Partie IV : Formalisation et opérationnalisation

1. Introduction

Les méthodes formelles jouent un rôle crucial dans le développement des technologies du Web sémantique et ont pour but d'assurer leur fiabilité et leur sécurité. Les techniques de modélisation et de vérification peuvent être utiles dans les différents niveaux de la conception et du déploiement des ontologies (**Chaâbani & al, 2009**). Parmi ces méthodes les logiques de description en sont un outil très puissant par la variété de ses différents langages.

Dans une ontologie formalisée, nous pouvons vérifier la consistance, et calculer la hiérarchie des classes, elle permet aussi de compléter et valider le modèle construit.

A notre connaissance, c'est la première fois qu'une ontologie en arabe est formalisée avec la logique de description pour que, une fois terminée elle puisse être intégrée dans n'importe quelle application et qu'on puisse raisonner dessus. La formalisation peut bien être faite avec les frames[53], les graphes conceptuels[54], le formalisme-Z[55], LIFE[56]ou une logique de description[57].

2. Formalisation des concepts

2.1. Les logiques de description

Les logiques de descriptions sont une famille de formalisme pour la représentation des connaissances dans différents domaines notamment dans les ontologies. Dans une base de connaissance en logique descriptive, on distingue la TBox (niveau Terminologique) et la ABox (niveau Assertionnel). La première contient tous les axiomes définissant les concepts du domaine, comme la définition de « **رسول** » qui est un « **نبي**» qui a en plus « *a un message de Dieu à transmettre à un peuple*» par exemple. La ABox contient les assertions sur les individus en spécifiant leurs classes et leurs attributs. C'est dans la ABox qu'on trouvera que « **عيسى** » est un « **نبي** » qui en plus a un message à un peuple « **بنو_اسرائيل** ». Dans la TBox on est

[53] http://www.learningwebdesign.com/pdf/frames.pdf
[54] http://www.jfsowa.com/cg/cg_hbook.pdf
[55] http://www.ppig.org/papers/14th-triffitt.pdf
[56] http://www.cs.uiowa.edu/~fleck/lifeIntro.pdf
[57] http://www.cs.man.ac.uk/~horrocks/Publications

intéressé à savoir si tous les concepts sont consistants, par exemple que si deux classes « **مؤمن** » et « **كافر** » sont disjointes, on ne doit trouver une sous-classe communes aux deux. C'est aussi dans la TBox qu'on exprime la relation de subsomption, par exemple si on a que « **نبي**» *est_un* « **إنسان** » et qu'on a que « **رسول** » *est_un* « **نبي** » alors on déduit automatiquement que « **رسول** » est un « **إنسان** ». Toutes fois il existe pour notre ontologie des inconsistances avec lesquelles on doit travailler comme tout « **إنسان** » descend de « **ذكر** » et « **أنثى** » et que « **عيسى** » est un « **إنسان** » mais ne descend pas de « **ذكر** » et « **أنثى** ». Outre cela, nous avons été appelés à compléter la hiérarchie avec des termes générique quand cela s'est avéré nécessaire.

2.2. AL : La base des logiques de description

Les logiques de description varient, de La base qui est le langage **AL (**attributive language**),** jusqu'à celles avec une complexité exponentielle comme c'est le cas pour SHIF or SHIQ (**Papini, 2002**).

Le degré d'expressivité du langage AL est limité, mais il peut convenir à une utilisation qui ne nécessite pas un haut degré d'expressivité. Les descriptions possibles dans le langage AL sont les suivantes (notons que les concepts ou rôles atomiques ou primitifs, constituent les entités élémentaires d'une TBox tels que « **نبي** » et « **رسول** », alors que les concepts et les rôles composés ou définis sont ceux combinés au moyen de constructeurs tels que « **نبي** ∩**رسول** » (**Napoli, 1997**)).

2.2.1. Syntaxe du langage AL

En supposant que A est un concept atomique et que C et D peuvent être atomiques ou complexes (**Gagnon, 2004**) nous avons:

A	Concept atomique
T	Concept universel
⊥	Concept impossible
¬ A	Négation atomique

$C \cap D$	Intersection de concepts
$\forall R.C$	Restriction de valeur
$\exists R.\ T$	Quantification existentielle limitée

Tableau 15: La syntaxe du langage AL

Le concept le plus générique est la racine désignée par T. $\perp$ est le concept le plus spécifique, le constructeur *et* $\cap$ définit une conjonction et le symbole $\neg$ exprime une négation. Le quantificateur universel *tous* ($\forall r.C$) donne le co-domaine du rôle r, alors que le quantificateur existentiel *quelque* ($\exists r$) exprime le fait qu'il y a au moins un couple d'individus reliés par la relation (ou rôle) r.

2.2.2. Sémantique du langage AL

La sémantique du langage AL fait appel à la théorie des ensembles. A chaque concept est associé un ensemble d'individus. Une interprétation suppose l'existence d'un ensemble non vide Δ qui représente des entités du monde décrit. Soit une fonction d'interprétation $\mathcal{I}$, qui associe à chaque description un sous-ensemble de Δ. On suppose que pour chaque concept atomique A, la fonction $\mathcal{I}(A)$ associe un sous-ensemble $A^{\mathcal{I}} \subseteq \Delta$, et pour chaque relation atomique R, une relation binaire $R^{\mathcal{I}} \subseteq \Delta x \Delta$. La fonction d'interprétation est définie ainsi :

$\mathcal{I}(T) = \Delta$
$\mathcal{I}(\perp) = \Phi$
$\mathcal{I}(\neg A) = \Delta \setminus A^{\mathcal{I}}$
$\mathcal{I}(C \cap D) = \mathcal{I}(C) \cap \mathcal{I}(D)$
$\mathcal{I}(\forall R.C) = \{a \in \Delta \mid \forall b.\ (a, b) \in \mathcal{I}(R) \rightarrow b \in \mathcal{I}(C)\}$
$\mathcal{I}(\exists R.\ T) = \{a \in \Delta \mid \exists b.\ (a, b) \in \mathcal{I}(R)\}$

Tableau 16: Sémantique du langage AL

De plus on définit le concept d'axiome terminologique qui est en fait toute formule de la forme suivante :

C⊆D ou **C≡D**

La première forme déclare que toute entité de la classe C appartient aussi à la classe D, alors que la seconde indique que les concepts C et D sont équivalents, c'est-à-dire que si un individu b appartient à la classe C, il appartient nécessairement à la classe D et vice versa.

