

Université Saad DAHLAB - Blida 1



Faculté des sciences

Département d'Informatique

Mémoire présenté par :

Mr BOUDEKAK MOHAMED

Pour l'obtention du diplôme de Master

Domaine : Mathématique et Informatique

Filière : Informatique

Spécialité : Traitement Automatique de la Langue

**Conception et réalisation d'un outil d'indexation automatique de la
tradition musulmane (Sunna) basé sur l'extraction des alfadh al hadith**

Soutenu le : 24-12-2020

Devant le jury composé de :

- Mme Mezzi M., Présidente
- M. Kamèche H. A., examinateur
- M. Chérif-Zahar A., Promoteur

Résumé

« Alfadh el Hadith » sont des portions de discours portant en elles un sens indépendamment du reste du discours. Dans un hadith comme :

{ "حدثنا يحيى بن يحيى التميمي, وأبو بكر بن أبي شيبة ومحمد بن العلاء الهمداني عن أبي هريرة, قال قال رسول الله صلى الله عليه وسلم" } {من نفس عن مؤمن كربة من كرب الدنيا نفس الله عنه كربة من كرب يوم القيامة (1) ومن يسر على معسر يسر الله عليه في الدنيا والآخرة (2) ومن ستر مسلما ستره الله في الدنيا والآخرة (3) والله في عون العبد ما كان العبد في عون أخيه (4) ومن سلك طريقا يلتمس فيه علما سهل الله له به طريقا إلى الجنة (5) وما اجتمع قوم في بيت من بيوت الله يتلون كتاب الله ويتدارسونه بينهم إلا نزلت عليهم السكينة (6) وغشيتهم الرحمة (7) وحفتهم الملائكة (8) وذكرهم الله فيمن عنده (9) ومن بطأ به عمله لم يسرع به نسبه (10), }.

On remarque qu'il est composé de deux parties essentielles (sont entre {}).

On remarque aussi qu'il y a plusieurs portions porteuses de sens (numérotées dans le texte).

Le présent mémoire décrit un modèle qui permet d'extraire automatiquement ces portions en vue d'une indexation.

Le présent travail a décroché une précision de 34%

Mots clés: Alfadh el hadith, Sunna, 'ilm al hadith, Sanad, Matn, Segmentation, Algorithme, Grammaire, Machine Learning, Tal.

Abstract

“Alfadh el Hadith” are portions of speech that carry meaning independent of the rest of the speech. In a hadith as:

{|"} حدثنا يحيى بن يحيى التميمي, وأبو بكر بن أبي شيبه ومحمد بن العلاء الهمداني عن أبي هريرة, قال قال رسول الله صلى الله عليه وسلم"} {من نفس عن مؤمن كربة من كرب الدنيا نفس الله عنه كربة من كرب يوم القيامة (1) ومن يسر على معسر يسر الله عليه في الدنيا والآخرة (2) ومن ستر مسلما ستره الله في الدنيا والآخرة (3) والله في عون العبد ما كان العبد في عون أخيه (4) ومن سلك طريقا يلتمس فيه علما سهل الله له به طريقا إلى الجنة (5) وما اجتمع قوم في بيت من بيوت الله يتلون كتاب الله ويتدارسونه بينهم إلا نزلت عليهم السكينة (6) وغشيتهم الرحمة (7) وحفتهم الملائكة (8) وذكرهم الله فيمن عنده (9) ومن بطأ به عمله لم يسرع به نسبه (10), {."}

Note that it is composed of two essential parts (they are between {}).

It is noted that there are several meaningful portions (numbered in the text).

This memo describes a model that automatically extracts these portions for indexing.

The present work has achieved an precision of 34%.

Keywords: Alfadh el hadith, Sunna, ‘ilm al hadith, Sanad, Matn, Segmentation, Algorithm, Grammar, Machine Learning, NLP.

ملخص

"ألفاظ الحديث" هي أجزاء من الكلام تحمل معاني مستقلة عن بقية الخطاب. في حديث نبوي مثل: {حدثنا يحيى بن يحيى التميمي, وأبو بكر بن أبي شيبة ومحمد بن العلاء الهمداني عن أبي هريرة, قال قال رسول الله صلى الله عليه وسلم} {من نفس عن مؤمن كربة من كرب الدنيا نفس الله عنه كربة من كرب يوم القيامة (1) ومن يسر على معسر يسر الله عليه في الدنيا والآخرة (2) ومن ستر مسلما ستره الله في الدنيا والآخرة (3) والله في عون العبد ما كان العبد في عون أخيه (4) ومن سلك طريقا يلتمس فيه علما سهل الله له به طريقا إلى الجنة (5) وما اجتمع قوم في بيت من بيوت الله يتلون كتاب الله ويتدارسونه بينهم إلا نزلت عليهم السكينة (6) وغشيتهم الرحمة (7) وحفتهم الملائكة (8) وذكرهم الله فيمن عنده (9) ومن بطأ به عمله لم يسرع به نسبه (10), "}.
يلاحظ أنه يتكون من جزأين أساسيين (بينهما {}).
يلاحظ أن هناك عدة أجزاء ذات معنى (مرقمة في النص).
تصف هذه المذكرة نموذجًا يقوم تلقائيًا باستخراج هذه الأجزاء للفهرسة.
العمل الحالي حقق دقة 34٪ .

كلمات مفتاحية: ألفاظ الحديث، السنة، علم الحديث، سند، المتن، تجزئة، خوارزمية، قواعد، تعلم الآلة، المعالجة الألية للغات.

Dédicaces

Nous aimerions dédier ce mémoire en guise de consécration d'une vie d'études à nos parents bien-aimés qui nous ont aidés à bien des égards au cours de ce voyage, que Dieu les bénisse.

À nos sœurs, superviseurs et amis qui ont partagé leurs conseils et leurs encouragements pour terminer ce travail.

Et enfin, à nous-mêmes pour être dans ce long voyage et pour l'avoir fait jusqu'au bout.

Mohamed

Remerciement

Tout d'abord, nous tenons à remercier Allah qui nous a donné la force et le courage de terminer ce travail, et nos parents d'être notre plus grand soutien.

Nous tenons également à remercier tous ceux qui nous ont aidés à faire ce travail, notre superviseur Mr CHERIF ZAHAR en particulier pour être toujours solidaire quelle que soit la situation.

Table des matières

Introduction Générale.....	17
1 Problématique.....	18
2 Qu'est-ce que les alfadh el hadith ?.....	22
3 Modélisation du concept des alfadh el hadith	23
3.1 Définition :	23
3.2 Extraction des alfadh:	27
4 Objectifs.....	28
Chapitre I : ETAT DE L'ART	29
1 Introduction.....	30
2 Présentation des travaux manuelle.....	30
2.1 تحفة الأشراف بمعرفة الأطراف:	30
2.2 إتحاف المهرة بأطراف العشرة:	30
2.3 أطراف المسانيد العشرة	31
2.4 ذخائر المواريث في الدلالة على مواضع الحديث:	31
2.5 أطراف المسند المعتلى بأطراف المسند الحنبلي:	31
2.6 أطراف الحديث المختارة للضياء المقدسي:	31
2.7 أطراف مسند الفردوس:	32
2.8 أطراف صحيح ابن حبان:	32
2.9 موسوعة أطراف الحديث النبوي الشريف:	32
3 Limites des travaux manuels	36
4 Présentation des travaux informatique	37
5 Limites des travaux informatiques :	43
6 Conclusion	44
Chapitre II : Conception de l'outil d'extraction.....	45

1.	Introduction.....	46
2	Rappel des notions de la linguistique arabe	46
2.1	L'analyse des mots	46
2.2	Les Phrases	47
2.2.1	Phrase Verbale :	47
2.2.2	Phrase nominale :.....	47
2.3	El Mouaaraf et Nakira(المعرف ب ال و النكرة):	47
2.4	Jar et Majrur (الجار و المجرور):.....	47
2.5	El idafa (الإضافة) :.....	48
2.5.1	El modaf (مضاف).....	48
2.5.2	El modaf ilihe (مضاف إليه).....	48
2.5	Charte (أسلوب الشرط)	48
2.6	Tachbihe (تشبيه):.....	48
3	Conception générale :.....	49
3.1	Extraction des matn (Séparation sanad et matn):.....	50
3.1.1	Introduction :.....	50
3.1.2	'an (عن):.....	51
3.1.3	Harf nidaa (حرف نداء):.....	52
3.1.4	Anna (أن):	53
3.1.5	Verbes :	54
3.1.6	Résumé de la première opération.....	57
3.1.7	Déroulement	58
3.1.8	Indexation matn et sanad	62
3.1.9	Conclusion	62
3.2	Extraction des alfadh (Création des alfadh):.....	63

3.1.1	Introduction.....	63
3.1.2	Waw(واو) :.....	66
3.1.3	Aw et Thoma (أو, ثم):	68
3.1.4	Illa (إلا):.....	69
3.1.5	Hatta (حتى) :.....	69
3.1.6	Kaana (كأن):.....	70
3.1.7	Inna (إن):.....	70
3.1.8	Damir kasd (ضمير القصد):.....	71
3.1.9	Kala (قال)	71
3.1.10	Résumé	72
3.1.11	Conclusion	72
4	Conclusion	73
Chapitre III : Mise en œuvre et expérimentation		74
1	Introduction.....	75
2	Outils et bibliothèques.....	75
2.1	Python.....	75
2.2	Flask	75
2.3	Tashaphyne.....	76
2.4	Black-Box.....	76
2.5	ReportLab	77
2.6	Arabic_Reshaper	77
2.7	Kacstqurn.....	77
3	Présentation des résultats	78
3.1	Présentation du dataset.....	78
3.2	Présentation des données de teste.....	79

3.3	Présentation des interfaces	80
3.4	Résultats	81
4	Conclusion	83
	Conclusion générale	84
1	Sommaire.....	85
2	Perspective	86
	Références bibliographiques	87
	Bibliographie	88

Liste des figures

Figure 1: Illustration explicative de ce que sont les alfadh	20
Figure 2: Exemple explicatif des cas d'utilisation des alfadh	22
Figure 3: Exemple de alfadh tirés de l'annexe de l'encyclopédie [32]	26
Figure 4: Exemple de alfadh tirés de l'encyclopédie [32]	26
Figure 5: Notre vision d'un hadith	27
Figure 6: Liste des symboles utilisés dans l'encyclopédie	33
Figure 7: Aperçu de l'organisation de l'encyclopédie Alif mamdoua	34
Figure 8: Aperçu de l'organisation de l'encyclopédie Inna	34
Figure 9: Aperçu de l'organisation de l'encyclopédie Yaa	35
Figure 10: Aperçu de l'organisation de l'encyclopédie dans l'annexe	35
Figure 11: application de l'algorithme se basant sur l'HTML	38
Figure 12: conception du travail se basant sur l'HTML	39
Figure 13: conceptions du travail des n-gramme	40
Figure 14: Conception de l'outil d'extraction	49
Figure 15: Aperçu de la particule 'an	51
Figure 16: Aperçu du nidaa	52
Figure 17: Aperçu de Anna cas simple	53
Figure 18: Aperçu de Anna cas complexe	53
Figure 19: Aperçu du cas simple des verbes	54
Figure 20: Aperçu du 1er cas complexe des verbes	55
Figure 21: Aperçu du 2eme cas complexe des verbes	56
Figure 22: Aperçu du 3eme cas complexe des verbes	56
Figure 23: Aperçu général de l'algorithme de séparation	57
Figure 24: Aperçu des possibilités de commencement du matn	64

Figure 25: Aperçu des types de condition de alfadh.....	65
Figure 26:Aperçu des conditions de succession du waw	66
Figure 27:Aperçu des cas de traitement du waw.....	67
Figure 28: Aperçu des cas de charte.....	69
Figure 29: Aperçu des cas de ka-ana	70
Figure 30: Aperçu des cas de inna	70
Figure 31: Aperçu générale de l’algorithme des alfadh	72
Figure 32: Aperçu du logo de python	75
Figure 33: Aperçu du logo de Flask.....	76
Figure 34: Aperçu de l’architecture du dataset	78
Figure 35: fichier d’importation pour test.....	79
Figure 36: Interface d’importation du fichier.....	80
Figure 37: Interface de téléchargement du fichier	80
Figure 38: résultats de l’échantillon	81
Figure 39: Formule de précision	82
Figure 40: Formule de précision des alfadh	82

Liste des Tableaux

Tableau 1: Exemple d'alfadh.....	20
Tableau 2: Exemple d'un sanad et matn d'un hadith	23
Tableau 3: Exemple des alfadh, matn et sanad d'un hadith (d'après l'encyclopédie [31])	24
Tableau 4: Revue des Précédente Recherches sur la Segmentation du Hadith.....	41
Tableau 5: Mots de transmission.....	50
Tableau 6: Déroulement de l'algorithme de séparation (1 ^{er} exemple)	59
Tableau 7: Déroulement de l'algorithme de séparation (2 ^{ème} exemple)	60
Tableau 8: Liste des particules pris en compte	64
Tableau 9: Liste des différentes apparitions du waw.....	66
Tableau 10: Ahadith utilisé pour le 1 ^{er} teste	79
Tableau 11: résultats finaux.....	82

Introduction Générale

1 Problématique

La Sunna est, avec le Coran, l'une des deux sources de la religion islamique. La sunna représente la vie du prophète en quelque sorte : ses paroles, ses actions, ses consentements.

La sunna est représentée sous la forme de narrations qu'on appelle les ahadiths (pluriel de hadith) qui ont été transmises de génération en génération par des garants (ruwât) narrant chacun ce qu'il a entendu de ces prédécesseurs directs et ceux jusqu'au locuteur terminal : Le prophète ou bien un de ses compagnons.

Quelques temps après la mort du prophète, la Sunna a été regroupée dans des recueils de crainte qu'elle ne se perde.

Des docteurs de la loi (imams) se sont chargés de la sauvegarder et ont ainsi donné naissance à des recueils qui diffèrent par la façon avec laquelle ils ont été groupés, par les critères retenus pour garder ou rejeter une narration etc.

Comme certains imams avaient le dessein de sauvegarder les narrations sans se soucier si elles étaient recevables ou non, cela a donné naissance quelques temps après à tout une science élaborée dite 'ilm al-hadith qui a posé les fondements pour dire si une narration est recevable ou non. Cela a englobé une sous discipline qui est appelé 'ilm al-djarh wa-t-ta'dîl qui est un système d'enquêtes faites sur la probité ('adala) des garants (ruwât) et leur aptitude à transmettre fidèlement la sunna (dabt).

Les spécialistes de hadith pour trancher si un hadith est recevable ou non doivent avoir à leur disposition les différentes narrations d'un hadith. Cela permet de savoir si le hadith est fort (sahih), acceptable (hassan) ou faible (da'if).

Le travail devient difficile dès lors qu'il faut naviguer dans une quantité énorme d'informations et de plus, les narrations ne se font pas toujours dans le langage intégral du prophète mais dans le sens de ce qu'il a dit. Quelques fois, un narrateur se rappelle d'une tournure de phrase tandis qu'un autre se rappelle d'une autre, variante de la première. Exemple : certains rapportent que le prophète a dit : « ما نقص مال من صدقة » alors que d'autres ont rapporté : « ما نقصت صدقة من مال ». Pour voir si les différentes narrations se confortent au besoin, il faut les recenser -à supposer qu'on les connaisse-.

Introduction générale

Jadis des oulémas ont travaillé sur une indexation de la sunna par les propositions : segments de phrases ayant un sens complet dans un hadith et qu'on appelle alfadh (pluriel de lafdh, extrémité). Cela permettait de se retrouver facilement. Toutefois, leurs efforts étaient proportionnels aux moyens qu'ils avaient et personne ne pouvait faire un travail complet.

Seul l'avènement de l'ordinateur permet cela.

On peut, si on trouve l'algorithme adéquat, créer cet indexe complet qui permettra aux spécialistes d'avancer rapidement dans leurs travaux.

