

Extraction de tâches dans les e-mails : une approche fondée sur les rôles sémantiques

Melissa Mekaoui, Guillaume Tisserant, Mathieu Dodard, Cédric Lopez

Emvista

<https://www.emvista.com>

Cap Oméga, Rond-point Benjamin Franklin, 34960 Montpellier

prenom.nom@emvista.com

Résumé. Il est estimé qu'en 2019 environ 1,4 milliard d'e-mails sont envoyés chaque jour en France (293 milliards dans le monde). Les e-mails augmentent sensiblement le volume des communications dans les entreprises. Il en découle une difficulté pour les salariés de consulter la totalité des messages reçus afin d'identifier les tâches à mener. À la fin des années 1990, les premiers systèmes visant à repérer les tâches dans les e-mails ont fait leur apparition. De nombreux travaux ont été réalisés à ce sujet (toujours d'actualité) fondés sur des méthodes d'apprentissage, des méthodes symboliques, et des méthodes hybrides. Deux approches sont communément adoptées : 1) classification d'actes de langages au niveau du message dans son intégralité ou de ses segments textuels, 2) extraction d'informations fondée sur des patrons morphosyntaxiques. Nous proposons et expérimentons une nouvelle approche fondée sur les rôles sémantiques afin de repérer les tâches dans les e-mails et de les structurer. L'évaluation de notre système sur des e-mails professionnels montre la pertinence notre proposition.

1 Introduction

Le substantif "tâche" est défini par le CNRTL¹ ainsi : "Travail défini et limité, imposé par autrui ou par soi-même, à exécuter dans certaines conditions". En contexte industriel, de nombreuses tâches sont décrites textuellement à travers des outils de communication interne tels que Slack² ou Teams³. Ces outils n'ayant évidemment pas pour objectif de véhiculer uniquement les tâches, celles-ci se trouvent noyées dans une multitude d'informations que l'utilisateur doit être en mesure de traiter.

À la fin des années 1990, les premiers systèmes visant à repérer les tâches dans les e-mails ont fait leur apparition. De nombreux travaux ont été réalisés à ce sujet, fondés sur des méthodes d'apprentissage (généralement un SVM), des méthodes symboliques (utilisant des informations linguistiques, généralement au niveau syntaxique), ou des méthodes hybrides (apprentissage + traits linguistiques).

1. <https://www.cnrtl.fr/definition/t%C3%A0che>

2. <https://slack.com/>

3. <https://teams.microsoft.com/>