Leur sémantique est :

$\mathcal{I}(C \subseteq D) = \text{vrai}$ si $\mathcal{I}(C) \subseteq \mathcal{I}(D)$

$\mathcal{I}(C \equiv D) = \text{vrai}$ si $\mathcal{I}(C) = \mathcal{I}(D)$

Une définition est un axiome de la forme C ≡D où C est un concept atomique. Elle sert à associer un nom à un concept complexe.

Nous pouvons bien sûr définir d'autres constructeurs pour obtenir d'autres langages tells que:

ALU = AL ∪ (C ∪ D): disjonction (Union).

ALC =AL∪ {¬c} c est un concept défini ou combiné. ALC (AL avec le Complément) est le langage le plus important.

2.2.3. Les deux niveaux de description

Dans ce qui suit nous allons présenter la formalisation des concepts et relations Avec le langage AL. Comme nous l'avons mentionné plus haut, en logique des descriptions il existe deux niveaux: le terminologique et l'assertionnel.

2.2.3.1. Le niveau terminologique ou TBox

La TBox décrit la connaissance générale d'un domaine particulier, elle inclut les définitions des concepts et des rôles et contient le modèle du monde en termes de concepts leurs propriétés et les relations entre les concepts.

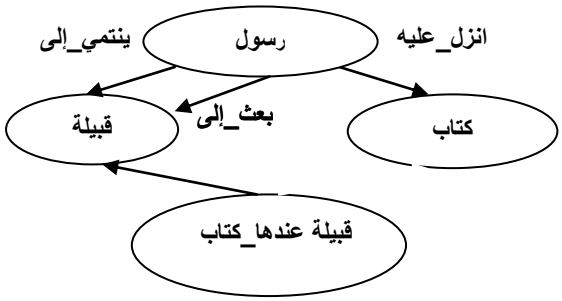

Figure 40: Représentation des concepts et relations (Zaidi & al., 2012b)

Nous pouvons exprimer ces concepts et ces relations ainsi:

رسول $\subseteq \forall$ **ينتمي-إلى . قبيلة** (un messager appartient seulement à une tribu)

قبيلة $\subseteq$ **قبيلة _عندها_ كتاب** (une tribu_avec_un_livre est une tribu)

نبي $\subseteq \exists$ **بعث_إلى . قبيلة** (un prophete est envoyé à au moins une tribu)

Nous avons notés le type de relations suivantes:

- L'identité: (كتاب، كتاب)
- La synonymie : (العرش، الكرسي), (جبريل ،روح_القدس), (إسرائيل، يعقوب), (مكة، بكه)
- La classification (*الرسول، النبي*)
- L'antonymie : (ذكر،انثى), (مؤمن، كافر)
- L'équivalence : (نعمة ، فضل _ من _ الله)

Et quelques propriétés comme la réflexivité dans :

- **(قوم، قوم) لا_يسخر_من**
- **(مؤمن، مؤمن) أخ**

2.2.3.2. Le niveau assertionnel (factuel) ou ABox

Il décrit les individus en les nommant et en spécifiant les assertions, en termes de concepts et de rôles. Plusieurs ABox peuvent être associées à une même TBox. Chacune d'elles montre une configuration constituée par des individus et utilise les concepts et les rôles de la TBox pour l'exprimer.

Considérons l'exemple suivant, tiré de notre corpus:

Concept1	**Relation**	**Concept2**
رسول	أنزل_عليه	كتاب

Avec « **عيسى**» comme instance du concept « **رسول** » et « **الإنجيل** » comme instance du concept « **كتاب** » et la relation « **انزل_عليه** ». La TBOx et la ABOx correspondantes sont:

T-BOX
رسول ≐ **انسان** ∩ ∃**اسمه**. string
∩ ∃**قبيلة** . string
∩ ∃**كتاب .انزل عليه**
⊆ ∃ **قبيلة .بعث_إلى**
ABOX
عيسى :رسول
الإنجيل: كتاب
(بنو_اسرائيل ،عيسى) بعث_إلى
(الإنجيل، عيسى) انزل_عليه

Figure 41: Exemple de TBox et de ABox associée

Soit l'interprétation suivante:

Δ = {محمد، عيسى، موسى، إبراهيم ،داوود، شعيب، صالح، القرآن، الإنجيل، التوراة، الزابور، الصحف،**}**بنو_إسرائيل، مدين، قريش، ثمود،العراق،

نبي^{I} **=**{ محمد، عيسى، موسى، إبراهيم ،داوود، شعيب، صالح**}**

رسول^{I} **=**{ محمد، عيسى، موسى، إبراهيم ،داوود**}**

قبيلة^{I} **=**{ بنو_إسرائيل، مدين، قريش، ثمود،العراق**}**

انزل_عليه^{I} **=**{**(**إبراهيم، الصحف**),(**داوود، الزابور**),(**موسى،التوراة**),(**عيسى، الإنجيل**),(**محمد،القرآن**)}**

ينتمي_إلى^{I} **=**{**(**بنو_إسرائيل ،داوود**) ,(** بنو_إسرائيل ،موسى**),(**بنو_إسرائيل ،عيسى**) ,(** قريش ،محمد**)**
}(ثمود ،صالح) **(** العراق ،إبراهيم**) ,(**مدين ،شعيب

Alors :

(نبي¬∩¬رسول) ^{I}**=**{ شعيب، صالح **}**

¬نبي^{I} **=** {القرآن، الإنجيل، التوراة، الزابور، الصحف، بنو_إسرائيل، مدين، قريش، ثمود،العراق**}**

(∃بعث_إلى. **(**قبيلة**) =**^{I}**{(**قريش ،محمد**)}**

(∀إلى_بعث. **(**قبيلة**)=**^{I}**{ (**عيسى ،بنو_إسرائيل**),(**موسى ،بنو_إسرائيل **) ,(**داوود ،بنو_إسرائيل**),(** مدين **)**

(شعيب،**), (**إبراهيم ،العراق **) (**صالح ،ثمود**) }**

Pour formaliser les concepts, un éditeur a été spécialement développé parce que tous les éditeurs existants ne pouvaient supporter les caractères arabes avec les symboles mathématiques requis par la logique de descriptions.