Si nous prenons ce hadith :

حدثنا يحيى بن بكير، قال حدثنا الليث، عن عقيل، عن ابن شهاب، عن عروة بن الزبير، عن عائشة أم المؤمنين، أنها قالت أول ما بدئ به رسول الله صلى الله عليه وسلم من الوحي الرؤيا الصالحة في النوم، فكان لا يرى رؤيا إلا جاءت مثل فلق الصبح، ثم حُبب إليه الخلاء، وكان يخلو بغار حراء فيتحنث فيه وهو التعبد الليالي ذوات العدد قبل أن ينزع إلى أهله، ويتزود لذلك، ثم يرجع إلى خديجة، فيتزود لمثلها، حتى جاءه الحق وهو في غار حراء، فجاءه الملك فقال اقرأ. قال " ما أنا بقارئ ". قال " فأخذني فغطني حتى بلغ مني الجهد، ثم أرسلني فقال اقرأ. قلت ما أنا بقارئ. فأخذني فغطني الثانية حتى بلغ مني الجهد، ثم أرسلني فقال اقرأ. فقلت ما أنا بقارئ. فأخذني فغطني الثالثة، ثم أرسلني فقال {اقرأ باسم ربك الذي خلق * خلق الإنسان من علق * اقرأ وربك الأكرم} ". فرجع بها رسول الله صلى الله عليه وسلم يرجف فؤاده، فدخل على خديجة بنت خويلد رضي الله عنها فقال " زملوني زملوني ". فزملوه حتى ذهب عنه الروع، فقال لخديجة وأخبرها الخبر " لقد خشيت على نفسي ". فقالت خديجة كلا والله ما يخزيك الله أبدا، إنك لتصل الرحم، وتحمل الكل، وتكسب المعدوم، وتقري الضيف، وتعين على نوائب الحق. فانطلقت به خديجة حتى أتت به ورقة بن نوفل بن أسد بن عبد العزى ابن عم خديجة وكان امرأ تنصر في الجاهلية، وكان يكتب الكتاب العبراني، فيكتب من الإنجيل بالعبرانية ما شاء الله أن يكتب، وكان شيخا كبيرا قد عمي فقالت له خديجة يا ابن عم اسمع من ابن أخيك. فقال له ورقة يا ابن أخي ماذا ترى فأخبره رسول الله صلى الله عليه وسلم خبر ما رأى. فقال له ورقة هذا الناموس الذي نزل الله على موسى صلى الله عليه وسلم يا ليتني فيها جذعا، ليتني أكون حيا إذ يخرجك قومك. فقال رسول الله صلى الله عليه وسلم " أومخرجي هم ". قال نعم، لم يأت رجل قط بمثل ما جئت به إلا عودي، وإن يدركني يومك أنصرك نصرًا مؤزرا. ثم لم ينشب ورقة أن توفي وفتر الوحي.

On retrouve plusieurs parties du hadith où le locuteur est le prophète lui-même impliquant l'existence de plusieurs alfadh en l'occurrence cinq alfadh existent. Dans cet exemple de hadith, on a :

Tableau 1: Exemple d'alfadh

Premier	Deuxième	Troisième	Quatrième	Cinquième
"ما أنا بقارئ"	فأخذني فغطني حتى بلغ مني الجهد، ثم أرسلني فقال اقرأ. قلت ما أنا بقارئ. فأخذني فغطني الثانية حتى بلغ مني الجهد، ثم أرسلني فقال اقرأ. فقلت ما أنا بقارئ. فأخذني فغطني الثالثة، ثم أرسلني فقال {اقرأ باسم ربك الذي خلق * خلق الإنسان من علق * اقرأ وربك الأكرم}	زملوني زملوني	لقد خشيت على نفسي	أومخرجي هم

حدثنا يحيى بن بكير، قال حدثنا الليث، عن عقيل، عن ابن شهاب، عن عروة بن الزبير، عن عائشة أم المؤمنين، أنها قالت أول ما بدئ به رسول الله صلى الله عليه وسلم من الوحي الرؤيا الصالحة في النوم، فكان لا يرى رؤيا إلا جاءت مثل فلق الصبح، ثم حيب إليه الخلاء، وكان يخلو بغار حراء فيتحنث فيه وهو التعبد الليالي ذوات العدد قبل أن ينزع إلى أهله، ويتزود لذلك، ثم يرجع إلى خديجة، فيتزود لمثلها، حتى جاءه الحق وهو في غار حراء، فجاءه الملك فقال اقرأ. قال " ما أنا بقارئ ". قال " فأخذني فغطني حتى بلغ مني الجهد، ثم أرسلني فقال اقرأ. قلت ما أنا بقارئ. فأخذني فغطني الثانية حتى بلغ مني الجهد، ثم أرسلني فقال اقرأ. فقلت ما أنا بقارئ. فأخذني فغطني الثالثة، ثم أرسلني فقال {اقرأ باسم ربك الذي خلق * خلق الإنسان من علق * اقرأ وربك الأكرم} ". فرجع بها رسول الله صلى الله عليه وسلم يرجف فؤاده، فدخل على خديجة بنت خويلد رضى الله عنها فقال " زملوني زملوني ". فزملوه حتى ذهب عنه الروع، فقال لخديجة وأخبرها الخبر " لقد خشيت على نفسي ". فقالت خديجة كلا والله ما يخزيك الله أبدا، إنك لتصل الرحم، وتحمل الكل، وتكسب المعدوم، وتقري الضيف، وتعين على نوائب الحق. فانطلقت به خديجة حتى أتت به ورقة بن نوفل بن أسد بن عبد العزى ابن عم خديجة وكان امرأ تنصر في الجاهلية، وكان يكتب الكتاب العبراني، فيكتب من الإنجيل بالعبرانية ما شاء الله أن يكتب، وكان شيخا كبيرا قد عمي فقالت له خديجة يا ابن عم اسمع من ابن أخيك. فقال له ورقة يا ابن أخي ماذا ترى فأخبره رسول الله صلى الله عليه وسلم خبر ما رأى. فقال له ورقة هذا الناموس الذي نزل الله على موسى صلى الله عليه وسلم يا ليتني فيها جذعا، ليتني أكون حيا إذ يخرجك قومك. فقال رسول الله صلى الله عليه وسلم " أومخرجي هم ". قال نعم، لم يأت رجل قط بمثل ما جئت به إلا عودي، وإن يدركني يومك أنصرك نصرًا مؤزرا. ثم لم ينشب ورقة أن توفي وفتر الوحي.

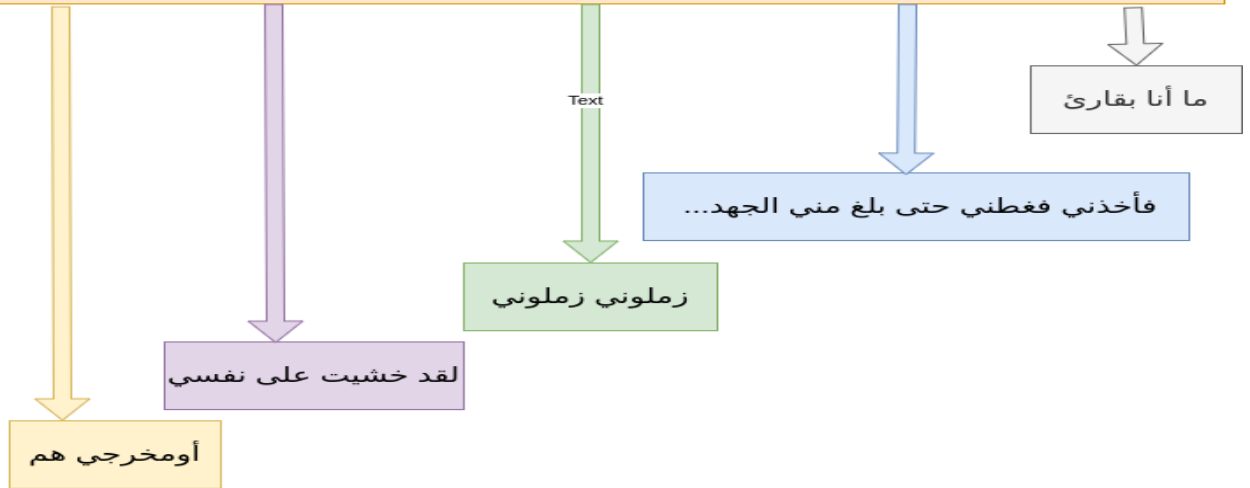


Figure 1: Illustration explicative de ce que sont les alfadh

Introduction générale

La sunna regroupe comme nous l'avons précisé plus haut les ahadiths du prophète Mohamed et de ses compagnons. Ils sont catégorisés en trois types :

I. **Sunna kawlia « سنة قولية »** : Ce sont tous les discours émanant du prophète ou se rapportant à chose qu'il aurait dite verbalement.

Exemple :

حدثنا الحميدي عبد الله بن الزبير، قال حدثنا سفيان، قال حدثنا يحيى بن سعيد الأنصاري، قال أخبرني محمد بن إبراهيم التيمي، أنه سمع علقمة بن وقاص الليثي، يقول سمعت عمر بن الخطاب رضي الله عنه على المنبر قال سمعت رسول الله صلى الله عليه وسلم يقول " إنما الأعمال بالنيات، وإنما لكل امرئ ما نوى، فمن كانت هجرته إلى دنيا يصيبها أو إلى امرأة ينكحها فهجرته إلى ما هاجر إليه ".

II. **Sunna fi'lia « سنة فعلية »** : Ce type de Sunna représente les **Actions** du prophète dans une situation précise telles que les ont rapportés les compagnons du prophète, compagnons incluant ses épouses et ses proches également.

Exemple :

حدثنا عبد الله بن مسلمة، عن مالك، عن سعيد المقبري، عن أبي سلمة بن عبد الرحمن، أنه سأل عائشة رضي الله عنها كيف كانت صلاة رسول الله صلى الله عليه وسلم في رمضان قالت ما كان يزيد في رمضان ولا غيره على إحدى عشرة ركعة، يصلي أربع ركعات فلا تسأل عن حسنهن وطولهن، ثم يصلي أربعاً فلا تسأل عن حسنهن وطولهن، ثم يصلي ثلاثاً فقلت يا رسول الله تنام قبل أن توتر قال " تنام عيني ولا ينام قلبي ".

III. **Sunna takriria « سنة تقريرية »** : Ce type de Sunna (Ahadith) représente les sentences induites par un consentement du prophète Mohamed dans une situation donnée.

Exemple :

حدثنا شعبة، قال أخبرني سعد بن إبراهيم، قال سمعت حفص بن عاصم، قال سمعت رجلاً، من الأزد يقال له مالك ابن بحينة أن رسول الله صلى الله عليه وسلم رأى رجلاً وقد أقيمت الصلاة يصلي ركعتين، فلما انصرف رسول الله صلى الله عليه وسلم لاث به الناس، وقال له رسول الله صلى الله عليه وسلم " الصبح أربعاً، الصبح أربعاً ".

2 Qu'est-ce que les alfadh el hadith ?

Alfadh el hadith sont vue dans la vie réelle par de nombreux expert comme les substance des ahadith qu'une personne remémore, en effet l'un des points les plus importants dans les alfadh el hadith c'est qu'elles sont en grande partie les segments mémorisés des ahadith par une personne car généralement une personne ne mémorise qu'une partie du hadith (sauf peut-être dans le cas des spécialistes dans le rassemblement des ahadith. Un lafdh très répandu dans les bouches des gens ou qui les a marqués par exemple.

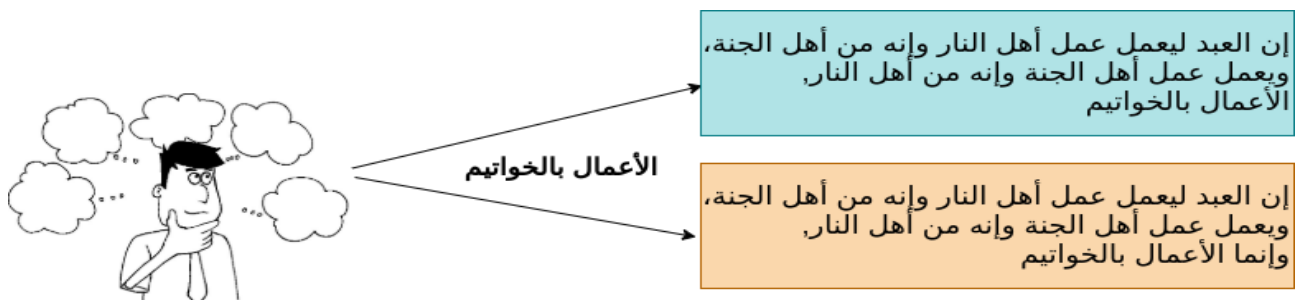


Figure 2:Exemple explicatif des cas d'utilisation des alfadh

Et c'est ce qui rend alfadh el hadith très intéressent pour la recherche en Sunna mais aussi en traitement informatique, car une personne ne se rappelant pas du hadith ou ne se souvenant pas exactement du hadith ou ne connaissant pas le hadith mais connaissant un lafdh de ses alfadh peut retrouver le hadith et toutes les informations nécessaires l'accompagnant grâce à un lafdh, et c'est ce qui fait la puissance et l'importance des alfadh.

Durant des années des traditionnistes (spécialistes de la tradition = sunna) ont mis en œuvre des manuscrits des atrafe el hadith rassemblés de différentes sources généralement, des six livres majeures des ahadiths connu sous le nom de « al kutub-e-sita » et de leurs annexes considérées comme les livres fondamentaux en terme de hadith. Ces livres sont : Sahih al-Bukhari, Sahih Muslim, Sunnan Abî Dawûd, Sunnan al-Nasâ'î, Sunnan at-Tirmidhi, et Sunnan ibn Majah.

Il n'existe pas que ces reccueils. Certains ont rassemblé les hadiths qui leur sont parvenus et les ont classés selon le narrateurs : ce sont les musnad. D'autres ont regroupés uniquement les hadiths liés au droit (fiqh) comme c'est le cas d'Ibn al-Djarûd. D'autres ont regroupés les

Introduction générale

hadiths liés aux engagements militaires du prophète (kutub al-maghâzis...), les qualités du prophètes (kutub al-Shama'il)...

Certains chercheurs considère al kutub-e-sita comme **المراجع الأصلية** car regroupant l'essentiel de la Sunna et le reste comme **المراجع الفرعية** (regorupant le plus souvent des textes dont la substance est déjà présente dans la première catégorie).

3 Modélisation du concept des alfadh el hadith

3.1 Définition :

Les alfadh sont vus comme des phrases ou segments de phrase qui doivent porter une information complète et se distinguent du contexte (de ce qui suit ou de ce qui précède). Ce n'est donc pas un complément d'information mais une information indépendante.

Mais d'abord il faut savoir que les segments de phrase qu'on a appelé alfadh ne sont pas partout dans un hadith ils sont dans la deuxième partie appelé **matn**.

Tableau 2: Exemple d'un sanad et matn d'un hadith

Hadith	
حدثنا الحميدي عبد الله بن الزبير، قال حدثنا سفيان، قال حدثنا يحيى بن سعيد الأنصاري، قال أخبرني محمد بن إبراهيم التيمي، أنه سمع علقمة بن وقاص الليثي، يقول سمعت عمر بن الخطاب رضي الله عنه على المنبر قال سمعت رسول الله صلى الله عليه وسلم يقول " إنما الأعمال بالنيات، وإنما لكل امرئ ما نوى، فمن كانت هجرته إلى دنيا يصيبها أو إلى امرأة ينكحها فهجرته إلى ما هاجر إليه "	
matn	sanad
إنما الأعمال بالنيات، وإنما لكل امرئ ما نوى، فمن كانت هجرته إلى دنيا يصيبها أو إلى امرأة ينكحها فهجرته إلى ما هاجر إليه	حدثنا الحميدي عبد الله بن الزبير، قال حدثنا سفيان، قال حدثنا يحيى بن سعيد الأنصاري، قال أخبرني محمد بن إبراهيم التيمي، أنه سمع علقمة بن وقاص الليثي، يقول سمعت عمر بن الخطاب رضي الله عنه على المنبر قال سمعت رسول الله صلى الله عليه وسلم

Introduction générale

Ce résultat représente le hadith ou le prophète a dit quelque chose, donc pas de dialogue avec une autre personne ou une histoire racontée par un compagnon, donc théoriquement tout le matn est un lafdh ou un ensemble de lafadh.

