

Corriger, analyser et représenter le texte Synapse Développement

Patrick Séguéla et Dominique Laurent

Synapse Développement, 33 rue Maynard, 31000 Toulouse

patrick.seguela@synapse-f.com dlaurent@synapse-fr.com

RÉSUMÉ

Synapse Développement souhaite échanger avec les conférenciers autour des technologies qu'elle commercialise : correction de textes et analyse sémantique.

Plusieurs produits et démonstrateurs seront présentés, notre but étant d'instaurer un dialogue et de confronter notre approche du TAL, à base de méthodes symboliques et statistiques influencées par des contraintes de production, et celles utilisées par les chercheurs, industriels ou passionnés qui viendront à notre rencontre.

ABSTRACT

Checking, analysing and representing texts

Synapse Développement would like to demonstrate its grammar checker and semantic analysis technologies to open exciting discussions with natural language specialists. We are particularly interested in discussing the scientific issues we have to face and solve according to our industrial needs.

MOTS-CLÉS : Correction grammaticale, analyse syntaxique, analyse sémantique, analyse d'opinions.

KEYWORDS : Grammar checker, POS tagging, semantic analysis, opinion mining.

1 Correction de texte

Une première démonstration portera sur le moteur de correction grammaticale du français. Le fonctionnement de ce dernier a déjà été présenté aux conférences TALN 2009 (Laurent et al, 2009).

Le moteur met en œuvre 53 000 règles de grammaire basées sur 1 460 000 informations grammaticales et sémantiques. Au-delà de l'aspect fonctionnel, visible sur le produit *Cordial*, nous montrerons comment sont construites ces règles et les outils que nous mettons en place pour toujours les faire évoluer en fonction des opportunités technologiques nouvelles (réseaux sociaux, sites participatifs avec mémoire d'édition, etc.) (Laurent, 2012), (Beaufort et al, 2010), (Wisniewski et al, 2010).

D'un point de vue plus technique, nous présenterons comment nous arrivons à maintenir une vitesse de traitement supérieure à 10 000 mots/seconde sur une machine du commerce en utilisant toujours plus de ressources linguistiques dans notre moteur.

2 Analyse de texte

Synapse Développement analyse les textes écrits, non structurés et les représente sous forme d'objets. Ces objets sont ensuite organisés pour proposer des applications à haute

valeur ajoutée. L'idée de cette seconde démonstration est de présenter plusieurs applications autour de technologies d'analyse de texte et d'échanger sur la technologie elle-même ainsi que sur les différentes visualisations des résultats issus de ces technologies. Ces démonstrations sont disponibles sur le "lab" de Synapse.

Les technologies présentées seront l'étiquetage morpho-syntaxique (Laurent et al, 2009), l'analyse conceptuelle, l'Extraction d'entités nommées et l'analyse d'opinion (Chardon, 2013).

Les applications seront les suivantes :

1. Anonymisation de textes
2. Création automatique de métadonnées pour l'indexation, la recherche et la mise en relation de contenus proches
3. À partir de commentaires de restaurants : reconnaissance des plats servis, préférés et déconseillés pour un établissement. Évaluation automatique des restaurants sur plusieurs critères reconnus automatiquement : cuisine, ambiance, service, etc.
4. Analyse de la tonalité de commentaires sur une marque à partir des réseaux sociaux et flux RSS.
5. Création de parcours de lecture de textes issus commentaires : orientation (positive/négative), intensité, nature (conseil, jugement, sentiment).

Références

BEAUFORT R. ROEKHAUT S. COUGNON L. et FAIRON C. (2010). Une approche hybride traduction/correction pour la normalisation des SMS. *Actes de TALN 2010 (Traitement automatique des langues naturelles)*, Montréal. ATALA.

CHARDON B. BENAMARA F. POPESCU V MATHIEU Y. et Asher N. (2013). Measuring the Effect of Discourse Structure on Sentiment Analysis. *In Proceedings of CICLING 2013 (Conference on Intelligent Text Processing and Computational Linguistics)*, Samos.

LAURENT D. (2012). Les vraies difficultés du français. ÉDITIONS SYNAPSE DÉVELOPPEMENT.

LAURENT D. NÈGRE S. et SÉGUÉLA P. (2009). L'analyseur syntaxique Cordial dans Passage. *Actes de TALN 2009 (Traitement automatique des langues naturelles)*, Senlis. ATALA, LIPN.

LAURENT D. NÈGRE S. et SÉGUÉLA P. (2009). Apport des cooccurrences à la correction et à l'analyse syntaxique. *Actes de TALN 2009 (Traitement automatique des langues naturelles)*, Senlis. ATALA, LIPN.

WISNIEWSKI G. MAX A. et YVON F. (2010). Recueil et analyse d'un corpus écologique de corrections orthographiques extraits des révisions Wikipedia. *Actes de TALN 2010 (Traitement automatique des langues naturelles)*, Montréal. ATALA.