Le nouvel éditeur a une syntaxe spécifique qui respecte la grammaire de la LD. Il contient les deux parties Terminologique (TBox) et factuelle (ABox) et il permet d'importer un fichier LD pour le réutiliser (**Zitouni, 2010**).

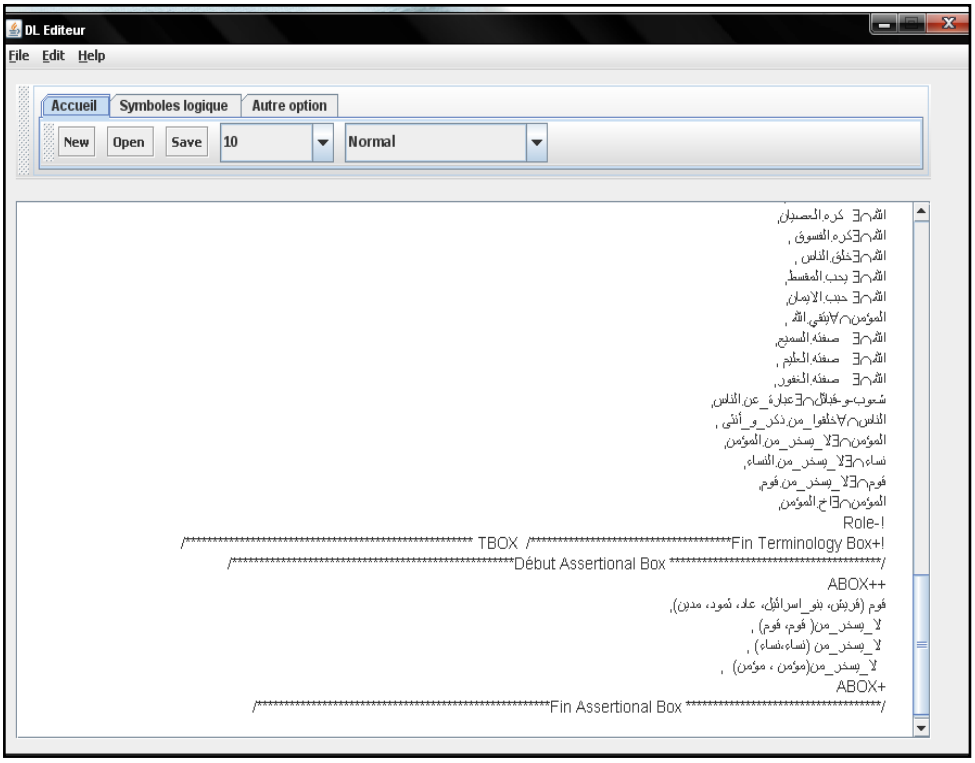

Figure 42: Editeur pour la manipulation des caractères arabes et des symboles mathématiques

3. Opérationnalisation de l'ontologie

Pour la phase d'opérationnalisation et afin de tester notre fichier LD obtenu, nous avons essayé de l'utiliser dans une application (**Djabourabi, 2011**), pour montrer que notre ontologie, une fois complétée et formalisée, peut être intégrée dans n'importe quel système dans le but d'exploiter le fichier LD. Pour cela une application simple a été créée dans le but d'opérationnaliser l'ontologie cette application prend en entrée un fichier LD qui doit représenter l'ontologie

formalisée et produit en sortie la visualisation de la hiérarchie des concepts, les relations, les propriétés et les instances et qui permet de rechercher un concept dans l'arbre conceptuel ou mettre à jour la structure ontologique par l'ajout de nouveaux concepts, la modification ou la suppression et ce soit directement de l'interface de l'application ou en mettant à jour le fichier LD à l'aide de l'éditeur conçu à cette fin.

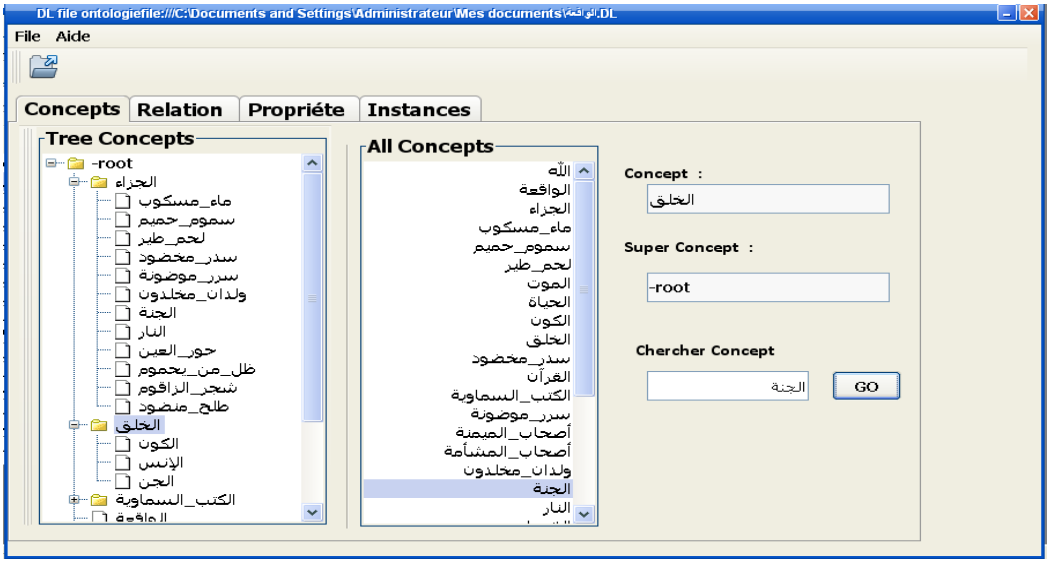

Figure 43: Interface utilisateur du navigateur de l'ontologie

Cette application n'est qu'un essai, l'ontologie peut être intégrée dans d'autres systèmes comme l'expansion d'une requête en rajoutant des hyperonymes afin