Tableau 3: Exemple des alfadh, matn et sanad d'un hadith (d'après l'encyclopédie [32])

Hadith	
<p>حدثنا يحيى بن بكير، قال حدثنا الليث، عن عقيل، عن ابن شهاب، عن عروة بن الزبير، عن عائشة أم المؤمنين، أنها قالت أول ما بدئ به رسول الله صلى الله عليه وسلم من الوحي الرؤيا الصالحة في النوم، فكان لا يرى رؤيا إلا جاءت مثل فلق الصبح، ثم حبيب إليه الخلاء، وكان يخلو بغار حراء فيتحنث فيه وهو التعبد الليالي ذوات العدد قبل أن ينزع إلى أهله، ويتزود لذلك، ثم يرجع إلى خديجة، فيتزود لمثلها، حتى جاءه الحق وهو في غار حراء، فجاءه الملك فقال اقرأ. قال " ما أنا بقارئ ". قال " فأخذني فغطني حتى بلغ مني الجهد، ثم أرسلني فقال اقرأ. قلت ما أنا بقارئ. فأخذني فغطني الثانية حتى بلغ مني الجهد، ثم أرسلني فقال اقرأ. فقلت ما أنا بقارئ. فأخذني فغطني الثالثة، ثم أرسلني فقال {اقرأ باسم ربك الذي خلق * خلق الإنسان من علق * اقرأ وربك الأكرم} ". فرجع بها رسول الله صلى الله عليه وسلم يرجف فؤاده، فدخل على خديجة بنت خويلد رضی الله عنها فقال " زملوني زملوني ". فزملوه حتى ذهب عنه الروع، فقال لخديجة وأخبرها الخبر " لقد خشيت على نفسي ". فقالت خديجة كلا والله ما يخزيك الله أبدا، إنك لتصل الرحم، وتحمل الكل، وتكسب المعدوم، وتقري الضيف، وتعين على نوائب الحق. فانطلقت به خديجة حتى أتت به ورقة بن نوفل بن أسد بن عبد العزى ابن عم خديجة وكان امرأ تنصر في الجاهلية، وكان يكتب الكتاب العبراني، فيكتب من الإنجيل بالعبرانية ما شاء الله أن يكتب، وكان شيخا كبيرا قد عمي فقالت له خديجة يا ابن عم اسمع من ابن أخيك. فقال له ورقة يا ابن أخي ماذا ترى فأخبره رسول الله صلى الله عليه وسلم خبر ما رأى. فقال له ورقة هذا الناموس الذي نزل الله على موسى صلى الله عليه وسلم يا ليتني فيها جذعا، ليتني أكون حيا إذ يخرجك قومك. فقال رسول الله صلى الله عليه وسلم " أومخرجي هم ". قال نعم، لم يأت رجل قط بمثل ما جئت به إلا عودي، وإن يدركني يومك أنصرك نصرا مؤزرا. ثم لم ينشب ورقة أن توفي وفتر الوحي</p>	
matn	sanad
<p>أول ما بدئ به رسول الله صلى الله عليه وسلم من الوحي الرؤيا الصالحة في النوم، فكان لا يرى رؤيا إلا جاءت مثل فلق الصبح، ثم حبيب إليه الخلاء، وكان يخلو بغار حراء فيتحنث فيه وهو التعبد الليالي ذوات العدد قبل أن ينزع إلى أهله، ويتزود لذلك، ثم يرجع إلى خديجة، فيتزود لمثلها، حتى جاءه الحق وهو في</p>	<p>حدثنا يحيى بن بكير، قال حدثنا الليث، عن عقيل، عن ابن شهاب، عن عروة بن الزبير، عن عائشة أم المؤمنين أنها قالت</p>

Introduction générale

<p>غار حراء، فجاءه الملك فقال اقرأ. قال " ما أنا بقارئ " . قال " فأخذني فغطني حتى بلغ مني الجهد، ثم أرسلني فقال اقرأ. قلت ما أنا بقارئ. فأخذني فغطني الثانية حتى بلغ مني الجهد، ثم أرسلني فقال اقرأ. فقلت ما أنا بقارئ. فأخذني فغطني الثالثة، ثم أرسلني فقال {اقرأ باسم ربك الذي خلق * خلق الإنسان من علق * اقرأ وربك الأكرم} ". فرجع بها رسول الله صلى الله عليه وسلم يرجف فؤاده، فدخل على خديجة بنت خويلد رضی الله عنها فقال " زملوني زملوني ". فزملوه حتى ذهب عنه الروع، فقال لخديجة وأخبرها الخبر " لقد خشيت على نفسي ". فقالت خديجة كلا والله ما يخزيك الله أبدا، إنك لتصل الرحم، وتحمل الكل، وتكسب المعدوم، وتقري الضيف، وتعين على نوائب الحق. فانطلقت به خديجة حتى أتت به ورقة بن نوفل بن أسد بن عبد العزى ابن عم خديجة وكان امرأ تنصر في الجاهلية، وكان يكتب الكتاب العبراني، فيكتب من الإنجيل بالعبرانية ما شاء الله أن يكتب، وكان شيخا كبيرا قد عمي فقالت له خديجة يا ابن عم اسمع من ابن أخيك. فقال له ورقة يا ابن أخي ماذا ترى فأخبره رسول الله صلى الله عليه وسلم خبر ما رأى. فقال له ورقة هذا الناموس الذي نزل الله على موسى صلى الله عليه وسلم يا ليتني فيها جذعا، ليتني أكون حيا إذ يخرجك قومك. فقال رسول الله صلى الله عليه وسلم " أومخرجي هم ". قال نعم، لم يأت رجل قط بمثل ما جئت به إلا عودي، وإن يدركني يومك أنصرك نصرًا مؤزرًا. ثم لم ينشب ورقة أن توفي وفتر الوحي.</p>	
--	--

alfadh

أومخرجي هم	لقد خشيت على نفسي	زملوني زملوني	فأخذني فغطني حتى بلغ مني الجهد، ثم أرسلني فقال اقرأ. قلت ما أنا بقارئ. فأخذني فغطني الثانية حتى بلغ مني الجهد، ثم أرسلني فقال اقرأ. فقلت ما أنا بقارئ. فأخذني فغطني الثالثة، ثم أرسلني فقال {اقرأ باسم ربك	ما أنا بقارئ
------------	-------------------	---------------	--	--------------

Introduction générale

			الذي خلق * خلق الإنسان من علق * اقرأ وربك الأكرم {	
--	--	--	--	--

Dans le **Tableau 3** nous avons un meilleur exemple représentant les compositions du hadith mais aussi une meilleure vision sur *alfadh el hadith*.

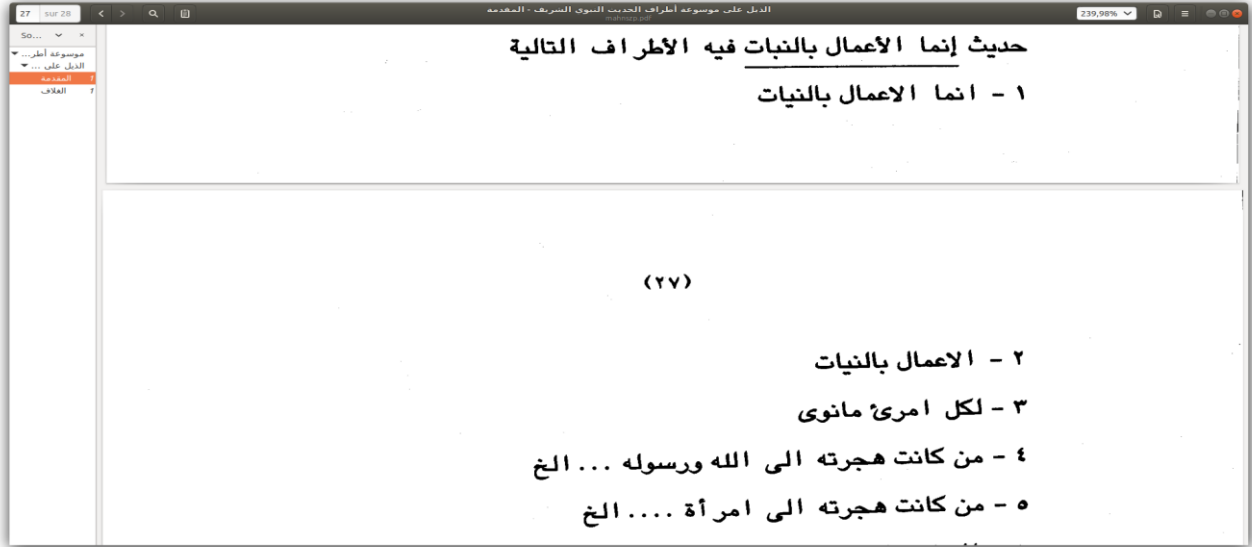


Figure 3: Exemple de *alfadh* tirés de l'annexe de l'encyclopédie [32]

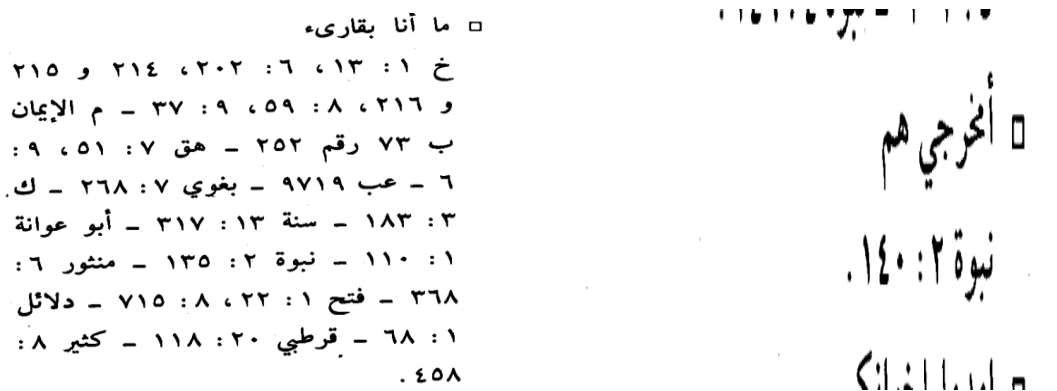


Figure 4: Exemple de *alfadh* tirés de l'encyclopédie [32]

3.2 Extraction des alfadh:

Comme nous l'avons précisé le hadith est composé d'un sanad et d'un matn, qui lui est composé de segments de phrase certains considérés comme alfadh hadith d'autre non, mais le plus important dans notre cas c'est : est-il possible de dresser un algorithme permettant l'extraction des différent composant du hadith ? Et est-il possible d'extraire alfadh el hadith automatiquement ? Et bien la réponse rapide courte est « **Oui** » on peut.

D'abord il faut se rappeler que le hadith est un ensemble de phrases issues de la linguistique arabe donc la première chose à faire est de rassembler, apprendre ou d'obtenir des connaissances en linguistique arabe, c'est d'ailleurs l'élément clé de notre travail.

Nous baserons notre travaille sur les relations entre terme d'abord pour l'identification du sanad, puis par la suite dans l'extraction des alfadh el hadith du matn.

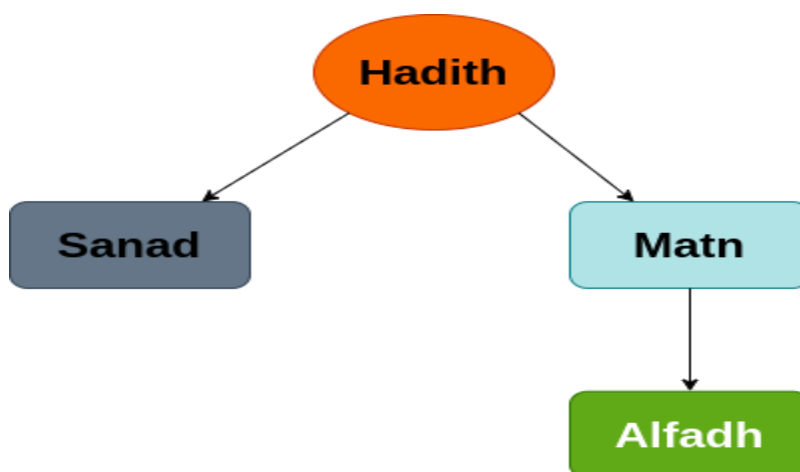


Figure 5: Notre vision d'un hadith

4 Objectifs

Dans ce mémoire, nous allons développer un outil informatique utilisant le traitement automatique du langage pour traiter la langue arabe et extraire au final des alfadh puis les indexer pour, à la fin, créer une base donnée des alfadh el hadith.

Si nous ne disposons pas d'une pareille base, il faudra faire de la recherche « plein texte ». Or, en plus d'avoir l'inconvénient d'être très lente, elle a aussi le tort de nécessiter une correspondance exacte. C'est aussi le cas des moteurs de recherche.

Le travail consiste à séparer automatiquement sanad et matn puis extraire à partir du matn les différents segments porteurs d'information et enfin l'indexation proprement dite.

Chapitre I : ETAT DE L'ART

1 Introduction

Dans ce chapitre nous présenterons les travaux relatifs à notre sujet, nous mettrons le point sur les travaux manuels, comme sur les travaux informatiques.

Nous verrons aussi les critiques concernant ces travaux, dans les deux cas (manuelle et informatique).

2 Présentation des travaux manuelle

Certains auteurs postérieurs à l'âge de la collecte du hadith se sont penché sur la nécessité d'indexer le hadith et ont ainsi écrit des livres servant d'index.

Parmi les livres les plus célèbres des alfadh el hadith on trouve les suivants :

2.1 تحفة الأشراف بمعرفة الأطراف:

Pour l'Imam Abi Al-Hajjaj Yusuf bin Abd Al-Rahman Al-Mizzi, décédé en (742 هـ),

Il a rassemblé alfadh el hadith des six livres dit majeurs, référence, ou plus populairement «**al-kutub-e-sita** » et certains de leurs annexes, parmi eux on trouve :

1. مقدمة صحيح مسلم
2. المراسيل لأبي داود السجستاني
3. العلل الصغير للترمذي
4. الشمائل للترمذي
5. عمل اليوم والليلة للنسائي

Il a indexé chacun de ses livres en leurs donnant à chacun un index spécifique qu'il a dénoté dans l'introduction de son livre, l'indexation se fait par ordre alphabétique des noms des compagnons c'est pour cela qu'on retrouve en premier « أبيض بن حمال »

2.2 إتحاف المهرة بأطراف العشرة:

Pour Al-Hafiz Ibn Hajar Al-Asqalani « ابن حجر العسقلاني » ce que l'on entend par « العشرة » est :

الموطأ	مسند أحمد	مستدرک الحاكم	صحيح بن حبان	مستخرج أبي عوانة	شرح معاني الآثار
--------	-----------	---------------	--------------	------------------	------------------

Chapitre 1 : Etat de l'art

منتقى بن الجارود	مسند الدراي	مسند الشافعي	سنن الدار قطني	صحيح بن خزيمة	
------------------	-------------	--------------	----------------	---------------	--

Le nombre ici est onze. En fait, le Mouwatta' de l'imam Malik a complété une partie manquante de Sahih Ibn Khuzaymah « صحيح ابن خزيمة » dont un quart seulement avait été compilé.

2.3 أطراف المسانيد العشرة

Pour l'imam lashehab adin abi abbasse ahmed ibn abo-bakre el kennani elbousayri achafi'i « لشهاب الدين أبي العباس أحمد بن أبي بكر الكناني البصيري الشافعي » décéder l'année (840 hijri), il entend lui aussi par « المسانيد العشرة » un ensemble de dix masanid c'est-à-dire des livres de ahadith attribué à un rawi qui a vécu l'ère du prophète Mohamed :

مسند إسحاق بن راهويه		مسند مسدد بن مسرهد		
مسند أبي يعلى الموصلي	مسند الحارث بن محمد بن أبي أسامة	مسند عبد بن حميد	مسند أحمد بن منيع	مسند أبي بكر بن أبي شيبة

2.4 ذخائر الموارث في الدلالة على مواضع الحديث:

Pour cheikh abd el ghani naboulsi « عبد الغاني النابلسي » décéder l'année (1143 hijri), il a rassemblé dans son livre les alfadh des ahadith apparue dans les 6 livres majeurs en plus du mowataa « الموطأ », et a utilisé l'ordre de « تحفة الأشراف » comme s'il était son résumé, mais à la différence, il excelle dans l'indexation ou il a remarqué la diversité dans les noms des compagnons (sahaba), et il utilise cela pour la division de son livre en sept sections.

2.5 أطراف المسند المعتلى بأطراف المسند الحنبلي:

Pour l'imam el hafid ibn hajare « ابن حجر », il a créé son livre à partir des ahadiths rassemblés du mousnad el imam ahmed ibn hanbal « أحمد ابن حنبل » à partir de son livre « إتحاف المهرة »

2.6 أطراف الحديث المختارة للضياء المقدسي:

Pour l'imam ibn hajar el Asaklani « ابن حجر العسقلاني »

Chapitre 1 : Etat de l'art

2.7 أطراف مسند الفردوس:

Pour l'imam el Asaklani aussi « ابن حجر العسقلاني »

2.8 أطراف صحيح ابن حبان:

Pour l'imam Abi el fadle el Iraki « أبي الفضل العراقي »

2.9 موسوعة أطراف الحديث النبوي الشريف:

Considérer comme un trésor, elle représente la plus répondu encyclopédie des alfadh el hadith, en effet l'encyclopédie des alfadh el hadith est reconnue comme la révolution en matière des alfadh car elle rassemble plus de cent-cinquante référence d'annuaires de sunna, sira, fikhe et beaucoup d'autres sciences islamique en matière de sunna dans le **Asl** (composé de 11 manuscrit) plus environ deux-cent cinquante références dans l'annexe (composé de quatre manuscrits) ce qui nous fait un total d'environ quatre-cents référence.

