République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université des Sciences et de la Technologie Houari Boumediene

Faculté d'Informatique

Département d'Intelligence Artificielle et Sciences des Données

Mémoire de Master

Spécialité: Master Systèmes Informatiques Intelligents

Thème

Deep learning pour la recherche des experts scientifiques

Sujet Proposé par :
M. BELLAZZOUGUI Djamal
M. CHAA Messaoud

Présenté par :
HOUATI Chakib Mouloud
MEZIANI Serine

Soutenu le : 20 /09/2022

Devant le jury composé de:

M. HAMMAL Youcef	Président
M. KHENNAK Ilyes	Membre

N° Projet : SII_E_012 / 2022

Résumé

La recherche d'experts scientifiques (*scientific expert finding*) est une tâche récurrente dans le milieu académique. En effet, s'exprime souvent le besoin de trouver des encadreurs, membres de jury pour des projets de fin d'études ou de thèses, évaluateurs de propositions de projets de recherches, référés pour des articles soumis à des revues scientifiques ou des membres de comités de programmes de conférences. Cette recherche est rendue très difficile par l'augmentation du nombre de chercheurs et des travaux scientifiques ainsi que par la grande diversification et spécialisation des domaines de recherches scientifiques.

L'objectif de ce travail sera, de concevoir un système de recherche d'experts par des requêtes exprimées en langage naturel. Ce système sera basé sur les données des articles publiés par les chercheurs (titre, résumé, mots-clés). Le système proposé intégrera les techniques les plus récentes de traitement du langage naturel (NLP) basées sur le deep-learning, en particulier la technique d'apprentissage automatique à base de *transformer BERT* (*SciBERT* et *RoBERTa*), développée par *Google*, et publiée en 2018. Nous présenterons trois contributions principales : une nouvelle approche de l'indexation, l'expansion des requêtes à l'aide de la définition, et une méthode de distribution des scores. En plus de cela, la construction d'un nouveau corpus de test basé sur l'ACM.

"Une plateforme nationale de recherche d'experts académiques a été créée à l'issue de ce projet. Ce portail intègre les méthodes que nous avons proposées, et les valide sur le cas de la production scientifique algérienne.

Mots clés :
Scientific Expert Finding ; Deep Learning ; Word Embeddings ; Bibliographic Databases ; Voting Models ; Query Expansion ; BERT ; SciBERT ; RoBERTa ; FAISS.

Abstract

Scientific expert finding is a recurrent task in the academic world. Indeed, it is often necessary to look for supervisors, jury members for end-of-study or thesis projects, evaluators for research project proposals, referees for articles submitted to scientific journals or members of conference program committees. This research is made very difficult by the increase in the number of researchers and scientific work as well as the great diversification and specialization of the scientific research fields.

The objective of this work will be to design an expert finding system that enhanced through the expansion of queries expressed in natural language. This system will be based on the data of the articles published by the researchers (title, abstract, keywords). The proposed system will integrate the most recent techniques in natural language processing (NLP) based on deep-learning, in particular the transformer-based machine learning technique *BERT* (*SciBERT* and *RoBERTa*), developed by *Google*, and published in 2018. We will present three main contributions : a new indexation, expansion of queries by using the definition, and a score distribution method. In addition to this, the construction of a new test corpus based on ACM.

A national platform for expert research was created at the end of this project. This portal integrates the methods that we proposed, and validates them on the case of the Algerian scientific production.

Key words :
Scientific Expert Finding ; Deep Learning ; Word Embeddings ; Bibliographic Databases ; Voting Models ; Query Expansion ; BERT ; SciBERT ; RoBERTa ; FAISS.

Remerciements

It is with genuine gratitude and warm regard that I dedicate this thesis to my loving parents, ***Abdelaziz*** *and* ***Salima****, who have been always by my side throughout my life and especially in my educational journey. To my beloved siblings,* ***Tahar*** *and little* ***Nonat****.*

I would also like to acknowledge my appreciation and gratefulness to my supervisors ***Mr. BELLAZZOUGUI Djamal*** *and* ***Mr. CHAA Messaoud*** *for their valuable advises and orientations that allowed us to carry out our work successfully.*

I sincerely thank the members of the jury, ***Mr. HAMMAL Youcef*** *and* ***Mr. KHENNAK*** *Ilyes for the honor they give us by accepting to judge this work.*

Finally, a special thanks to my aunt ***Souhila****, my partner* ***Chakib****, and to all the people who have reviewed this thesis.*

Remerciements

J'adresse mes sincères remerciements à mes deux encadreurs, ***Mr. BELLAZZOUGUI Djamal*** *et* ***Mr. CHAA Messaoud*** *pour toutes leurs remarques précieuses qui ont guidé notre travail tout au long de notre stage au CERIST.*

Je remercie également le juré composé de ***Mr. HAMMAL Youcef*** *et* ***Mr. KHENNAK Ilyes*** *pour nous avoir honorer de juger notre projet.*

Je tiens à exprimer toute ma reconnaissance à mes deux chers parents, ma petite soeur ***Amel****, et mes grand-parents pour leur soutien inconditionnel durant tout mon parcours académique.*

J'exprime toute ma gratitude à ceux qui ont revu notre mémoire et corrigé les fautes, incluant ***Mme. Souhila****,* ***Mlle Ikram*** *ainsi que mes amis* ***Nesrine*** *et* ***Djalil****.*

Je remercie spécialement ***Dr. BOUCENNA Fateh****, qui fut le responsable de notre découverte du sujet de notre PFE. Ainsi que mon oncle* ***Moulou****d qui a grandement contribué au bon déroulement du projet*

Je voudrais Pour finir, exprimer mes vifs remerciements envers mes amis ***Hocine****,* ***Ines****,* ***Chakib****,* ***Zo****,* ***Yacine****,* ***Lint****, sans oublier mon cousin* ***Raouf****, et ma binôme* ***Serine*** *qui m'ont apporté leur soutien moral et intellectuel.*

À ma très chère regrettée grand-mère

Table des matières