

Sémantique et contrastes de corpus multilingues

Dans cette communication, on se propose d'illustrer comment la Sémantique Interprétative, assistée par des outils de lexicométrie, peut permettre l'analyse sémantique de corpus volumineux, en particulier pour des objectifs de détection de contenus propres à certains émetteurs.

Je présenterai des résultats obtenus dans deux contextes de recherche :

* le projet C-mantic (en cours) dont un des partenaires, *l'Alliance contre le tabac* souhaite des propositions lui permettant d'innover et d'atteindre mieux sa cible dans ses campagnes de lutte contre le tabac et de prévention du comportement tabagique

* le projet Princip, qui, dans le cadre du Plan d'Action [Safer Internet](#), a développé le seul système de filtrage qui ne se fonde pas sur la technique des mots-clés, pour identifier le racisme, mais sur une analyse linguistique du contenu des pages de la toile.

Pour l'aspect multilinguisme, je proposerai quelques analyses effectuées dans le cadre d'un projet pédagogique, à l'Ertim : les étudiants de Master devaient constituer leur corpus de discours journalistique sur une période du début de la crise financière (dans 5 langues, chinois, japonais, arabe, allemand, anglais) et analyser dans quelle mesure les points de vue exprimés dans les différents journaux variaient avec les lignes éditoriales et les univers culturels.

Evelyne Bourion (CNRS-ERTIM)