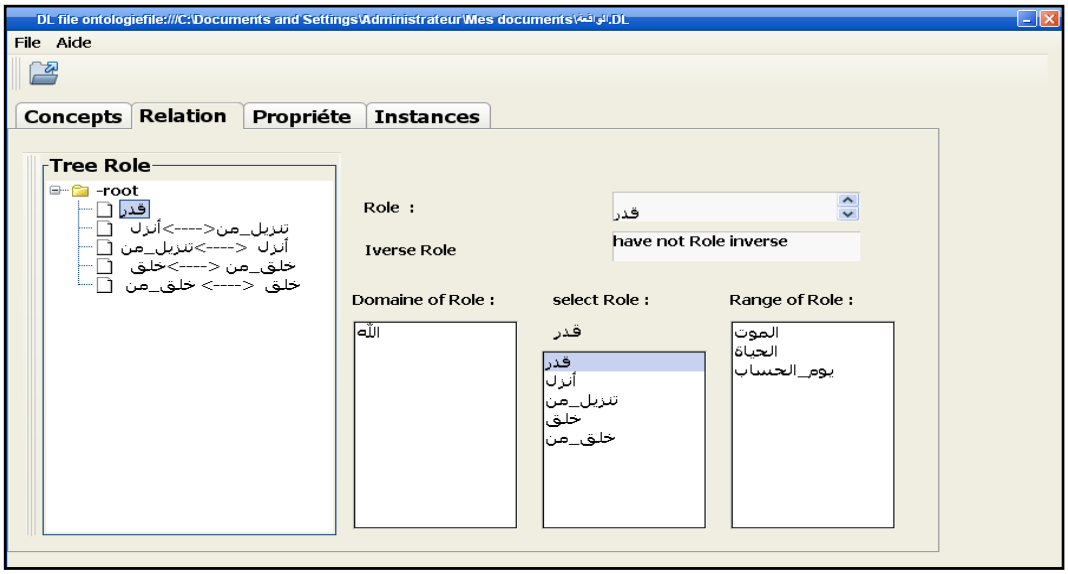

Figure 44: Visualisation des relations

d'améliorer le rappel ou des hyponymes pour améliorer la précision. Elle serait également d'une aide considérable pour des applications de brainstorming dans le sens où elle fournit à l'utilisateur des termes en relation avec un éventuel projet.

4. Conclusion

Nous avons présenté dans ce chapitre l'essentiel de notre travail qui se résume en quatre parties, la première nous l'avons dédiée à l'extraction des termes simples à l'aide d'une approche statistique basée sur tf-idf, notons qu'on aurait pu utiliser GATE pour retrouver des noms en tant que termes et nous pouvons filtrer les résultats avec tf-idf. La seconde partie concerne la recherche des collocations avec une méthode hybride d'abord linguistique puis les résultats obtenus sont filtrés avec une méthode statistique utilisant l'information mutuelle comme mesure de cohésion entre les termes. La troisième traite de l'extraction de relations sémantiques entre termes simples ou collocations préalablement retrouvés, elle repose aussi sur la succession des deux mêmes approches mentionnées. Dans la dernière partie nous avons présenté la formalisation des concepts avec l'un des formalismes de représentation de connaissances les plus puissants, qui est la logique de description. Enfin nous avons contribué à la conception d'une application pour expérimenter l'utilisation du fichier LD créé avec un éditeur développé à cette fin, puisque les éditeurs existants ne permettaient pas la manipulation des caractères arabes avec les symboles mathématiques utilisés en logique de description. L'objectif étant de rendre l'ontologie construite opérationnelle, cette application permet la recherche, la visualisation et la mise à jour des concepts.

5. Conclusion et Perspectives

Les travaux dans le domaine d'extraction de connaissances, dans leur majorité, ont traité un seul type de termes soit simples soit composés, nous avons essayé d'en extraire et les simples et les collocations. Quant aux relations, certains travaux se sont orientés dans l'extraction de la relation de subsomption, d'autres ont traité les relations de causalité. Nous avons opté pour une démarche intuitive, nous nous sommes intéressés aux relations qui paraissaient les plus évidentes en fonction des termes dont nous disposons et dont les marqueurs sont souvent utilisés dans tout type de corpus.

Dans le cadre de cette thèse, nous nous sommes intéressés aux méthodes d'extraction des termes simples et sous forme de collocations ainsi qu'à l'extraction de relations à partir de corpus arabes. Nous avons présenté le problème de la construction semi-automatique d'ontologies et nous l'avons appliqué sur le texte coranique. Nous avons souligné les difficultés rencontrées lors des différentes étapes de constructions comme la structure complexes des phrases devant le manque d'outils adéquats pour la désambiguïsation et la résolution des coréférences. Le manque de travaux dans ce domaine a fait que nous ne pouvions opter en faveur d'une technique au détriment d'une autre. C'est essentiellement pour cette raison que, nous avons essayé une méthode dans la deuxième partie, puis nous l'avons abandonnée parce qu'elle ne donnait aucune amélioration.

Au niveau de la formalisation, nous avons été contraints de construire un éditeur qui puisse supporter les caractères arabes et les symboles mathématiques, les éditeurs existants ne pouvaient supporter les deux. La formalisation des relations a été une tâche difficile parce qu'on ne disposait d'aucune référence. Ce fût un travail minutieux et itératif pour obtenir un exemple de fichier LD acceptable, plus pénible encore la conception de l'application pour l'exploiter et la réaliser. A la fin, nous avions une idée globale et plus claire sur ce qu'il fallait compléter ou qui restait à améliorer.

Nous avons considéré la sourate comme un document et la différence de taille a fait que ces documents étaient très hétérogènes par rapport à leur taille, les

méthodes comme celle basée sur tf-idf exigeait des documents approximativement de même longueur, en effet la sourate Al-Baqarah compte 286 versets alors que Al Kawthar ne comporte que 3versets.

