., Torroni, P., & Villata, S. (2018, September). Argument Mining on Clinical Trials. In *COMMA* (pp. 137-148).

Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Comput. Surv.* 55, 9, Article 195 (September 2023), 35 pages. <https://doi.org/10.1145/3560815>

Yang, S., Zhang, Y., Cui, L., & Zhang, Y. (2022). Do Prompts Solve NLP Tasks Using Natural Language?. *arXiv preprint arXiv:2203.00902*.