

10 ans de JeuxDeMots : un gros réseau lexico-sémantique obtenu par crowdsourcing

Mathieu Lafourcade

LIRMM, Université de Montpellier & CNRS
mathieu.afourcade@lirmm.fr

Résumé : Le projet JeuxDeMots a pour objet de construire un réseau lexical de sens commun (et de spécialité) en français à l'aide de jeux (gwaps - games with a purpose), d'approches contributives mais également de mécanismes d'inférences. Une dizaine de jeux ont été conçus dans le cadre du projet, chacun permettant de collecter des informations spécifiques ou encore de vérifier la qualité de données acquise via un autre jeu. Cet exposé s'attachera à décrire la nature des données que nous avons collectées et construites depuis le lancement du projet durant l'été 2007.

Nous décrirons en particulier les aspects suivant : la structure de réseau lexical obtenu, les types de relations sémantiques représentées (ontologiques, subjectives, rôles sémantiques, associations d'idées), les questions liées à l'activation et l'inhibition de termes et relations, l'annotation de relations (méta-informations), les raffinements sémantiques (gestion de la polysémie), la création d'agglomérations permettant la représentation de connaissances plus riches.

Ce réseau lexical, distribué sous licence libre, est exploité dans de nombreux laboratoires de recherche et entreprises. Les applications en cours utilisant le réseau JeuxDeMots concernent principalement l'interprétation sémantique de textes, la compréhension de l'écrit, la recherche d'information, l'inférence de faits, l'analyse d'opinions et de sentiments - et ce dans des domaines comme la radiologie, le tourisme, la nutrition. Construit à partir d'une liste de 150 000 termes sans aucune relation entre eux, le réseau lexical de JeuxDeMots contient maintenant plus de 1000 000 termes et plus de 80 millions de relations.