Comme perspectives nous envisageons de partitionner le texte coranique non plus en sourate mais en hizb ou en jouz'a, cela pourrait résoudre le problème de l'homogénéité des documents.

Le travail présenté ici, a été effectué étape par étape en quatre parties, comme travail futur, qui sera un vrai travail se synthèse, nous proposons le développement d'une plateforme complète à partir de laquelle nous pouvons accéder à toutes les fonctionnalités du système, en l'occurrence : la construction de corpus à partir de documents sur disque dur ou sur le Web, l'analyse et le prétraitement de ce corpus, l'extraction des termes simples, l'extraction des termes composés, l'identification des relations entre termes simples d'un coté et entre termes composés de l'autre, la formalisation, l'édition de l'ontologie, sa mise à jour et sa restructuration pour corriger d'éventuelles anomalies structurelles ou sémantiques, ce point est un travail innovateur et d'actualité. En plus de la possibilité de l'intégrer dans une application pour l'amélioration de la recherche d'information sur le Web. Nous proposons également le travail avec étiquetage plus affinés et l'écriture de règles JAPE plus complexes pour améliorer la précision. Nous proposons aussi d'utiliser d'autres métriques que les tf-idf et l'IM, puis les comparer pour voir laquelle donnerait les meilleurs résultats aussi bien pour les termes que pour les relations. Concernant l'utilisation du concordancier, nous proposons soit la recherche de concordancier plus développé avec des fonctionnalités complexes soit le traitement du corpus avant l'utilisation du concordancier pour avoir des fréquences plus significatives.

Si le titre porte le mot *plateforme*, ce qui était l'objectif au début du travail, les choses se sont avérées beaucoup plus difficiles, vu le manque de références et d'outils spécialisés, mais ce qui reste à faire est juste un travail de synthèse et quelques améliorations à différents niveaux pour obtenir un produit fini et fonctionnel dans les domaines utilisant la langue arabe. C'est la suite que nous allons donner à ce travail de recherche.

6. Références et bibliographie

مصحف التجويد كلمات القرآن تفسير و بيان مع فهرس مواضيع القرآن، دار المعرفة الطبعة التاسعة مطبعة الثريا دمشق 2010.

Abdelali A., **Cowie J.R., Farwell D., Ogden W.C.,** (2004). *UCLIR: a Multilingual Information Retrieval Tool*. Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial 8(22): 103-110.

Amari S., (2009). *Extraction d'information à partir des textes arabes à l'aide de l'outil Gate,* mémoire de master, Université d'Annaba.

Assadi H., (1998). *Construction d'ontologies à partir de textes techniques, Application aux systèmes documentaires*. Thèse de doctorat. Université Paris 6.

Assadi H., Bourigault D., (1996). Ac*quisition de connaissances à partir de textes: Outils informatiques et éléments méthodologiques*. Actes du dixième congrès Reconnaissance de Formes et Intelligence Artificielle (RFIA' 96), pp 505-514. Rennes.

Atwell E. et Al-Sulaiti L., Al-Osaimi S., Abu Shawar B., (2004). *A Review of Arabic Corpus Analysis Tools.* JEP-TALN 2004, Arabic Language Processing, Fez, 19-22 April 2004 School of Computing, University of Leeds, Leeds LS2 9JT, England.

Azé J., Heitz T., (2004). *Cours sur la Fouille de textes et Apprentissage*, (2004), disponible sur : http://www.lri.fr/~aze/enseignements.php.

Bachimont B. (2000). *Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances*. In R. Teulier, J. Charlet & P. Tchounikine, Coordinateurs, Ingénierie des connaissances, chapitre 19. Paris

Baloul, S., (2003). *Développement d'un système automatique de synthèse de la parole à partir du texte arabe standard voyellé,* Thèse de doctorat, Université du Maine, Académie de Nantes, France.

Baneyx A., (2007). *Construire une ontologie de la pneumologie, aspects théorique, modèles et expérimentations*. Thèse de doctorat, Université Pierre et Marie Curie.

Baneyx A., & Charlet, J., (2006). *Évaluation, évolution et maintenance d'une ontologie en médecine: état des lieux et expérimentation*, Revue I3 ; SI 2006 special issue on Ontological ressources.

Banouni M., Lazrek A., Sami K., (2002). *Une translittération arabe/roman pour un e-document*, Trans-Tec, CFD'02 - CIDE'5.

Béchet N., (2009). Extraction et regroupement de descripteurs morpho-syntaxiques pour des processus de Fouille de Textes, Thèse de doctorat, Université de MontpellierII.

Beguin A., Jouis C., Widad M., (1997). *Evaluation d'outils d'aide à la construction de terminologie et de relations sémantiques entre termes à partir de corpus*. Premières Journées Scientifiques et Techniques (JST) du réseau Francophone de l'ingénierie de langue de l'AUPELF-UREF, pp 419-425. Avignon.

Benmazou S., (2009). *Adaptation d'UNITEX pour l'extraction de concordances à partir de textes arabes,* mémoire de master, Université d'Annaba.

Bernhard D., (2006). *Apprentissage de connaissances morphologiques pour l'acquisition automatique de ressources lexicales*, Thèse de doctorat, Université Joseph Fourier – Grenoble.

Biebow B., Szulman S., (2000). *Une approche terminologique pour la construction d'ontologie de domaine à paritr de textes : TERMINAE*. Actes du douzième congrès Reconnaissance de Formes et Intelligence Artificielle (RFIA' 2000), pp 81-90. Paris.

Bodson C., (2004). *Termes et relations sémantiques en corpus spécialisés : rapport entre patrons de relations sémantiques (PRS) et types sémantiques (TS)* Thèse de doctorat, Université de Montréal.

Boulaknadel S. (2008). *Traitement Automatique des Langues et Recherche d'Information en langue arabe dans un domaine de spécialité : Apport des connaissances morphologiques et syntaxiques pour l'indexation*, Thèse de doctorat, Université de Nantes.

