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'Electronique et d'Informatique
Département Informatique

Mémoire de Master

Option
Ingénierie de logiciel

Thème

Extraction d'événements à partir de textes écrits en langue arabe

---

Encadreur :

Dr. ALIANE H., CERIST

Présenté par :

GUENDOUZI Wassila.
MOKRANI Amina.

Devant le jury composé de :

Pr.	A. Guessoum	USTHB	Président
Mme.	M. Mameri	USTHB	Membre
Mme	Benchouk	USTHB	Membre

Binôme N°/0542012

Remerciements

*Tout d'abord, nous remercions **DIEU** le tout puissant de nous avoir aidées pour accomplir ce modeste travail.*

Nous tenons à remercier les membres de jury : Pr. Guessoum, Mme Mameri et Mme Benchouk de nous avoir fait l'honneur de juger ce travail.

*Un Grand Merci et une reconnaissance sincère à notre promotrice **Mme ALIANE** pour nous avoir fait découvrir le domaine passionnant du TAL ainsi que pour sa disponibilité, sa bienveillances et son aide fructueuse qui nous a permis de mener à bien ce travail.*

Une reconnaissance particulière à toutes les personnes qui ont contribué de loin ou de prêt à l'aboutissement de ce travail, pour leurs encouragements, suggestions et remarques ; en gardant toujours un grand sourire.

Enfin, nous ne pouvons qu'être infiniment reconnaissantes envers nos parents pour leur soutien indescriptible, leur patience, leur confiance et leurs nombreux sacrifices. Nous leur dédions avec plaisir ce travail Qu'ils sachent que nous sommes conscientes de ce que nous leurs devons et qu'ils trouvent ici l'immense témoignage de notre affection

Résumé

Avec l'augmentation sans cesse grandissante des volumes de linformation textuelle sur le web, le problème de l'extraction de connaissances à partir de données textuelles devient crucial et incontournable. Le processus d'extraction de connaissances inclut généralement plusieurs étapes, dont la reconnaissance d'entités et d'événements dans lesquelles elles sont impliquées, ainsi que les informations temporelles, spatiales et autres qui peuvent leur être rattachées. Néanmoins, bien que le problème de lidentification des entités nommées soit relativement bien compris, lidentification d'événements et de leurs relations temporelles demeure un problème qui est encore peu abordé. Or, que ce soit pour les applications de réponses à des questions, ou la simple recherche d'information, on est souvent appelé à rechercher ou annoter dans les documents les éléments qui nous permettent de répondre à des questions de la forme suivante : Quand un certain événement a-t-il eu lieu ? Que s'est-il passé pendant une certaine période de temps donnée ? Dans quels événements une personne a-t-elle été impliquée ?

Les évènements peuvent être désignés par des verbes mais on peut aussi avoir des évènements nominaux désignés comme des déverbaux. Une étape du travail dextraction des évènements est de déterminer les expressions désignant potentiellement un événement, puis de déterminer en contexte si c'est bien le cas.

Nous proposons dans ce travail une approche pour l'extraction des entités linguistiques désignant potentiellement des événements ainsi que les expressions pouvant y afférer à savoir les expressions temporelles et les expressions de lieu. Nous n'utilisons pas de corpus annoté ni de lexiques mais notre approche est fondée sur l'élaboration d'un ensemble de règles qui utilisent des indicateurs formels issus de lobservation du corpus de travail. Nous travaillons sur un corpus de dépêches de presse écrits en Arabe Standard Moderne, et les résultats que nous obtenons sont comparables aux résultats des travaux que nous avons étudiés pour les autres langues.

Mots clés :

TALN, langue arabe, segmentation des mots arabes, annotation en partie de discours, extraction de connaissance, fouille de texte, extraction d'événement, expressions temporelles expressions de lieu..

Liste des tableaux

Liste des figures

Table de matières