Cette dernière regroupe les ahadiths de « المراجع الفرعية » et « المراجع الأصلية » et c'est ce qui fait son importance.

Créer par Cheikh Mohamed Essaid Basoni Zaghlole « الشيخ محمد السعيد بسيوني زغلول » en collaboration avec le professeur Mohamed Ali Baydone « محمد علي بيضون ». Il a lancé le projet de création d'une encyclopédie rassemblant les livres les plus reconnue et les plus importants en Sunna.

La quantité de donnée rassemblée étant énorme, le travail étant ardent l'encyclopédie fut enfin créer en 1996 et devenue alors le trésor dont tout le monde recherchait depuis des millénaires et fut donc la référence de tous amateur, chercheur, connaisseur, ou toutes personne s'intéressant à Sunna.

Le plus remarquable dans ces anciens manuscrits c'est qu'elles ont toutes était dresser manuellement en utilisant des indexations parfois différentes et parfois complémentaire ce qui est pour le moins impressionnant quand on voie la quantité de donnée allant de ressource générale vers les alfadh elles même, comme pour le cas de « تحفة الأشراف بمعرفة » «الأطراف» ou l'écrivain a donné un indexe à chaque livre de ses référence puis a ordonnée son livre par ordre alphabétique des nom des compagnons(Sahaba).

Chapitre 1 : Etat de l'art

« الشيخ محمد السعيد بسيوني زغلول » a annoter une indexation dans son introduction s'inspirant des conceptions des livres précédents, ou on retrouve dans l'introduction page 16-21 les symboles qu'il a donnée à chaque livre dont il s'est inspiré (dénotant à côté du symbole le nom du livre et son édition), puis à ordonner les alfadh ou le début d'un hadith par ordre alphabétique pure.

The screenshot shows a PDF document with a table of contents. The title is 'قائمة بأسماء الكتب والمصنفات التي احتوتها موسوعة الأطراف الأخرى'. The table has three columns: 'الرمز' (Symbol), 'اسم الكتاب' (Book Name), and 'الطبعة' (Edition). The symbols are numbered 1 through 8.

الرمز	اسم الكتاب	الطبعة
١ - تحاف	تحاف السادة المتقين للزبيدي	تصوير بيروت
٢ - تحافات	الإتحافات السنوية	الكليات الأزهرية
٣ - اخبار المنتظر	عقد الدرر	تصوير دار الكتب العلمية
٤ - اخلاق	أخلاق النبوة	النهضة المصرية
٥ - اذكار	الأذكار النووية	عيسى الحلبي
٦ - أذكيا	الأذكيا لابن الجوزي	مؤسسة الرسالة
٧ - اسرار	الأسرار المرفوعة لعلي القاري	المجلس الاعلى للشؤون الإسلامية
٨ - استذكار	الاستذكار لابن عبد البر	

Figure 6: Liste des symboles utilisés dans l'encyclopédie

Il avait commencé par la lettre **Alif Mamdouha** dans le premier tome ou il rassemblée tous les alfadh commençant par la lettre après l'organisation de ces alfadhe était faite selon l'ordre alphabétique de la langue arabe aussi, puis en dessous de chaque lafdh on trouve sa ou ses symboles et références.

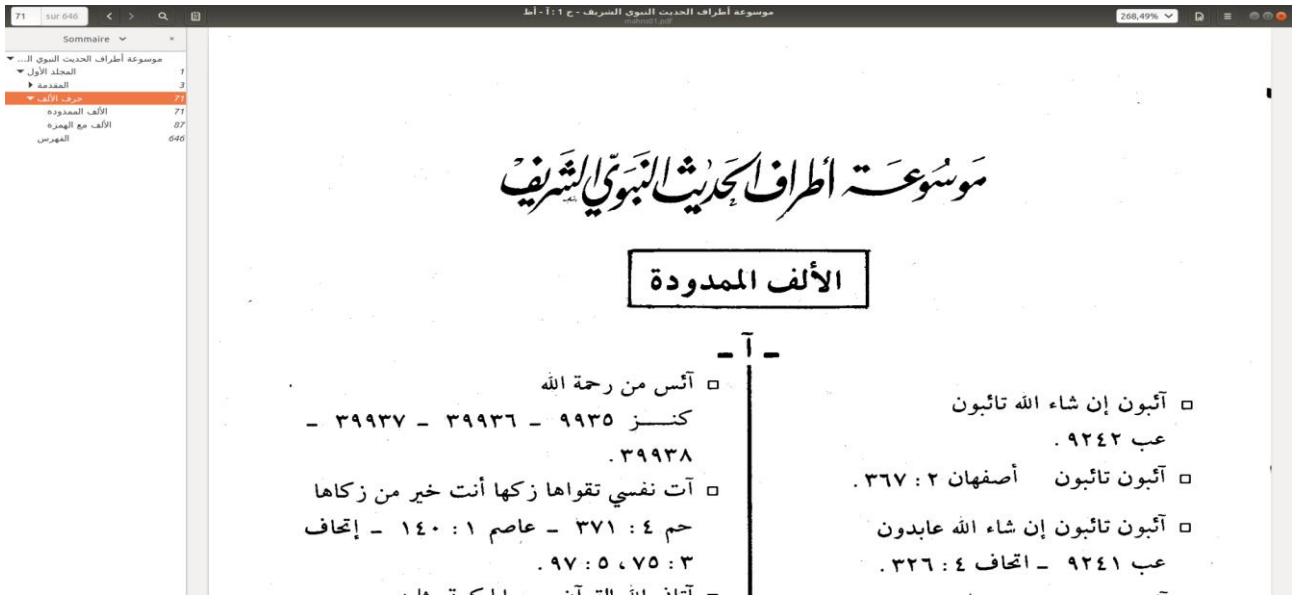


Figure 7: Aperçu de l'organisation de l'encyclopédie Alif mamdouha

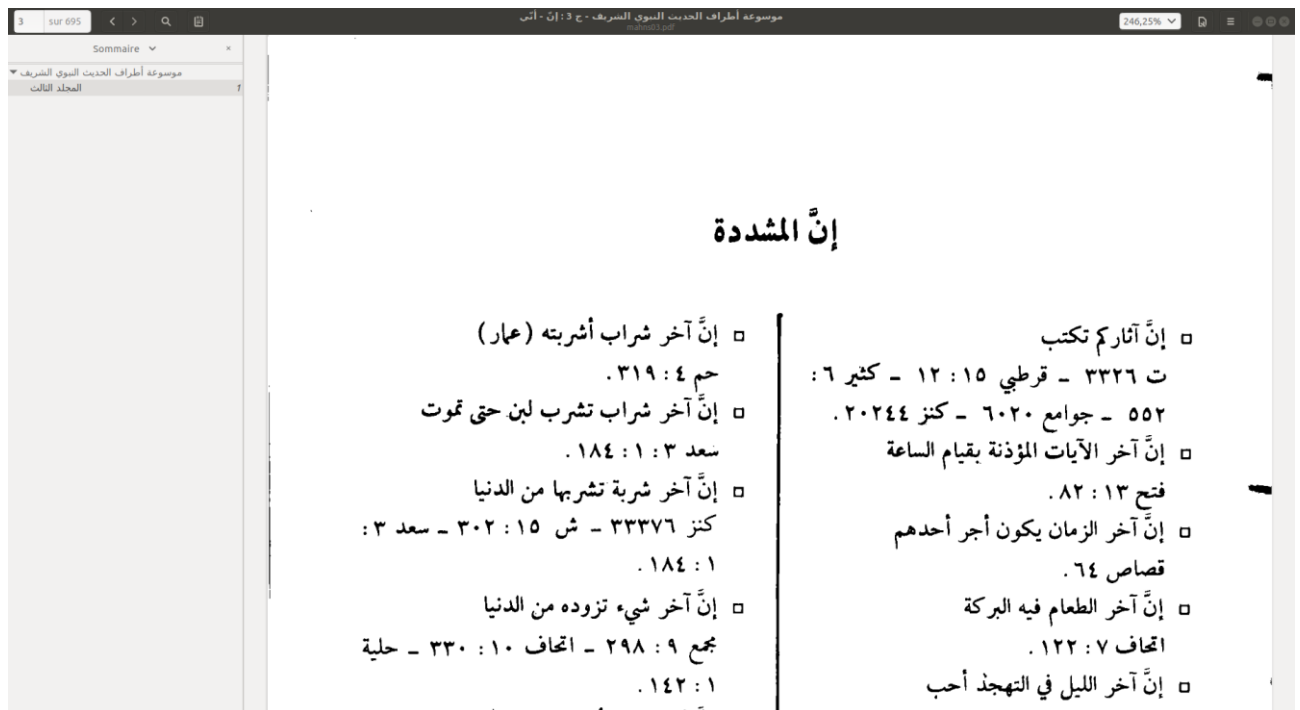


Figure 8: Aperçu de l'organisation de l'encyclopédie Inna

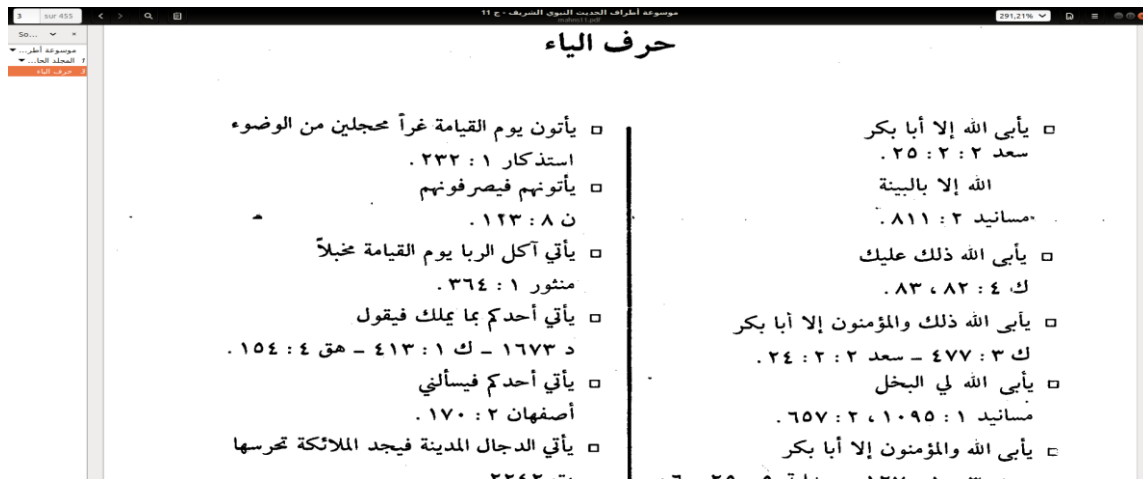


Figure 9: Aperçu de l'organisation de l'encyclopédie Yaa

Les figure présenter au-dessus représente un bref aperçu de l'encyclopédie et de sa puissance en termes d'indexation aussi en termes de classification, et malgré sa ces figures ne représente que le **Asl** c'est-à-dire les cent-cinquante première référence tirés des livres, le reste on le trouvera dans l'**annexe**, dans les figures qui suivent nous auront un bref aperçu de sa aussi

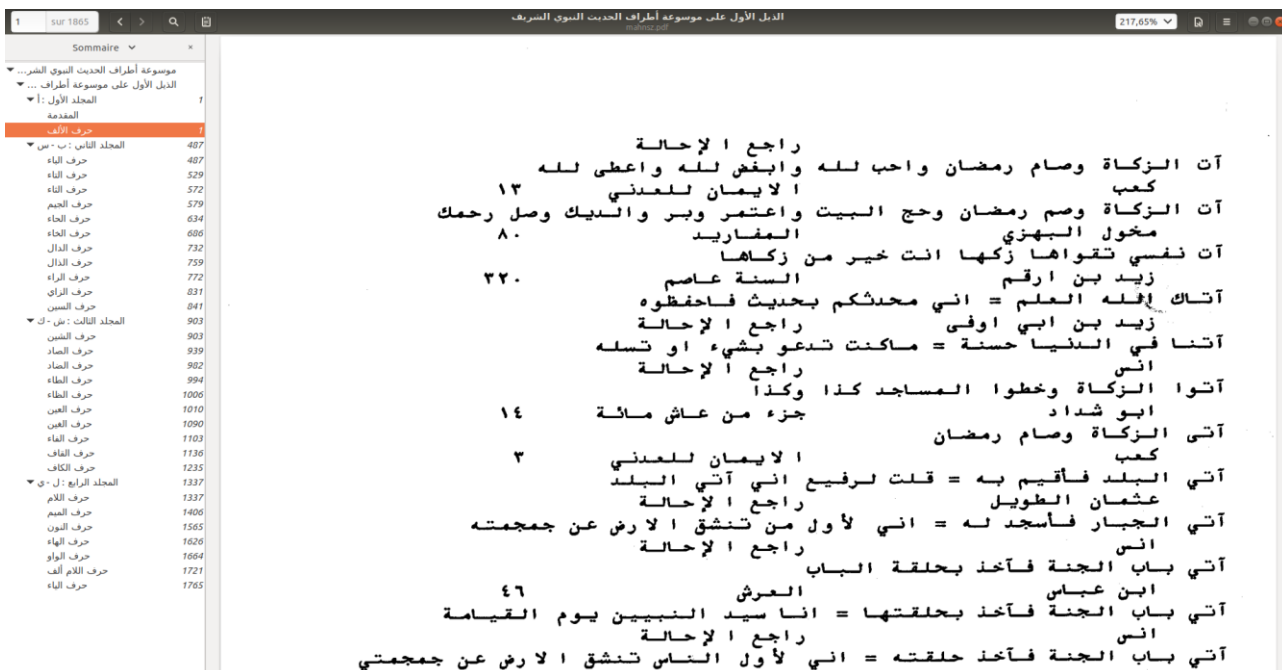


Figure 10: Aperçu de l'organisation de l'encyclopédie dans l'annexe

Chapitre 1 : Etat de l'art

L'importance de l'encyclopédie étant énorme, elle en a au moins certains défauts que les experts le lui ont reproché, à commencer par le fait qu'un hadith apparaissant dans « المراجع الفرعية » et apparaissant aussi dans « المراجع الأصلية » ne devrait pas être indexer plusieurs fois, un index de « المراجع الأصلية » suffit.

3 Limites des travaux manuels

Ces livres considérés comme référence durant des millénaires, ont permis une transmission correcte, intégrée et fiable des ahadith du prophète Mohamed, créant ainsi diverses sciences en Sunna conduisant à une créativité en matière d'indexation et de présentation des ahadith. Ceci étant dit, ces livres ne sont plus la première référence à utiliser pour le grand public (voire peut-être même les spécialistes du domaine), ils penchent plutôt vers les moteurs de recherches, due au temps important de recherche dans ces livres (qui peut prendre des heures, voire des jours), en plus de l'obligation de connaître exactement les mots du hadith. Les moteurs de recherche ne sont pas sans faille, ils présentent certains défauts comme les livres, il est obligatoire de connaître exactement les mots du hadith, dans le cas contraire on se retrouve avec des résultats faiblement pertinents.

4 Présentation des travaux informatiques

La recherche dans le traitement automatique de la langue arabe a longtemps été liée à deux sources de donnée, le Coran et le Hadith mais les plus grandes avancées et les plus grandes recherches et expérimentations, ont été faites sur le Coran en laissant à côté la Sunna (Hadith) et ce du fait de la non disponibilité des corpus de donnée contenant les Ahadith. Heureux pour nous avons réussi à trouver une source mettant à disposition un dataset bien structuré réunissant tous les ahadiths des Sita-e-Sihah que nous allons voir plus loin dans les prochains chapitres.

Malgré ces difficultés on retrouve quelques recherches dans le monde liées au Sunna en l'occurrence le traitement des ahadith certain en Anglais et d'autre en Arabe, mais ce qui les réunis ce n'est qu'aucun d'eux n'avait comme objectif de mettre en place un système d'extraction des ahadith à partir des matn, et certain n'utilise pas une approche liée au traitement automatique du langage comme nous l'auront espéré.

Le hadith étant composé de deux aspects majeurs **sanad** et le **matn** représentant le contenu du hadith, ces deux aspects ont été pour les chercheurs suivant une source d'inspiration :

Aldahlan et Al (Aldhlan, K.A, Zeki, AM, 2010): a présenté sa compréhension pour la collection des connaissances islamique de différentes ressources, son objectif était de présenter les données dans une meilleure façon possible pour cela il utilise les ahadiths comme donnée d'apprentissage puis il a utilisé un algorithme supervisé pour faire la classification des ahadiths.

