Nom et Prénom : KOURTIN ASMAA
Date de soutenance : 23/11/2024
Directeur de Thèse : MOURCHID MOHAMMED

Sujet de thèse :

Modélisation Informatique des Expressions Figées Arabes dans le Traitement Automatique du Langage Naturel basée sur l'Approche Lexique-Grammaire et Applications

Résumé:

Cette recherche explore l'importance des analyses syntaxiques et sémantiques dans le traitement automatique du langage naturel (TALN). L'approche lexique-grammaire, visant à décrire le lexique à travers des tables éditables manuellement, est intégrée à la plateforme NooJ. Cependant, la génération manuelle de dictionnaires à partir de ces tables est chronophage et sujette aux erreurs. Afin de remédier à cela, l'étude propose le développement d'un outil automatisé de génération de dictionnaires, maximisant ainsi l'efficacité de cette approche. L'approche lexique-grammaire, bien que cruciale en linguistique, rencontre des défis d'intégration dans les plateformes de TALN. La normalisation de ces tables est présentée comme une solution, permettant leur compatibilité avec ces plateformes. Cette normalisation est suivie de l'intégration des tables normalisées dans NooJ, en générant leurs dictionnaires et en créant leurs grammaires syntaxiques. Ces derniers capturent les informations lexicales, syntaxiques et sémantiques pour évaluer l'efficacité des dictionnaires. La recherche explore également les expressions figées en arabe, en créant leurs tables lexique-grammaires et en les implémentant dans NooJ. Ces expressions sont analysées en vue de leur reconnaissance et annotation automatiques, même en présence de discontinuité. Cette approche offre des solutions potentielles pour des défis dans le TALN, y compris la traduction automatique. Enfin, la complexité de la traduction des expressions figées entre l'arabe, le français et l'anglais est abordée. En améliorant les tables existantes avec des traductions et en les convertissant en dictionnaires, un traducteur automatique est développé et évalué à travers des tests approfondis sur divers textes et corpus. Cette approche offre une contribution significative à la traduction automatique des expressionsfigées

Abstract:

This research explores the importance of syntactic and semantic analyses in automatic natural language processing (NLP). The lexicon-grammar approach, aimed at describing the lexicon through manually editable tables, is integrated into the NooJ platform. However, the manual generation of dictionaries from these tables is time-consuming and error-prone. To resolve this, the study proposes the development of an automated dictionary generation tool, maximizing the efficiency of this approach. Lexicon-grammar approach, while crucial in linguistics, faces integration challenges on NLP platforms. These tables' standardization is presented as a solution, enabling their compatibility with these platforms. This normalization is followed by the integration of the normalized tables into NooJ by generating their dictionaries and creating their syntactic grammar. These capture lexical, syntactic, and semantic information to assess the effectiveness of the dictionaries. The research also explores frozen expressions in Arabic by creating their lexicon-grammar tables and implementing them in NooJ. These expressions are analyzed for automatic recognition and annotation, even in the presence of discontinuities. This approach offers potential solutions for challenges in NLP, including machine translation. Finally, the complexity of translating frozen expressions between Arabic, French, and English is addressed. By enhancing existing tables with translations and converting them into dictionaries, an automatic translator is developed and evaluated through extensive testing on various texts and corpora. This approach offers a significant contribution to translation