Bourigault D., (1994). *LEXTER, Un logiciel d'Extraction de TERminologie. Application à l'acquisition de connaissances à partir de textes*. Thèse de doctorat. EHESS.

Bourigault D., Jacquemin C., (2000). *Construction de ressources terminologiques*. In Ingénierie des langues, pp 215-230, ed. J.M. Pierrel. Hermes Sciences.

Chaabani M., Mezghiche M., Strecer M. (2009). *Formalisation de la logique de description ALC dans l'assistant de preuve Coq*, JFO 2009 December 3-4 (2009) Poitiers, France.

Charlet J., Bachimont B., Bouaud J., Zweigenbaum P., (1996). *Ontologie et réutilisabilité :expérience et discussion*. In N. Aussenac-Gilles, P. Laublet & C. Reynaud, Coordinateurs, Acquisition et ingénierie des connaissances : tendances actuelles, chapitre 4, p. 69–87. Cepaduès-éditions.

Church K., Hanks P., (1989). *Word Association Norms, Mutual Information, and Lexicography*, dans Computational Linguistics, vol. 16, no 1, mars, p. 22-29.

Condamines A., Rebeyrolle J., (1997). *Construction d'une base de connaissances terminologiques à partir de textes : expérimentation et définition d'une méthode*. Actes des journées d'Ingénierie des Connaissances et Apprentissage Automatique (JICAA' 97), pp 191-206. Roscoff.

Corcho O., Fernandez-Lopez M., Gomez-Pérez A., Lopez-Cima A. (2005). *Building legal ontologies with METHONTOLOGY and WebODE*, In Law and the Semantic Web, number 3369 in LNAI, pages 142–157. Springer-Verlag, 2005.xviii, 109, 111
Disponible sur *http://www.cs.man.ac.uk/~ocorcho/documents/LawSemWeb2004_CorchoEtAl.pdf*.

Cowie, J., Jin, W., Abdelali, A., Mansouri Rad, H. (2004). *CRL Language Resources: Chinese and Arabic* . Memoranda in Computer and Cognitive Science MCCS-04-333.

Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V., Dimitrov, M., Aswani, N., Roberts, I., (2006). *Developing language processing components with Gate version 3.1 (a user guide).*

Cunningham H., Maynard D., Bontcheva K., and Tablan V., (2002). *GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications*. In Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02).

Daille B., (1994)**.** *Approche mixte pour l'extraction de terminologie : statistique lexicale et filtres linguistiques*. Thèse de doctorat, Université de Paris 7.

Daoust F., (1992). *SATO (Système d'Analyse de Textes par Ordinateur) version 3.6 : Manuel de Référence*. Centre ATO Université du Québec à Montréal.

David S., Plante P. (1990). *De la nécessité d'une approche morpho-syntaxique dans l'analyse de textes.* Intelligence Artificielle et Sciences Cognitives au Québec, 3(3), 140–154.

Debili F., Achour, H., (1998). *Voyellation automatique de l'Arabe,* Proceeding Semitic '98 Proceedings of the Workshop on Computational Approaches to Semitic Languages.

Dias G., (2002). *Extraction automatique d'associations lexicales à partir de corpora.* Thèse de Doctorat, Université d'Orléans.

Djabourabi M. (2011). *Utilisation d'un fichier LD pour la visualisation de concepts en arabe*, Mémoire de master, Université d'Annaba.

Douzidia F. S., (2004). *Résumé automatique de texte arabe*, Mémoire de M.Sc en informatique Université de Montréal, Québec.

Drouin P., (2002). *Acquisition automatique des termes : l'utilisation des pivots lexicaux spécialisés*, Thèse de doctorat, Université de Montréal.

Dubois J., Guespin L., Giacomo M., Marcellesi C., **MÉVEL J.**, (1994), *Dictionnaire de linguistique et des sciences du langage*. Collection Trésors du Français, Larousse. Paris. 1994.

Enguehard C., (1993). *ANA, Apprentissage Naturel Automatique d'un réseau sémantique*. Thèse de doctorat. Université de Compiègne.

Faure D., Poibeaut T., (2000). *Extraction d'information utilisant INTEX et des connaissances sémantiques apprises par ASIUM, premières expérimentations*. Actes du douzième congrès Reconnaissance de Formes et Intelligence Artificielle (RFIA' 2000), pp 91-100. Paris.

Fernandez M., Gomez-Pérez A., Juristo N., (1997). *Methontology: from ontological art towards ontological engineering*. In Spring Symposium Series on Ontological Engineering, National Conference of the American Association on Artificial Intelligence (AAAI).

Fotzo H.N., Gallinari P., (2004) *Information access via topic hierarchies and thematic annotations from document collections.* In International Conference on Enterprise Information Systems, pages 69-76.

Gagnon M., (2004). *Logique descriptive et OWL,*Cours disponible sur: http://www.cours.polymtl.ca/inf6410/Documents/logique_descriptive.pdf

Gandon F. (2002). *Ontology Engineering : a Survey and a Return on Experience*. Rapport interne 4396, INRIA. 181 p., ISSN 0249-6399

Garcia D., (1998). *Analyse automatique des textes pour l'organisation causale des actions, Réalisation du système informatique COATIS*. Thèse de doctorat, Université de Paris -Sorbonne.

Gomez-Pérez A. (2004). *Ontology Evaluation*, In S. Staab & R. Studer, Coordinateurs, Handbook on Ontologies, chapitre, p. 251–275. Handbooks in Information Systems. Springer.

Grefenstette G., (1994). *Explorations in automatic thesaurus discovery*, Kluwer Academic Publishers. Boston.

Gruber T. (1993). A *translation approch to portable ontology specifications*. Knowledge acquisition, 5(2), 199–220.

Hadj henni M. (2007). *Approche ontologique pour la modélisation sémantique, l'indexation et l'interrogation des documents Coraniques*, Mémoire de Magister, Ecole Supérieur d'Informatique (E.S.I) Alger.

Harrathi F., (2009). *Extraction de concepts et de relations entre concepts à partir des documents multilingues : approche statistique et ontologie*, thèse de doctorat, Institut national des sciences appliquées de Lyon.