Harrag (F. Harrag,, 2014) : a présenté un travail basé sur la reconnaissance des noms en l'occurrence les noms propres, puis extraction des entités des ahadiths.

Muazzam Siddiqui (Siddiqui, Muazzam Ahmed, Mostafa El-Sayed Saleh, Abobakr Ahmed Bagais, 2014) : a proposé un système d'extraction du sanad ou sanad (comme il l'a appelé) en utilisant une approche basée sur la reconnaissance et l'extraction des noms propres puis une classification dans une sorte de réseau (il a utilisé les approches de la machine Learning pour faire une classification basée sur les noms du sanad et les noms à l'intérieur du matn).

Mahmoud A : a proposé un Framework pour l'extraction des données de Sahih el Bukhari.

Chapitre 1 : Etat de l'art

Ahsan Mahmoud, Fawaz K Alafraj, Hikmat Ulah Khan, Muhammed Rawzane, Mahwish Ilyes (Ahsan Mahmood, Fawaz K. Alarfaj, Hikmat Ullah Khan*, Muhammad Ramzan, Mahwish Ilyas, 2018): ces chercheurs ont travaillé sur un projet de traitement de ahadiths en l'an 2018, leur travail a été publié par le journal 'International Journal Of Advanced Computer Science and Applications', leurs travaux ont porté sur un dataset des ahadith en Multilingue ou il ont utilisé le dataset des ahadiths du **cheikh El Albani**, l'objectif est de faire une extraction matn du hadith pour ensuite faire un moteur de recherches des mutun, d'après eux les algorithmes de classification actuelle ne prennent pas en charge entièrement la langue arabe ou ne donnent pas de résultat satisfaisant en vue de sa structure morphologique et orthographique très compliquée, ils ont donc opté pour un algorithme régulier leur permettant l'extraction du sanad et matn de chaque hadith non d'une structure linguistique naturelle mais plutôt des sites web.

Leurs plans de travail sont schématisés dans les figures : et

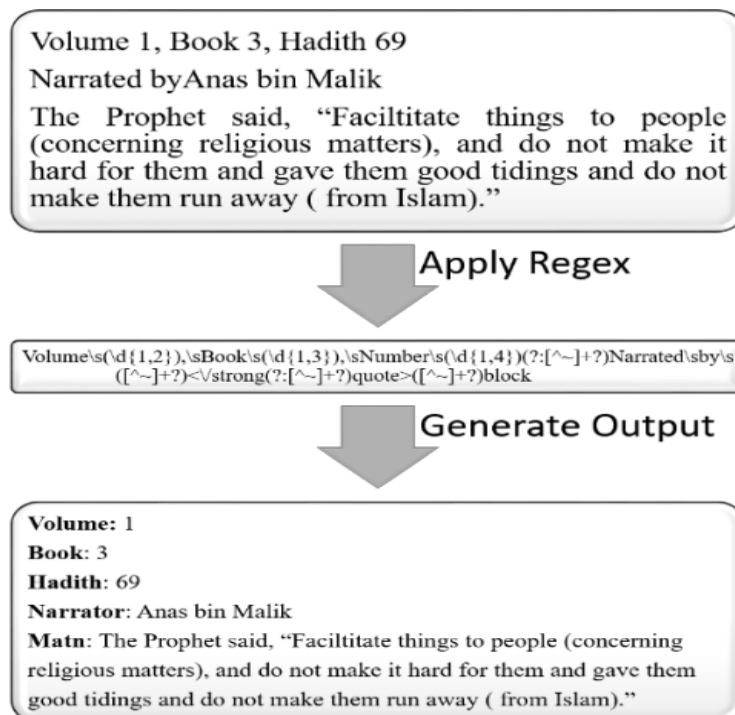


Figure 11: application de l'algorithme se basant sur l'HTML

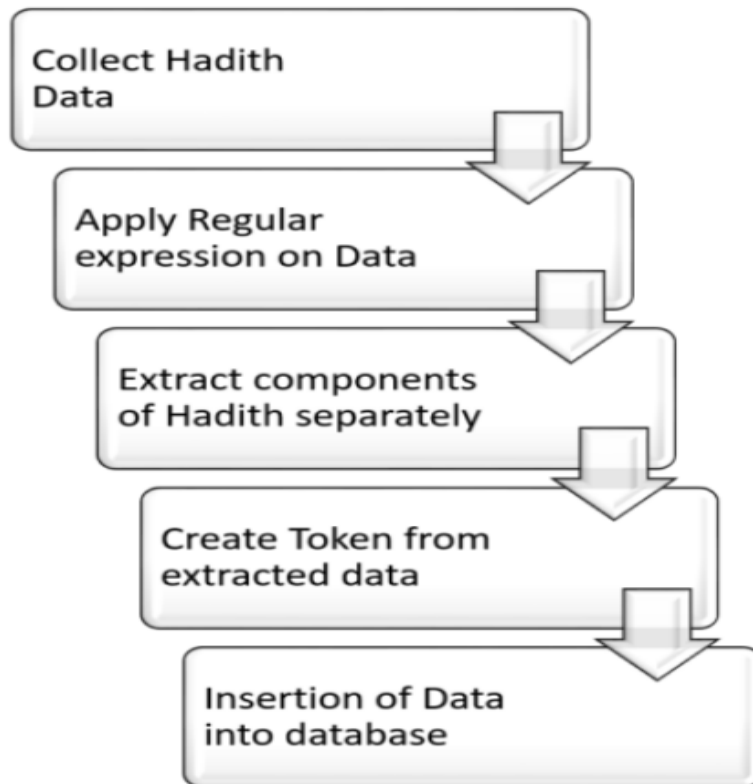


Figure 12: conception du travail se basant sur l'HTML

Maraoui et al. : Ils ont préparé une liste des mots qui viennent avant, après et entre les noms des narrateurs, puis ont préparé une liste de mots signant la fin de chaque hadith, avec sa ils sont parvenus à extraire le Matn du sanad, mais leurs algorithmes n'ont pas été testés sur un dataset de **ALBUKHARI**, donc il n'a pas été testé sur d'autres datasets pour observer sa réaction vis-à-vis.

Boella (Boella, 2011) : Ils ont présenté un outil d'extraction du matn en utilisant un algorithme régulier, ceci basé sur une liste de verbe et mot de transmission du sanad, La liste ne regroupe pas de mots littérale mais en des symboles identifiants des mots, puis ils ont transformé tous les mots du hadith selon la liste pour pouvoir identifier le sanad, leurs résultats n'étant pas très pertinents ils ont dressé une liste de cent caractères recueilli à partir des essais et erreurs passant leur permettant de vérifier si dans les prochains cent caractères il existe un mot de transmission sinon c'est le début du matn.

Azmi et bin Badia : Ont mis en terme un système permettant de créer l'arbre des narrateurs à partir de ahadiths, mais en premier lieu il fallait identifier puis extraire le sanad, pour cela ils sont passés par la voie du pré-processing ou pré-traitement en normalisant les ahadiths ceci

Chapitre 1 : Etat de l'art

en enlevant les ponctuations et les diacritise. Après le pré-traitement ils ont appliqué le concept d'Algorithme Libre (Free Grammar) pour identifier les segments du hadith se basant sur la comparaison des tokens à la liste lexicon qu'ils avaient compilé avant. Comme le but de leur étude était la création de l'arbre des narrateurs leurs résultats reflètes la réussite de leur système de création de l'arbre et non pas la segmentation.

Autre que la reconnaissance et extraction des noms d'autres techniques ont été présentées en l'occurrence la méthode de la prise en charge de donnée introuvable durant l'extraction des données du hadith nommée Missing Data Detector (**MDD**), l'auteur a proposé une méthode basé sur les méthodes de validité du sanad (sanad) dans la science du hadith. **Un autre outil** a été proposé utilisant les bases du Model de l'Espace Vectorielle pour permettre à l'utilisateur de rechercher un hadith particulier dans la base de données entière mais avec une meilleure précision, l'auteur a développé un système de classification des ahadiths basé sur le système de similarité entre eux.

Shatha Altammami, Eric Atwel, Ammar Alsalka : on propose un système se basant sur les n-gramme pour segmenter les matn et sanads, ils ne savaient pas qu'elle hadith était authentique ou non dans leur dataset, ils ont alors opté pour un choix de 40 hadith des six livres référence formant alors un dataset de 240 hadith dont ils sont qu'ils sont authentique, leur méthode a était utiliser dans l'extraction sanad et matn passant par un pré-traitement suivie d'une annotation des tokens pour à la fin retrouver les segments, ils ont schématiser leur travail dans la figure en dessous : **Figure 13** .

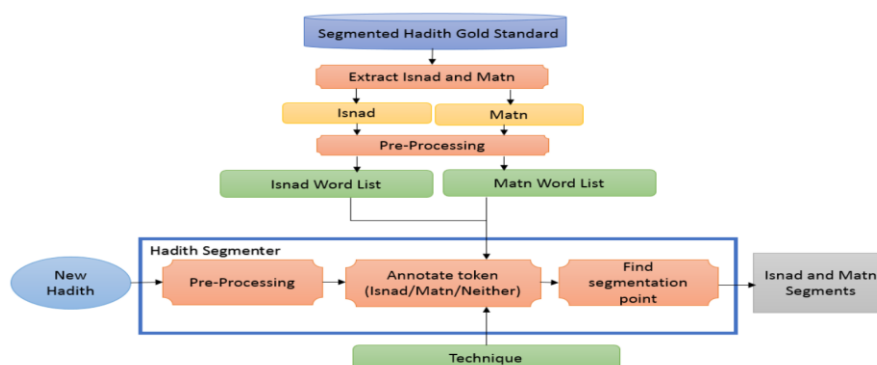


Figure 13: conceptions du travail des n-gramme

Chapitre 1 : Etat de l'art

Dans le **Tableau 4** nous mettons un résumer de ces travaux, pour faciliter l'analyse.

Tableau 4: Revue des Précédente Recherches sur la Segmentation du Hadith

Travail	Approches	Techniques	Pre-Traitement	Annotation Manuelle	Données	Résultat {Précision}
Muazzam Sidiqi (2014)	Machine Learning	Nave Base, Knn, arbre de décision	Enlever les diacritises, Stemming	Personne, Narrateur, autre	Albukhari, Musnad Ahmed	80%
Harrag (2014)	FST Finite State Transducers	-	La Tokenisation	-	Albukhari	44%
Azmi et Bin Badia (2010)	Basic	Context Free Grammar	Enlever les diacritises, les ponctuation Stemming	Hadith Lexicon	Albukhari	87%
Maraoui et al. (2018)	Basic		-	Hadith Lexicon	Albukhari	96%
Mahmood et al. (2018)	Basic	Expression Régulière	-	-	Albukhari Muslim, Abu Dawud, Imam Malik	98%
Boella (2011)	Basic	Expression Régulière	Transformation des mots selon	Hadith Lexicon	Albukhari	97%

Chapitre 1 : Etat de l'art

			une liste			
Adahlan et Al	Machine Learning	Knn	-	-	Albukhari	-
Shatha Altamma-mi, Eric Atwel, Ammar Alsalka	Basic	Segmentation, n-gramme, Annotation	Tokenisation	-	Sahih Albukhari, Sahih Muslim, Sunan Abu Dawood, Sunan Altarmithi, and Sunan Ibn Maja	92.5%

5 Limites des travaux informatiques :

Les précédentes recherches étaient toutes basées sur trois approches, l'approche Basic qui consistait à mettre des listes de mot ou autre pour identifier les noms du sanad, puis d'autres listes pour identifier le matn, la seconde approche quant à elle consiste à utiliser la machine learning utilisant des données annotées, algorithmes de sélection, entraînement et classification, la troisième approche consiste à utiliser le FST qui dépend du degré de consistance dans un texte bien structuré.

En revenant au tableau 2, on remarque que les approches Basic avaient de meilleurs résultats mais le problème est que ces approches ont été utilisées et testées sur un seul dataset représentant une certaine variété de livre, donc on ne peut pas garantir que le résultat sera le même avec d'autres datasets de d'autres livres, en réalité Mahmood et al. (2018) a créé une expression régulière pour différents livres de hadith ce qui démontre qu'une approche basic comme l'expression régulière sur tous si elle est créée à partir d'une source de données spécifique ne peut pas être universelle, de l'autre côté l'approche Machine Learning a besoin de données d'entraînement qui représentent une grande variété de hadith mais elle a aussi montré son échec au niveau de la performance car l'étude Muazzam Sadiqui (2014) a montré une chute de 10% dès que le modèle a été testé avec un dataset d'un autre livre de hadith, autre problème cette fois avec FST qui exige une uniformité dans la ponctuation du livre de hadith pour qu'il donne un minimum de résultat sinon la segmentation ne marchera pas, et donc on peut clairement déduire que ce modèle ne sera pas performant de le cas de changement de dataset.

Aussi dans la précédente Tableau on a aussi donné une mesure de précision des modèles (du moins les mesures qu'on était publié), mais il faut clarifier que ces approches ne sont pas comparables pour deux raisons, la première est que les données de test utilisées pour chaque étude sont différentes par rapport au type et à la taille des données, la seconde est que chaque étude a sa propre façon de mesurer sa performance, par exemple Muazzam Sadiqui a mesuré la précision de l'habileté du système à annoter les noms des personnes comme narrateur ou nom ceci que ce soit un nom du sanad ou matn, malgré que leur objectif n'est pas la segmentation mais l'identification des narrateurs qui peut être utilisée dans la segmentation.

Chapitre 1 : Etat de l'art

Ces études représente un travail remarquable et notre objectif n'est pas très loin des leurs, en réalité il les dépasse un peu car leur travailles a tous consistait à extraire le matn du sanad puis faire une classification ou autre chose dans le genre par auteur par exemple ou par livre .., dans notre cas nous allons aller plus loin en présentant un travail regroupant les six livre référence Esita-e-Sihah et extraire le Mtan du sunna el kawlia pour ensuite essayer de le segmenter selon les normes d'un lafdh hadith .

6 Conclusion

Dans ce chapitre nous avons présenté les travaux existant relative à notre sujet, ces travaux avaient chacun ses points fort et ses défaut, des défauts que nous essayerons de prendre en compte pour améliorer notre travail.

Chapitre II : Conception de l'outil d'extraction

1. Introduction

Dans ce chapitre nous allons commencer à expliquer la conception de notre travail. Les grammaires linguistiques utilisées et les observations pris en comptes. Bien sûr nous essayerons de schématiser tout ça. Mais avant tout ça un petit rappel sur la linguistique arabe s'impose.

2 Rappel des notions de la linguistique arabe

2.1 L'analyse des mots

La langue arabe a une morphologie très riche et différente par rapport aux langues occidentales. L'analyse morphologique d'un mot arabe, consiste principalement à déterminer la structure générale de ce mot, s'il existe, et les autres éléments utilisés pour construire ce mot (les affixes, les modèles).

2.1.1 Le shème

L'Arabe est une langue sémitique dont la morphologie fonctionne sur le croisement des racines (généralement trilitères) avec un nombre de formes déterminées pour produire sa gamme de vocabulaire. Ces formes, appelées moules ou schèmes, sont essentiellement constituées par

La langue arabe a une morphologie très riche et différente par rapport aux langues occidentales. L'analyse morphologique d'un mot arabe, consiste principalement à déterminer la structure générale de ce mot, s'il existe, et les autres éléments utilisés pour construire ce mot (les affixes, les modèles).

2.1.2 La racine

Une racine est purement consonantique, elle est formée par une suite de trois ou quatre (ou même cinq pour les noms) consonnes formant la base du mot. La racine est un élément important dans les langues dérivationnelles. En effet, à chaque racine correspond un champ sémantique et à l'aide de différents schèmes, on peut générer une famille de mots appartenant à ce champ sémantique.

Chapitre 2 : Conception de l'outil d'extraction

2.1.3 Les affixes

Les affixes sont des lettres qui s'ajoutent au début (les préfixes) ou à la fin des mots arabes (les suffixes). En général, ils sont utilisés pour accorder aux mots des éléments syntaxiques. Ils marquent l'aspect verbal, le mode, les propriétés transitives, etc. Leur nombre tourne aux alentours de **150** affixes.

2.2 Les Phrases

Dans la langue arabe on retrouve deux types de phrase :

2.2.1 Phrase Verbale :

Les phrases verbales sont les phrases commençant par un verbe, par quelque l'une de ses formes (passé, présent, impératif). En générale une phrase verbale est composée d'un verbe-sujet-complément, dans le cas d'un verbe transitif qui a besoin d'un complément pour exprimer son sens, par contre dans le cas d'un verbe non transitif, et bien la phrase se forme par un verbe-sujet seulement.

2.2.2 Phrase nominale :

Les phrases nominales sont les phrases où la partie nominative est présente, et sa structure se base sur deux parties : le sujet et le prédicat.

