



AVIS DE SOUTENANCE THESE DE DOCTORAT

Présentée par

Mr : MOHAMMED RAIS

Spécialité : Informatique

Sujet de la thèse : Etude et amélioration de la représentation du texte biomédical basée sur l'intégration des ontologies du domaine et l'enrichissement sémantique.

Formation Doctorale : Sciences de l'ingénieur Sciences Physiques, Mathématiques et Informatique.

Thèse présentée et soutenue le samedi 09 novembre 2019 à 10h au centre des conférences devant le jury composé de :

Nom Prénom	Titre	Etablissement	
Oussama CHERKAOUI	PES	Faculté des Sciences Dhar El Mehraz de Fès	Président
Zineb SERHIER	PES	Faculté de Médecine et de Pharmacie de Casablanca	Rapporteur
Abderrahim ELQADI	PES	Ecole Supérieure de Technologie de Salé	Rapporteur
Aicha MAJDA	PH	Faculté des Sciences et Techniques de Fès	Rapporteur
Noureddine CHENFOUR	PES	Faculté des Sciences Dhar El Mehraz de Fès	Examineur
Abdelmonaime LACHKAR	PES	Ecole Nationale des Sciences Appliquées de Tanger	Directeur de thèse

Laboratoire d'accueil : Laboratoire ISA.

Etablissement : Ecole Nationale des Sciences Appliquées de Fès.



Titre de la thèse : Etude et amélioration de la représentation du texte biomédical basée sur l'intégration des ontologies du domaine et l'enrichissement sémantique.

Nom du candidat : Mohammed RAIS

Spécialité : Informatique

Résumé de la thèse

Face à l'immense masse d'informations biomédicale qui ne cesse d'argumenter exponentiellement jours après jours, l'accès à l'information dans le domaine biomédicale devient de plus en plus ardu. Il est donc très intéressant de faire face à cette problématique, de proposer des méthodes, et de mettre en place des outils permettant ainsi d'assister les utilisateurs de l'information dans le domaine biomédical d'y accéder très rapidement et facilement pour satisfaire le mieux possible leurs besoins d'une manière très pertinente. Afin d'améliorer la performance des résultats de n'importe applications de Fouille de Texte Biomédicale, dans le cadre de cette thèse, nous nous sommes intéressés principalement au problème de la représentation du texte biomédical, ainsi nous traitons les défis de l'indexation des documents biomédicaux. Par ailleurs, nous nous sommes intéressés aussi à un troisième problème lié à la Navigation dans les résultats de recherche retournés par *PubMed*. La manière habituelle de représenter un texte est un paquet de mots BoW. Cette représentation souffre du manque de sens dans la représentation qui en résulte et ignore toute la sémantique qui réside dans le texte original, l'intégration de la sémantique est résolue par une représentation dite *Sac-de-Concepts BoC*. Néanmoins, en appliquant cette représentation conceptuelle, l'attribution d'un terme à ses concepts dans une ontologie peut générer de l'ambiguïté dans la représentation finale du document.

Pour faire face à la fois au problème de la Représentation du Texte et à celui de l'Indexation des Documents Biomédicaux, dans le cadre de cette thèse, nous avons proposé d'utiliser des ressources termino-ontologiques du domaine biomédical. En effet, nous avons d'abord présenté et discuter les différentes stratégies de conceptualisation en relation avec la résolution de l'ambiguïté. Ainsi, nous avons présenté un cadre de travail générique pour la conceptualisation de texte biomédical. Ensuite nous avons proposé deux nouvelles méthodes pour la résolution de l'ambiguïté des termes biomédicaux -WSD - basée sur le contexte du mot ambigu à travers l'utilisation les mesures de similarité entre les concepts de l'ontologie utilisée. De même au niveau de l'indexation des documents biomédicaux, nous avons proposé également d'utiliser une nouvelle méthode basée sur la représentation conceptuelle. Cette dernière consiste à intégrer un module pour la désambiguïté lexicale des termes ambigus dans l'architecture du système d'indexation et recherche d'information. En effet, dans ce module, nous avons exploité encore plus les solutions que nous avons déjà proposées pour surmonter le problème de la désambiguïté lexicale pour améliorer le processus de l'indexation et par suite augmenter la performance des systèmes de recherche d'information dédiés au domaine biomédical.

Par ailleurs, le problème de Navigation pose aussi un autre défi majeur pour l'accès à l'information dans le domaine biomédical. En effet, dans la phase de la Navigation offerte par le moteur *PubMed* ainsi que les autres moteurs de recherche tel que Google, Yahoo, Bing propose de retourner une liste ordonnée d'une dizaine de milliers de snippets (métas-données), les utilisateurs ne consultent que les premières pages, et par conséquent les documents situés à la fin de la liste ne sont jamais consultés bien qu'ils puissent être pertinents. Pour remédier à ce problème de navigation, nous avons proposé une autre contribution basée sur l'utilisation de l'Analyse Formelle de Concept (Formal Concept Analysis) FCA, cette dernière permet d'une part de produire un regroupement thématique des résultats de recherches, et d'autre part de présenter une interface de navigation hiérarchique sur deux niveaux permettant ainsi une navigation rapide, simple et efficace dans une structure arborescente des groupes thématiques ainsi générés.

Pour mettre en évidence l'intérêt de toutes les contributions que nous avons proposé dans le cadre de ce travail de thèse, nous avons bien mené plusieurs études comparatives. Les résultats obtenus sont très encourageants que ce soit pour la désambiguïté lexicale ou pour le cas de l'Indexation et la navigation.

Mots Clés

Fouille de Texte Biomédicale, Traitement du Langage Naturel, Sac de Concept, Résolution de l'Ambiguïté Lexicale, Représentation Conceptuelle, Enrichissement Sémantique, Indexation/Navigation.