Harris Z. S., (1968*). Mathematical structures of language*. Wiley, New York.

Hearst M., (1992). *Automatic Acquisition of Hyponyms from Large Text Corpora,* In Proceedings of the 13th international Conference On Computational Linguistics (COLING), pp 539-545. Nantes.

Heitz, T., (2006). *Modélisation du prétraitement des textes*, In Proceedings, JADT 8ème Journées internationales d'Analyse statistique des Données Textuelles, France.

Heitz T., (2008). *Une méthode pour le prétraitement des textes : dépendances entre traitements et leur intelligibilité*, Thèse de doctorat, Université Paris-Sud 11.

Isaac, A., (2005). *Conception et utilisation d'ontologies pour l'indexation de documents audiovisuels,* Thèse de doctorat, Université Paris IV – Sorbonne.

Jarrar M., Ayesh S., Al-Badawi M., Samara H. (2010). *Towards Building An Arabic Ontology*. Technical Report, Faculty of Information Technology, Birzeit University.

Jouis C., (1993*). Contribution à la Conceptualisation et à la Modélisation des connaissances à partir d'une analyse linguistique de textes. Réalisation d'un prototype: le système SEEK*. Thèse de doctorat, EHESS.

Khurshid A., (1996). *Language engineering and the processing of specialist terminology*, http://www.computing.surrey.ac.uk/ai/pointer/paris.html, 27 juin 1996.

Khurshid A., Fulford H., (1992). *Knowledge processing 4. Semantic relations and their use in elaborating terminology*, in Computing Sciences Technical Report CS-92-07, Guildford, Surrey.

Klai S., Khadir M-T., (2009). *Datat based Ontology Construction coupled to Expert System for Steam Turbine Aided Diagnostic,* Published in ewic journal:

Electronic Workshops in Computing Series (eWiC: http://ewic.bcs.org, ISSN 1477-9358), The British Computer Society (BCS).

Koeva S., Maurel D., Silberztein M., (2007). *Formaliser les langues avec l'ordinateur : de Intex à Nooj* Presses Univ. Franche-Comté, 2007 - 438 pages.

Lalaouna Y., (2009). *Adaptation de l'outil Exit pour l'extraction d'information arabe,* mémoire de master, Université d'Annaba.

Lanani F., (2010). *Utilisation d'Aramorph pour l'extarction de mots arabes,* Mémoire de Master, Université d'Annaba.

Lebart L., Salem A. (1988). *Analyse statistique des données textuelles*. Paris : Dunod, Bordas.

Lebhour F-Z, (2009). *Extraction d'information arabe à l'aide de NOOJ,* mémoire de master, Univeristé d'Annaba.

Le Priol F., Chevallet J -P., Brunadet M-F., Desclès J-P., (1998). *Intégration d'un système statistique (IOTA) et d'un système sémantique (SEEK) dans une chaîne de traitement permettant l'extraction de terminologies*, Actes Ingénierie des Connaissances (IC' 98), pp 33-40. Pont-à-Mousson.

L'Homme M.-C. (2001). *Nouvelles technologies et recherche terminologique. Techniques d'extraction des données terminologiques et leur impact sur le travail du terminographe*. In L'impact des nouvelles technologies sur la gestion terminologique, Toronto.

Lounis,L., (2009). *REMARAB :Un outil de recherche et d'extraction de mots à partir d'un texte arabe(Application sur le saint-Coran,* Mémoire de Master, Université d'Annaba.

Malaisé V., (2005). *Méthodologie linguistique et terminologique pour la structuration d'ontologies différentielles à partir de corpus textuels,* Thèse de doctorat, Université Paris 7 – Denis Diderot France.

Maynard D., Aswani, N. (2009). *Annotation and evaluation,* tutorial summer school, University of Sheffield.

Mesfar S. (2008). *Analyse morpho-syntaxique automatique et reconnaissance des entités nommées en arabe standard*, Thèse de doctorat, Université de Franche-Comte, France.

Mhiri M, Gargouri F, Benslimane D, (2006). *Détermination automatique des relations sémantiques entre les concepts d'une ontologie,* In Proceedings of INFORSID'2006. pp.627~642

Mizoguchi R. and Ikeda M. (1997). *Towards Ontology Engineering,* Technical Report AI-TR-96-1, I.S.I.R., Osaka University, Japan.

Morin E., (1999). *Extraction de liens sémantiques entre termes à partir de corpus de textes techniques.* Thèse de doctorat, Université de Nantes.

Napoli A. (1997). *Une introduction aux logiques de descriptions* N° 3314 Décembre 1997 thème 3 rapport de stage.

Nardi D., Brachman R. (2003). *The Description Logic Handbook : Theory, Implementation and Applications,* chapitre An introduction to description logics., p. 544. Cambridge University Press.

Noy N. F., McGuinness D. L., (2000). *Développement d'une ontologie 101 : Guide pour la création de votre première ontologie,* Université de Stanford.

Papini O., (2002). *Introduction au WEB Sémantique, Cours 3 : Introduction aux logiques de description,* ESIL Université de la méditerranée, http://odile.papini.perso.esil.univmed.fr/index.html.

Patil L., Dutta D., Sriram R. (2005).*Ontology formalization of product semantics for product lifecycle management.* Proc. ASME/IDETC CIE Conf., Long Beach, CA

Paumier S., (2009)., *Unitex2.0, user manual*, UniversitéParis-EstMarne-la-Vallée,

Perron, J. (1996). *ADEPTE-NOMINO : un outil de veille terminologique*, dans Terminologies nouvelles, no 15, juin et décembre, Bruxelles, RINT, p. 32-47.

Piwowarski, B, (2003). *Techniques d'apprentissage pour le traitement, d'informations structurées : Application à la recherche d'information,* Thèse de doctorat, Université Paris 6**.**

Plamondon, L., (2004). *L'ingénierie de la langue avec GATE,* RALI/DIRO, Université de Montréal.