2.3 El Mouaaraf et Nakira (المعرف ب ال و النكرة):

La langue arabe comme les autres langues, présente cet aspect, on peut le voir dans la langue anglaise comme « The » et dans la langue française comme « un (e) ». Inversement les entités dénuées de « ال » est dite « نكرة ».

2.4 Jar et Majrur (الجار والمجرور):

Le jar et le Majrur sont appelés en arabe une demi-phrase « شبه جملة », et la raison de ce nom est que le jar et le Majrur ne mènent pas à une signification indépendante dans le discours, mais plutôt le sens qu'ils exécutent est une sous-phrase. Verbale ou nominale.

Le jar : sont un ensemble de particule {في'على'من'إلى'...}

Chapitre 2 : Conception de l'outil d'extraction

El majrur : Quant au nom majrur, c'est un pronom ou un pronom continu avec lequel une préposition est introduite et s'y joint.

2.5 El idafa (الإضافة) :

El idafa dans la langue arabe indique l'imposition d'un rapport entre deux mots, en l'occurrence el modaf (مضاف) et modaf ilihe (مضاف إليه) :

2.5.1 El modaf (مضاف)

El modaf est chaque nom ajouté à un autre nom, car le premier mène au second, et le modaf ilihe y'est appelé.

2.5.2 El modaf ilihe (مضاف إليه)

El modaf ilihe est tout mot lui attribuant quelque chose par le biais d'un jar.

2.5 Charte (أسلوب الشرط)

Le style subjonctif est un style (clause conditionnelle) qui désigne l'intersection et la corrélation de deux phrases via un outil appelé l'outil conditionnel. De sorte que toute la réponse conditionnelle ne peut être remplie que si la phrase condition verbale est remplie. Autrement dit, l'occurrence et l'accomplissement de la phrase de la réponse sont conditionnels - dans la plupart des cas - de l'occurrence de la condition entière, et si la condition est remplie, la réponse est remplie.

La méthode de condition se compose de: l'outil de condition + la phrase de verbe de condition + la phrase de réponse de condition.

2.6 Tachbihe (تشبيه):

Une analogie est une comparaison entre deux parties ou deux choses qui partagent la même caractéristique et que l'une d'elles dépasse l'autre dans cette caractéristique, en utilisant une analogie.

3 Conception générale :

Notre travail portant sur la Sunna, la première chose à utiliser était la collection des ahadiths (sans ça on ne peut rien faire) puis viens l'enchaînement d'applications de l'outil d'extraction du matn puis des alfadh.

L'outil d'extraction comme nous l'avons dit est une composition d'outils, le premier pour l'extraction du matn, le deuxième pour l'extraction des alfadh.

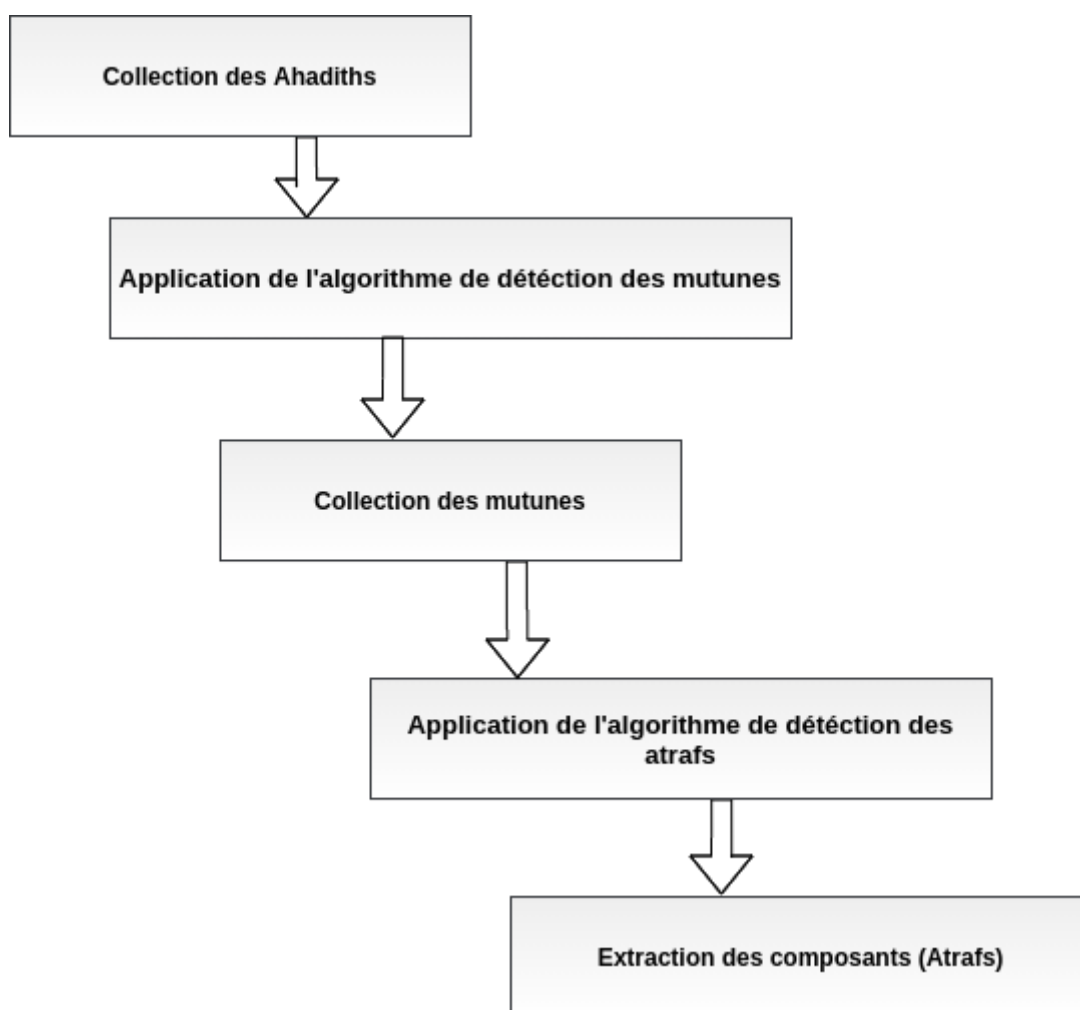


Figure 14: Conception de l'outil d'extraction

Chapitre 2 : Conception de l'outil d'extraction

3.1 Extraction des matn (Séparation sanad et matn):

3.1.1 Introduction :

Sanad : Le sanad représente l'enchaînement des personnes transmettant le hadith a travers le temps, c'est d'ailleurs l'une des conditions de fiabilité du hadith s'il est authentique ou non, de ce fait on peut dire que la transmission se fait par une variation de vocabulaire pas très grande ce qui nous a permis de cerner les mots de transmission.

Ces mots sont en général soit suivis d'un nom soit suivis d'autres mots représentant le début du matn.

matn : Le matn représente le contenu du hadith, nous y reviendront dessus dans les prochaines étapes.

L'algorithme d'extraction est basé sur la reconnaissance des noms propres du sanad. Une fois le sanad identifié, on considère le reste comme un matn, mais qu'elles sont les conditions à satisfaire ? Et bien il y a plusieurs façons d'extraire les noms propres, la plus probable serait de reconnaître le nom lui-même mais vu qu'on va extraire à partir d'un sanad on remarque qu'on a les mots qui définissent les garants suivant comme des noms propres (que nous appellerons transmetteurs) donc on peut connaître les noms propre grâce à cette approche.

Nous présenterons un exemple de ces mots dans le tableau suivant :

Tableau 5: Mots de transmission

حرف الجر	عن
حرف النداء	يا. ياأيها
أداة مصدرية	أن
الأفعال	قال. حدّث. أخبر. ذكر. سمع

Bien sûr nous prenons en compte leurs conjugaisons dans notre travail.

Chapitre 2 : Conception de l'outil d'extraction

En langue arabe les mots précédant le nom propre ne sont pas forcément des verbes comme : « حدثنا الأعمش », ils peuvent aussi être des particules comme « عن عائشة أم المؤمنين ». Ces mots introduisent obligatoirement un nom propre, mais comme toutes règles appartenant à la linguistique arabe, il y a des exceptions à ces règle-là, pourquoi ? Parce-que dans certains cas les mots transmetteurs réapparaissent successivement puis viens le nom propre comme par exemple : « قال قال رسول الله صلى الله عليه وسلم » ou dans d'autre cas l'apparition d'un autre transmetteur par exemple : « قال عن عمر رضي الله عنه ».

Nous allons traiter ces cas-là.

3.1.2 'an (عن):

Pour commencer, nous prendrons d'abord l'exemple de « عن » que nous traiterons :

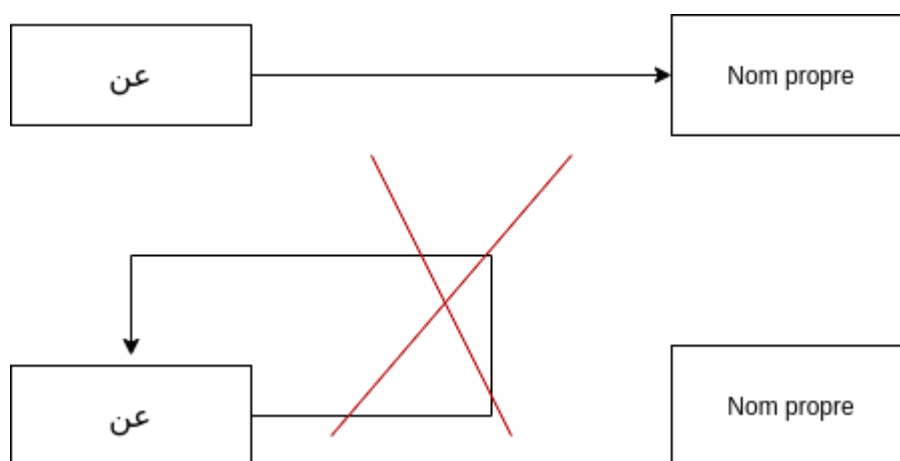


Figure 15:Aperçu de la particule 'an

Dans une phrase, la particule 'an définit obligatoirement son successeur comme un nom propre, nous verront en dessous un exemple de sanad pour le démontrer.

حدثنا يحيى بن بكير، قال حدثنا الليث، عن عقيل، عن ابن شهاب، عن عروة بن الزبير، عن عائشة أم المؤمنين

Le cas de la particule 'an est pratiquement le même pour harf nidaa qui exige lui aussi que son successeur soit un nom propre.

3.1.3 Harf nidaa (حرف نداء):



Figure 16: Aperçu du nidaa

A la différence de 'an, harfe nidaa peut en effet se passer du nom propre en le remplaçant par le surnom de l'intéressé, ceci en rajoutant potentiellement la particule « أياها » après la particule « يا ».

<p>عن أبي العباس عبد الله بن عباس رضي الله عنهما قال: كنت خلف النبي صلى الله عليه وسلم يوماً، فقال: يا غلام، إني أعلمك كلمات: احفظ الله يحفظك، احفظ الله تجده تجاهك، إذا سألت فاسأل الله، وإذا استعنت فاستعن بالله، واعلم أن الأمة لو اجتمعت على أن ينفعوك بشيء، لم ينفعوك إلا بشيء قد كتبه الله لك، وإن اجتمعوا على أن يضروك بشيء، لم يضروك إلا بشيء قد كتبه الله عليك، رفعت الأقلام وجفت الصحف</p>
<p>عن عبد الله بن عمر قال: قال رسول الله -صلى الله عليه وسلم:- يا أيها الناس، توبوا إلى الله، فإني أتوب في اليوم إليه مئة مرة</p>
<p>عبد الله بن مغفل فإنه قال: قال رجل: يا رسول الله، إني لأحبك، فقال صلى الله عليه وسلم: "انظر ما تقول" قال: والله إني لأحبك، ثلاث مرات، فقال عليه الصلاة والسلام: "إن كنت تحبني، فأعد للفقر تجفافاً، فإن الفقر أسرع إلى من يحبني من السيل إلى منتهاه"</p>
<p>عن معاذ بن جبل . رضي الله عنه . أن رسول الله . صلى الله عليه وسلم . أخذ بيده وقال: يا معاذ !والله إني لأحبك، أوصيك يا معاذ لا تدعني في دبر كل صلاة تقول: اللهم أعني على ذكرك وشكرك وحسن عبادتك</p>

3.1.4 Anna (أن):

L'autre cas des particules est celui de « أن » qui introduit un nom propre directement après son apparition.

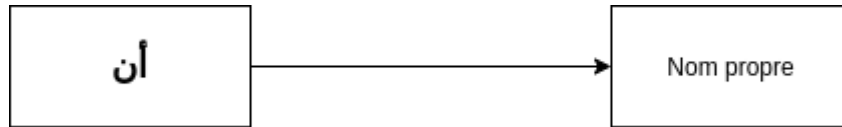


Figure 17: Aperçu de Anna cas simple

حدثنا أبو اليمان الحكم بن نافع، قال أخبرنا شعيب، عن الزهري، قال أخبرني عبيد الله بن عبد الله بن عتبة بن مسعود، أن عبد الله بن عباس، أخبره أن أبا سفيان بن حرب أخبره أن هرقل

Mais « أن » possède des variantes c'est-à-dire des dérivés qui prennent en compte ce qu'on appelle en langue arabe « ضمير متصل » représentant le précédent locuteur. Ces dérivés introduisent des comportements différents de leur origine.

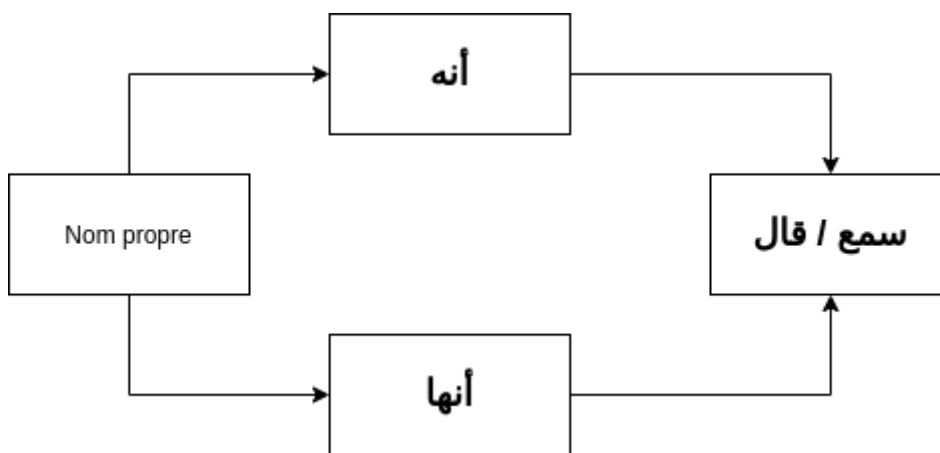


Figure 18: Aperçu de Anna cas complexe

حدثنا يحيى بن بكير، قال حدثنا الليث، عن عقيل، عن ابن شهاب، عن عروة بن الزبير، عن عائشة أم المؤمنين، أنها قالت

Chapitre 2 : Conception de l'outil d'extraction

حدثنا الحميدي عبد الله بن الزبير، قال حدثنا سفيان، قال حدثنا يحيى بن سعيد الأنصاري، قال أخبرني محمد بن إبراهيم التيمي، أنه سمع علقمة بن وقاص الليثي، يقول سمعت عمر بن الخطاب رضي الله عنه على المنبر قال سمعت رسول الله صلى الله عليه وسلم

3.1.5 Verbes :

Le dernier cas c'est celui des verbes qui ont un caractère bien différent des particules.

Certains verbes peuvent apparaître successivement ou faire appel à un autre verbe de la liste, ou voir appeler harfe nidaa ou 'an comme successeur.

