

Thèmes de recherche

Dernière mise à jour : 01-09-2017

- Extraction terminologique
- Alignement de termes en corpus comparable
- Unités lexicales complexes : termes complexes, variations, collocations
- Fouille d'opinion: analyse d'opinion à granularité fine

Extraction terminologique

L'extraction terminologique consiste à détecter dans un corpus monolingue relevant d'un domaine de spécialité les éléments lexicaux susceptibles d'être des termes, aussi appelés "termes candidats". Quelques questions relatives à l'extraction terminologique :

- le corpus spécialisé : la délimitation du domaine, la taille, la garantie d'une production langagière native, les genres de textes, les intentions communicatives ;
- les termes candidats : terme simple composé d'un seul mot ou termes complexes composés de plusieurs mots ;
- la méthode de repérage des termes candidats dans le corpus ;
- les mesures permettant d'ordonner ces candidats, les méthodes permettant d'éliminer de mauvais candidats, de grouper des candidats
- définir un cadre unifié d'évaluation.

Deux réalisations logicielles pour l'extraction terminologique disponibles dans le domaine public :

- Logiciel Termsuite développé dans le cadre du projet européen TTC (2010-2012) puis dans le projet ISTE services de base (2015-2017). TermSuite extrait termes simples et termes complexes pour les langues suivantes : anglais, français, espagnol, allemand, russe.
- Logiciel ACABIT extrait uniquement les termes complexes. Il a été développé pour le français et l'anglais. Il a par la suite été adapté au japonais, au malgache, à l'espagnol et à l'italien.

Variation terminologique

- Typologie des opérations linguistiques mises en oeuvre dans la production de variantes terminologiques
- Étude des relations sémantiques induites par les variations
- Étude de l'évolution terminologique au travers de ses variantes
- Étude des variations dans un contexte multilingue
- Détection des variantes implémentée dans le logiciel TermSuite

Alignement de termes en corpus comparable

Étant donné deux corpus relevant du même domaine de spécialité dans deux langues, l'alignement de termes consiste à trouver pour un terme dans une langue, ses traductions dans l'autre langue. Quelques questions relatives à l'extraction de terminologie bilingue à partir de corpus comparable :

- le corpus comparable : assurer et mesurer la comparabilité du corpus
- la méthode d'alignement en corpus comparable : dans le cadre du modèle vectoriel, l'obtention des paramètres optimaux pour obtenir les contextes les plus représentatifs d'un terme simple, le meilleur classement de ses traductions candidates
- l'exploitation du modèle vectoriel conjointement avec la compositionnalité pour calculer le meilleur contexte d'un terme complexe
- implémentation des méthodes d'alignement de termes dans le logiciel TermSuite.

Unités lexicales complexes

- Termes complexes
- Spécifications linguistiques des termes complexes
- Segmentation des termes complexes de nature morphologique
- Collocations

Fouille d'opinion

La fouille d'opinion consiste à détecter dans un texte les opinions émises, l'émetteur de l'opinion ainsi que sa cible. La fouille d'opinion à granularité fine se charge aussi de caractériser l'opinion en fonction de son caractère énonciatif, de son orientation positive ou négative, de son degré.

La fouille d'opinion à granularité fine nécessite des ressources diverses : grammaire de l'opinion, lexique de l'opinion dont il s'agit d'automatiser la production.

Dans le cadre du projet ANR Blogoscopie (2006-2008), réalisation du logiciel Apopsis chargé de détecter les opinions exprimées localement dans les blogs quelle que soit la thématique et de les catégoriser en fonction de leur orientation positive ou négative.