Roberts Andrew., Al-Sulaiti L., Atwell E., (2005). *aConCorde: towards a proper concordance for Arabic,* in P. Danielsson and M. Wagenmak ers (eds.) Proceedings of the Corpus Linguistics 2005 Conference, University of Birmingham, UK.

Roberts Angus, Gaizauskas R., Hepple M., Demetriou G., Guo Y., Setzer A., Roberts I., *Semantic Annotation of Clinical Text*: The CLEF Corpus AMIA Annu Symp Proc 2007:625–9.

Roche M., (2006). *Fouille de textes : enjeux, limites et perspectives des méthodes de classification* Exposé dans le cadre de la Journée Thématique : Information, Connaissance et Apprentissage du LIRMM, Montpellier.

Rousselot F., Frath P., Oueslati R., (1996). *Extracting Concepts and Relations from Corpora,* In Proceedings of ECAI Workshop on Corpus-Oriented semantic analysis. Budapest.

Sager J. C., (1990). *A Practical Course in Terminology Processing,* Amsterdam/Philadelphia, John Benjamins.

Séguéla P., (2001). *Construction de modèles de connaissances par analyse linguistique de relations lexicales dans les documents techniques*, Thèse de doctorat, Université de ToulouseIII.

Silberztein M. et Tutin A., (2004). *NooJ : un outil TAL de corpus pour l'enseignement des langues et de la linguistique Une application à l'étude des impersonnels* , Université de Franche-Comté.

Smadja, F. (1993). *Retrieving Collocations from Text: Xtrac*, Computational Linguistics 19(1), pp. 143-177.

Snow R., Jurafsky D., Andrew Y., (2004). *Learning syntactic patterns for automatic hypernym discovery*, In Advances in Neural information Processing Systems.

Sowa J. (2000). *Ontology, metadata and semiotics*. In 8th International Conference on Conceptual
Structures (ICCS'2000), volume 1867, p. 55–81 : Springer-Verlag LNCS.

Sundblad H., (2002) *Automatic Acquisition of Hyponyms and Meronyms from Question Corpora*, in Proceedings of the Workshop on Natural Language Processing and Machine Learning for Ontology Engineering at ECAI'2002, Lyon, France.

Thakker, D., Sman, T., Lakin, P., (2009). *GATE JAPE Grammar Tutorial*, Version 1.0, A,Photos, UK.

Toussaint Y., Royaute J., Muller C., Polanco X., (1997). *Analyse linguistique et infométrique pour l'acquisition et la structuration des connaissances*, Actes des deuxièmes rencontres Terminologie et Intelligence Artificielle (TIA'97) , pp 27-46. Toulouse.

Uschold M. & Grüninger M. (1996). *Ontologies : Principles, methods and applications*. Knowledge Engineering Review, 11(2).

Velardi P., Missikof M., Fabriani P. (2001). *Using text processing techniques to automatically enrich a domain ontology*. In Proceeding of ACM-FOIS.

Voutilainen A. (1993). *Nptool, a detector of English noun phrases*, In Proceedings of the Workshop on Very Large Corpora, June, Columbus, Ohio State University, p.48-57.

Welty C., & Guarino, N., (2001). *Supporting Ontological Analysis of Taxonomic Relationships* Data et Knowledge Engineering (39), pages 51-74, 2001.

Zaidi S., Abdelali A., Sadat F., Laskri, M-T., (2012a). *Hybrid approach for extracting collocations from Arabic Quran text, In Proceedings of LREC 2012, May 2012, Istanbul, Turkey*

Zaidi S., Abdelali A., Laskri, M-T., (2012b). *Extracting and Formalizing Terms and Relations to Build Ontology, Publication dans IJMSO (International Journal of Metadata, Semantics and Ontologies: En cours d'impression) ISSN (Online): 1744-263X - ISSN (Print): 1744-2621*

Zaidi, S., Abdelali A., Laskri, M-T., Eshennifi M.A., (2011). *Extraction des termes simples et composés à partir de textes arabes***,** Communications of the Arab Computer Society, Vol. 4 No.1, August, 2011 ISSN 1090-102X.

Zaidi, S., Abdelali, A. Laskri, M-T., (2010a), *Extraction des collocations à partir de textes arabes avec l'outil GATE (Application sur le Saint Coran)*, Journée d'étude sur le contenu numérique en arabe dans le cadre du système du e-gouvernement, Alger, Algérie. (*À paraitre comme chapitre dans un livre*) Editeur CSLA.

Zaidi, S., Laskri, M-T, Abdelali, A. (2010b). *Arabic collocations extraction based on JAPE rules,* In Proceedings, Acit Arab Conference of Information and Technology, Benghazi, Libya.

Zaidic, S., Laskri, M-T, Abdelali, A. (2010c). *Arabic collocations extraction using Gate,* In Proceedings, ICMWI'une conférence internationale sur les machines et le web intelligents IEEE, Algiers, Algeria.

Zaidi, S., Laskri, M-T, Abdelali, A. (2010d). *Étude d'adaptabilité d'outils de terminologie textuelle à l'Arabe,* In Proceedings COSI,colloque sur l'optimisation et les systemes d'information, Ouragla, Algerie.

Zaidi, S., Laskri, M-T, Abdelali, A. (2010e). *Utilisation de Gate pour l'extraction d'information à partir de corpus arabes*, In proceedings, JED, Journées Ecoles doctorales et Réseaux de recherche, Annaba, Algérie.

Zaidi, S., Laskri (2009). *Review of textual terminology tools for ontologies building,* In proceedings, MIC'09 Management International conference, 25- 28 november, 2009 Sousse Tunisia.

Zipf. G. K., (1949). *Human Behavior and the Principle of Least Effort*, New York, Harper, réédition 1966.

Zitoun, M. (2010). *Développement d'un éditeur pour la formalisation d'une ontologie en arabe,* mémoire de master, Université d'Annaba.

Zweigenbaum P., Bachimont B., Bouaud J., Charlet J., Boisvieux J.-F. (1995). *A multilingual architecture for building a normalised conceptual representation from medical language*. Journal of the American Medical Informatics Association, 2(suppl), 357–361.

Figures

Tableaux