3.1.5.1 Cas simple :

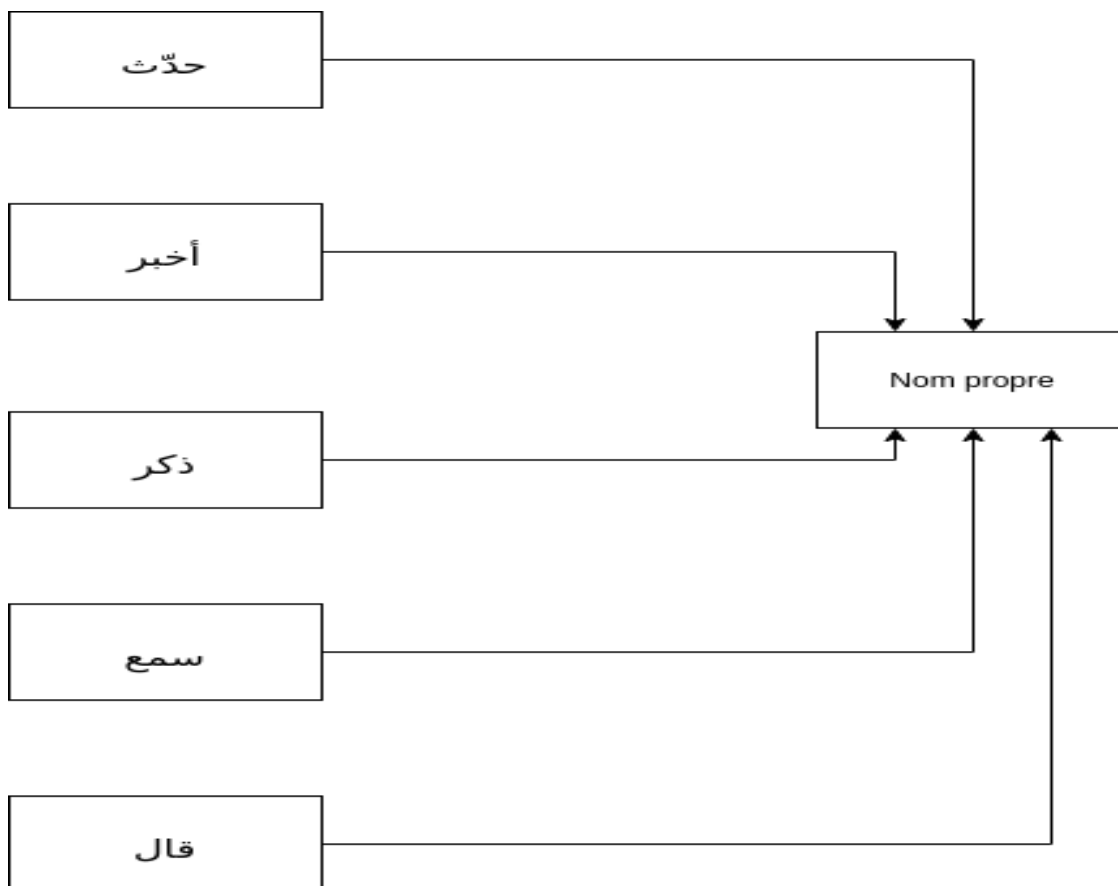


Figure 19: Aperçu du cas simple des verbes

Chapitre 2 : Conception de l'outil d'extraction

حدثنا الحميدي عبد الله بن الزبير، قال حدثنا سفيان، قال حدثنا يحيى بن سعيد الأنصاري، قال أخبرني محمد بن إبراهيم التيمي، أنه سمع علقمة بن وقاص الليثي، يقول سمعت عمر بن الخطاب رضى الله عنه على المنبر قال سمعت رسول الله صلى الله عليه وسلم يقول

La **Figure 19** représente le cas simple des verbes, c'est-à-dire un verbe transmetteur suivis d'un nom propre, mais comme nous l'avons dit en haut, certains verbes n'appliquent pas toujours la loi simple en l'occurrence « قال » qui prend des formes différentes dans une phrase, c'est-à-dire qu'il va introduire soit le discours de son précédent, soit il représente l'introduction du locuteur, soit il fait appel aux autres verbes ou particules.

3.1.5.2 Cas complexe :

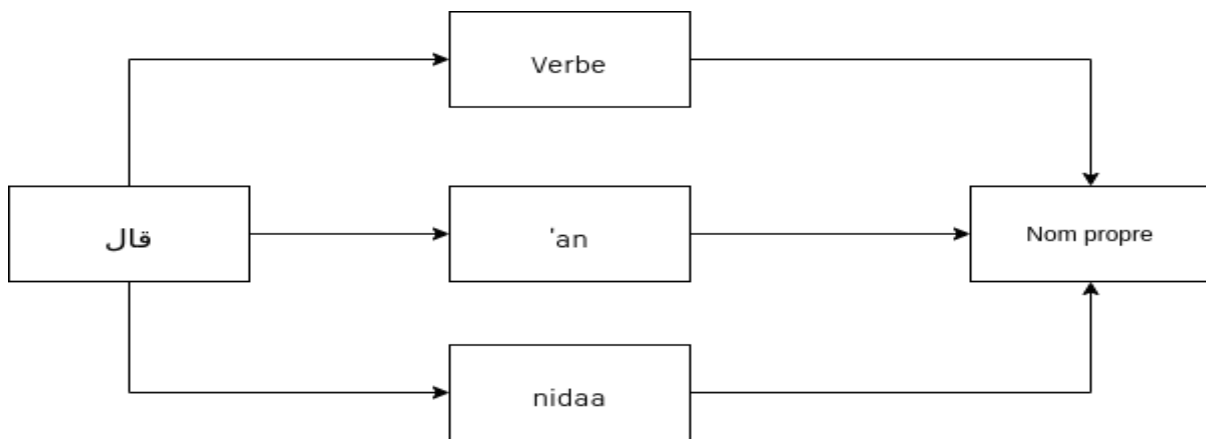


Figure 20: Aperçu du 1er cas complexe des verbes

حدثنا الحميدي عبد الله بن الزبير، قال حدثنا سفيان، قال حدثنا يحيى بن سعيد الأنصاري، قال أخبرني محمد بن إبراهيم التيمي، أنه سمع علقمة بن وقاص الليثي، يقول سمعت عمر بن الخطاب رضى الله عنه على المنبر قال سمعت رسول الله صلى الله عليه وسلم يقول

Nous remarquons dans l'exemple au-dessus que le verbe « قال » fait des appels aux autres verbes et particule pour qu'ils introduisent un nom propre.

Chapitre 2 : Conception de l'outil d'extraction

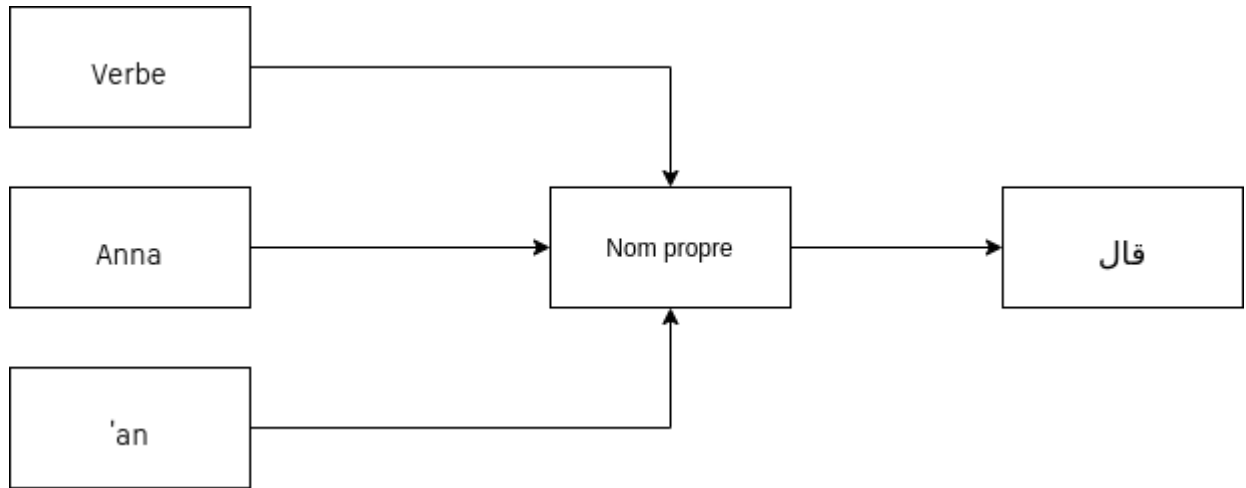


Figure 21: Aperçu du 2eme cas complexe des verbes

حدثنا عبد الله بن محمد، قال حدثنا أبو عامر العقدي، قال حدثنا سليمان بن بلال، عن عبد الله بن دينار، عن أبي صالح، عن أبي هريرة رضى الله عنه عن النبي صلى الله عليه وسلم قال

Nous remarquons dans l'exemple au-dessus que le verbe « قال » est appelé après l'apparition du nom propre qui est lui-même précédé d'un verbe transmetteur.

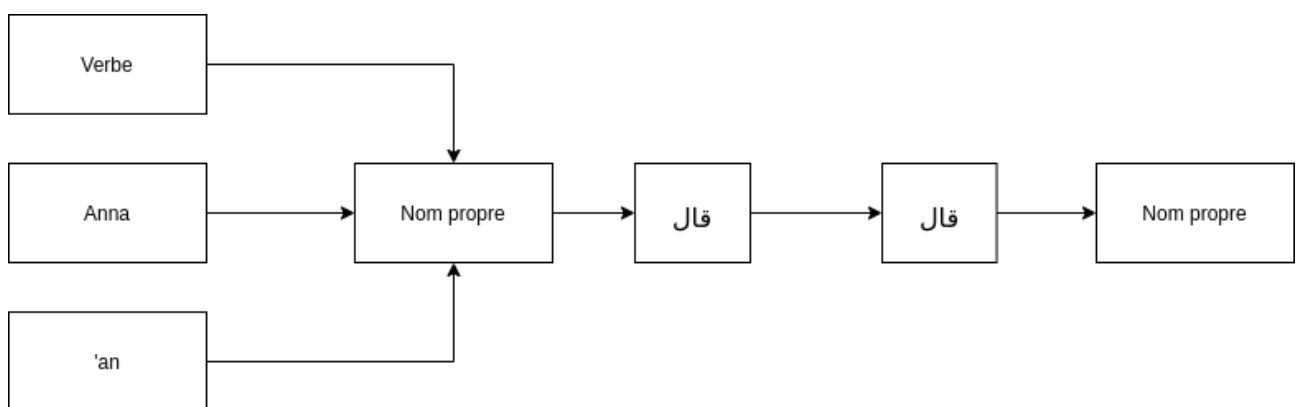


Figure 22: Aperçu du 3eme cas complexe des verbes

حدثنا يعقوب بن إبراهيم، قال حدثنا ابن علية، عن عبد العزيز بن صهيب، عن أنس، عن النبي صلى الله عليه وسلم وحدثنا آدم، قال حدثنا شعبة، عن قتادة، عن أنس، قال قال النبي صلى الله عليه وسلم

Chapitre 2 : Conception de l'outil d'extraction

L'exemple au-dessus démontre que le verbe « قال » peut faire un appel récursif a lui-même pour introduire par la suite un nom propre, pourquoi ? Est bien le premier « قال » introduit le discours de son locuteur précédent quant au deuxième « قال » et bien il introduit un nom propre et donc un nouveau locuteur.

3.1.6 Résumé de la première opération

On peut résumer se premier algorithme dans la figure qui suit (Figure 23) :

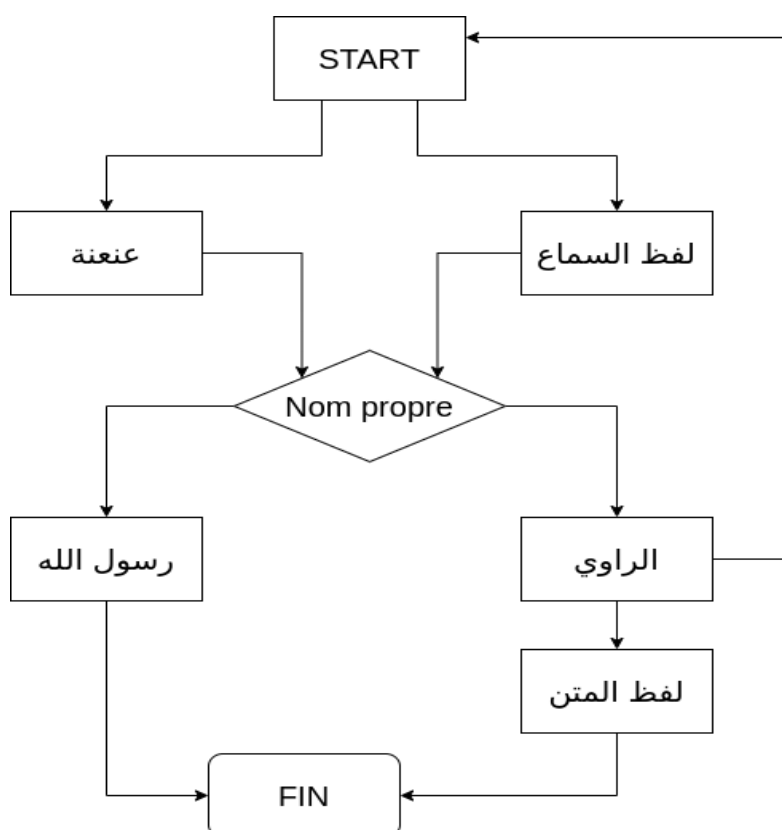


Figure 23: Aperçu général de l'algorithme de séparation

3.1.7 Déroulement

Pour éviter toute ambiguïté, nous allons procéder à un déroulement de l’algorithme :

Hadith
حدثنا قتيبة، حدثنا عبد العزيز بن محمد، عن العلاء بن عبد الرحمن، عن أبيه، عن أبي هريرة، أن رسول الله صلى الله عليه وسلم قال " ما نقصت صدقة من مال وما زاد الله رجلا بعفو إلا عزا وما تواضع أحد لله إلا رفعه الله "

Matn	Sanad
صلى الله عليه وسلم قال " ما نقصت صدقة من مال وما زاد الله رجلا بعفو إلا عزا وما تواضع أحد لله إلا رفعه الله "	(1) قتيبة (2) عبد العزيز بن محمد (3) العلاء بن عبد الرحمن (4) أبيه (5) أبي هريرة (6) رسول الله

Dans l’exemple au-dessus, nous remarquons que le sanad représente une liste de nom des transmetteurs du hadith dénué des locutions de transmission « أَلْفَاظُ الرِّوَايَةِ ». Pourquoi ? Et bien parce que durant nos recherches, nous avons cherché les noms propres en langue arabe afin d’augmenter la pertinence de notre travail. Nous avons aussi une autre approche basée sur les noms propres. Elle n’a pas pu voir le jour car il fallait un dataset des noms propres. Malheureusement nous n’avons pas trouvé un pareil dataset, donc nous avons opté pour cette méthode pour qu’elle soit utilisée dans le futur sur plusieurs critères de projet pas forcément relative à la religion islamique. Avec la méthode expliquée plus haut, le matn peut être précédé de quelque mot qui n’en fait pas partie. Nous traiterons cela lors de l’extraction des alfadh.

Cette erreur-là, nous l’avons mise exprès pour pouvoir gérer certaines exceptions qui sont apparues lors de nos précédents tests (car dans certains cas, le sanad prend quelque substance du matn est le contraire est vrai)

Chapitre 2 : Conception de l'outil d'extraction

Mais comment l'algorithme obtient-il ce résultat ? Pour faire simple nous allons prendre l'exemple au-dessus est faire un déroulement de l'algorithme :

Tableau 6: Déroulement de l'algorithme de séparation (1^{er} exemple)

i	Token du hadith	Transmetteur	Sanad
i=0	حدثنا قتيبة حدثنا عبد العزيز بن محمد عن العلاء بن عبد الرحمن عن أبيه عن أبي هريرة أن رسول الله صلى الله عليه وسلم قال ما نقصت صدقة من مال وما زاد الله رجلا بعفو إلا عزا وما تواضع أحد لله إلا رفعه الله	Vrai	قتيبة
i=2	حدثنا عبد العزيز بن محمد عن العلاء بن عبد الرحمن عن أبيه عن أبي هريرة أن رسول الله صلى الله عليه وسلم قال ما نقصت صدقة من مال وما زاد الله رجلا بعفو إلا عزا وما تواضع أحد لله إلا رفعه الله	Vrai	قتيبة عبد العزيز بن محمد
i=7	عن العلاء بن عبد الرحمن عن أبيه عن أبي هريرة أن رسول الله صلى الله عليه وسلم قال ما نقصت صدقة من مال وما زاد الله رجلا بعفو إلا عزا وما تواضع أحد لله إلا رفعه الله	Vrai	قتيبة عبد العزيز بن محمد العلاء بن عبد الرحمن
i=12	عن أبيه عن أبي هريرة أن رسول الله صلى الله عليه وسلم قال ما نقصت صدقة من مال وما زاد الله رجلا بعفو إلا عزا وما تواضع أحد لله إلا رفعه الله	Vrai	قتيبة عبد العزيز بن محمد العلاء بن عبد الرحمن أبيه
i=14	عن أبي هريرة أن رسول الله صلى الله عليه وسلم قال ما نقصت صدقة من مال وما زاد الله رجلا بعفو إلا عزا وما تواضع أحد لله إلا رفعه الله	Vrai	قتيبة عبد العزيز بن محمد العلاء بن عبد الرحمن أبيه أبي هريرة
i=17	أن رسول الله صلى الله عليه وسلم قال ما نقصت صدقة من مال وما زاد الله رجلا بعفو إلا عزا وما تواضع أحد لله إلا رفعه الله	Vrai	قتيبة عبد العزيز بن محمد العلاء بن عبد الرحمن أبيه

Chapitre 2 : Conception de l'outil d'extraction

			أبي هريرة رسول الله
--	--	--	------------------------

Une fois le nom du prophète détecté dans la liste du sanad, on considère que le sanad est terminé.

Voici un deuxième exemple qui devrait éclaircir le traitement :

Hadith
حدثنا عمرو بن خالد، قال حدثنا الليث، عن يزيد، عن أبي الخير، عن عبد الله بن عمرو رضى الله عنهما أن رجلا، سأل النبي صلى الله عليه وسلم أى الإسلام خير قال " تطعم الطعام، وتقرأ السلام على من عرفت ومن لم تعرف ".

Matn	Sanad
سأل النبي صلى الله عليه وسلم أى الإسلام خير قال " تطعم الطعام وتقرأ السلام على من عرفت ومن لم تعرف "	عمرو بن خالد الليث يزيد أبي الخير عبد الله بن عمرو رضى الله عنهما رجلا

Tableau 7: Déroulement de l'algorithme de séparation (2ème exemple)

i	Token du Hadith	Transmetteur	Sanad
i=0	حدثنا عمرو بن خالد قال حدثنا الليث عن يزيد عن أبي الخير عن عبد الله بن عمرو رضى الله عنهما أن رجلا سأل النبي صلى الله عليه وسلم أى الإسلام خير قال تطعم الطعام وتقرأ السلام على من عرفت ومن لم تعرف	Vrai	عمرو بن خالد
i=4 i=5	قال حدثنا الليث عن يزيد عن أبي الخير عن عبد الله بن عمرو رضى الله عنهما أن رجلا سأل النبي صلى الله عليه وسلم أى الإسلام خير قال تطعم الطعام وتقرأ السلام على من عرفت ومن لم تعرف	Vrai	عمرو بن خالد الليث

Chapitre 2 : Conception de l'outil d'extraction

i=7	عن يزيد عن أبي الخير عن عبد الله بن عمرو رضى الله عنهما أن رجلا سأل النبي صلى الله عليه وسلم أى الإسلام خير قال تطعم الطعام وتقرأ السلام على من عرفت ومن لم تعرف	Vrai	عمرو بن خالد الليث يزيد عن أبي الخير
i=12	عن عبد الله بن عمرو رضى الله عنهما أن رجلا سأل النبي صلى الله عليه وسلم أى الإسلام خير قال تطعم الطعام وتقرأ السلام على من عرفت ومن لم تعرف	Vrai	عمرو بن خالد الليث يزيد عن أبي الخير عبد الله بن عمرو رضى الله عنهما
i=20	أن رجلا سأل النبي صلى الله عليه وسلم أى الإسلام خير قال تطعم الطعام وتقرأ السلام على من عرفت ومن لم تعرف	Vrai	عمرو بن خالد الليث يزيد عن أبي الخير عبد الله بن عمرو رضى الله عنهما رجلا
i=22	سأل النبي صلى الله عليه وسلم أى الإسلام خير قال تطعم الطعام وتقرأ السلام على من عرفت ومن لم تعرف	Faux	

Dans l'exemple au-dessus nous remarquons que dès la détection d'un verbe interrogatif « سأل », l'algorithme comprend que nous sommes arrivés à l'extrémité du sanad et donc le reste est un matn.

Chapitre 2 : Conception de l'outil d'extraction

3.1.8 Indexation matn et sanad

Séparer le matn du sanad dans les ahadiths se fait en gardant le `hadith_id` du dataset dans chaque classe, en plus du matn obtenu dans la classe `matn` et du sanad dans la classe `sanad`. Bien-sûr, vu que les sanad et le matn sont les deux composantes d'un hadith, la relation `sanad-dataset` et `matn-dataset` est une relation d'appartenance ou un matn appartient à un et un seul hadith, et un sanad appartient à un et un seul hadith, de l'autre côté un hadith contient un et un seul matn et un hadith contient un et un seul sanad.

3.1.9 Conclusion

L'algorithme de séparation sanad-matn est le premier pas à franchir avant de pouvoir passer à l'extraction des alfadh, nous l'avons appliqué sur le dataset complet ce qui veut dire sur les six-kutub (34441 Hadith).

Nous verrons par la suite l'algorithme d'extraction des alfadh.

3.2 Extraction des alfadh (Création des alfadh):

3.1.1 Introduction

Il est très difficile de cerner tous les critères qui identifient un lafdh. Nous avons donc dû procéder de façon empirique c'est-à-dire consulté des livres de hadith et essayer de dégager les formes les plus récurrentes. Partant de cela, nous avons constaté les cas récurrents de début de matn suivants qui ne sont certainement pas les seuls :

Après obtention de nos matn, nous entameront la création des alfadh el hadith en se basant sur un algorithme créé à partir des lois de la linguistique arabe.

Tout matn est composé d'un ou plusieurs alfadh, cette relation est en effet l'objet de notre travail qui nous permettra par la suite de remonter dans la hiérarchie pour faire l'indexation, mais nous reviendront sur cela plus loin dans le mémoire.

L'algorithme développé pour l'extraction des alfadh a lui aussi été développé progressivement et adopte les lois de la linguistique arabe. Si nous pouvions résumer l'idée globale, ça serait que les alfadh -en l'occurrence segment des matn- doivent représenter une information complète.

Notre algorithme a pris en compte certains aspects linguistiques dans le cadre de création des alfadh. Il est basé sur deux aspects principaux **Global** et **Indépendant**. Dans la langue arabe, les particules ont certaines conditions d'utilisation, parmi les conditions que nous avons prises en compte vis-à-vis des particules, c'est le commencement d'une phrase. Si c'est Oui, alors il est indépendant, si c'est non, alors il sera lié à son prédécesseur.

Chapitre 2 : Conception de l'outil d'extraction

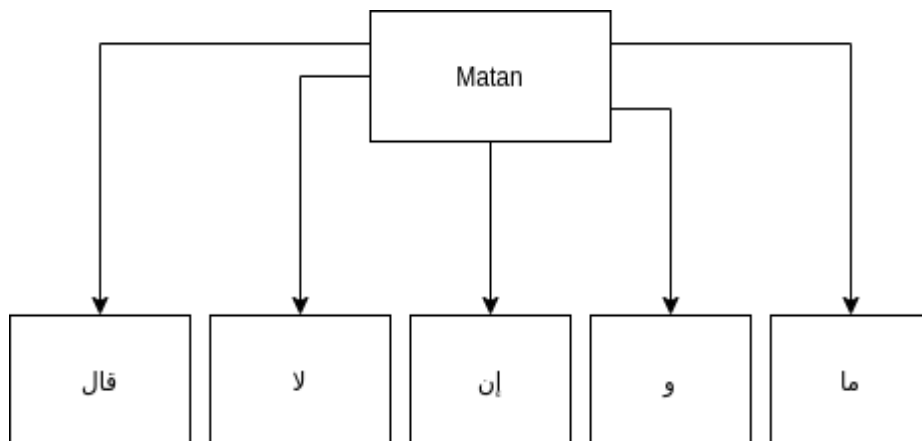


Figure 24: Aperçu des possibilités de commencement du matn

Tableau 8: Liste des particules pris en compte

حرف عطف	أو.ثم.الواو.
ضمير قصد	فوالذي.فالذي.فذلك.فذلكم
أداة نهي	لا
أداة مصدرية/ شرطية	ما
أداة توكيد	فإن.إن.
أداة شرطية	من. فمن.حتى.
حرف ناسخ	كأن
أداة إستثناء	إلا
لغة الخطاب	قال

Notre travail ayant pris les conditions nécessaires pour l'extraction des alfadh dans la meilleure forme linguistique possible, nous verrons par la suite comment procéder :

Chapitre 2 : Conception de l'outil d'extraction

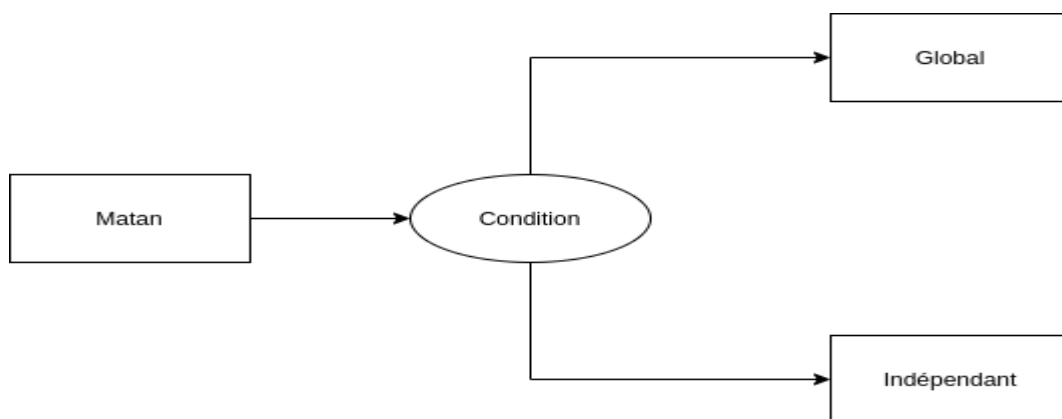


Figure 25: Aperçu des types de condition de alfadh

Si nous devons résumer ce que nous avons dit en haut, ça serait :

- Boucler sur les token du matn
- Évaluer les conditions posées
- Si condition vraie alors extraction du lafdh
 - Si condition d'indépendance vrai alors c'est bon
 - Sinon on crée/rattache le/au lafdh Global
- Sinon création lafdh globale

Les particules ne permettant pas l'indépendance d'un lafdh son «إلا.لا», ces deux particules sont en effet incapables de commencer une phrase du moins pour إلا, par contre لا viens toujours en début de phrase, et s'il y a réapparition, et bien il sera précédé d'un حرف عطف, et donc pas besoin de le prendre en condition seul.

Les alfadh indépendants peuvent contenir d'autres particules indépendantes selon leurs positions dans le matn, par la suite nous allons voir cela :

3.1.2 Waw(واو) :

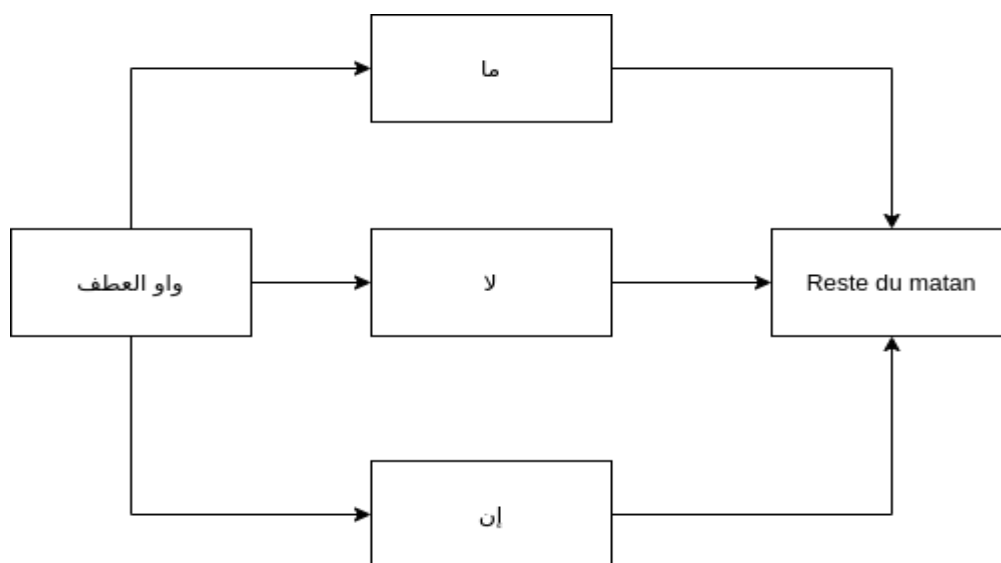


Figure 26:Aperçu des conditions de succession du waw

Vu comme ça, on peut dire que ça fait le bonheur d'une personne de voir ça, mais ce n'est pas le cas malheureusement, pourquoi ? Et bien c'est simplement dû au fait que الواو est la particule la plus compliquée à prendre en compte et c'est pas anodin. La langue arabe est connue pour être une langue très riche et très puissante sur tous les niveaux linguistiques, et ça devient logique qu'on voit qu'elle introduit le concept de polyvalence dit « كلمة وظيفية », Qu'est-ce que c'est ? Et bien la polyvalence c'est la capacité d'un mot ou une particule à prendre plusieurs sens et plusieurs valeurs grammaticales dans les phrases.

Nous verrons par la suite ces différentes raisons et aussi ses différentes variantes selon le contexte :

Tableau 9: Liste des différentes apparitions du waw

واو القسم (والله. وعزتي. وجلالي)
واو المعية (سرت والشاطيء)
واو العطف (حضر محمد وأخوه)

Chapitre 2 : Conception de l'outil d'extraction

حرف زائد للتأكيد (ما من أحد إلا وله مطمع)
حرف استئناف أو واو الابتداء (على الحاكم ألا يجور ويعدل)
واو الحال (جاء والشمس طالعة)
واو لصوق جملة الصفة بموصوفها (وعسى أن تكرهوا شيئاً وهو خير لكم)

Le tableau au-dessus ne représente qu'un échantillon de واو, démontrant la polyvalence dans la langue arabe. Ce phénomène n'est pas limité à seulement الواو mais pratiquement toutes les particules citées en haut.

Dans notre travail, nous avons pris en compte les cas que nous voyons possible à traiter vue la difficulté de distinction.

L'un des cas que nous avons rencontré est lorsque le lafdh contient plusieurs واو العطف. Ceci a causé des problèmes au niveau de l'algorithme qui le voyait comme un nouveau lafdh à chaque fois mais nous sommes parvenus à en tirer une sorte de loi, car, en général, quand cela arrive et bien le واو العطف est précédé par حرف جر par exemple ou حرف ناسخ (كأن) donc nous avons pris certaines mesures dans le cas où il est précédé par حرف جر et اسم مجرور.

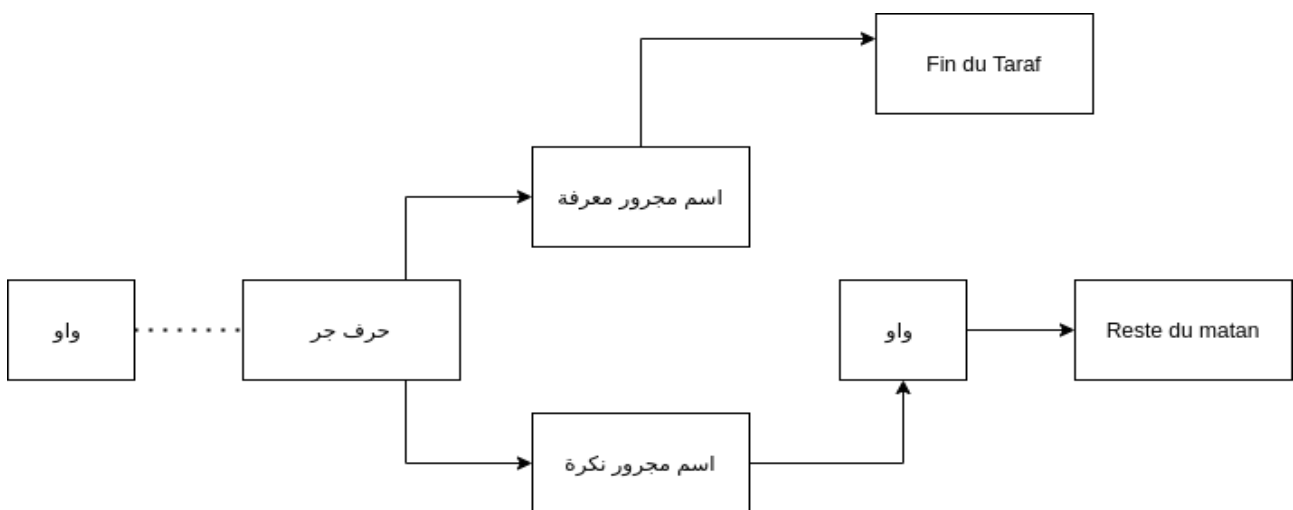


Figure 27: Aperçu des cas de traitement du waw

Chapitre 2 : Conception de l'outil d'extraction

En effet quand « اسم مجرور معرف » est introduit, il peut être suivi par un « اسم معرف », et devient donc un « اسم مجرور مضاف » tandis que le deuxième mot est un « مضاف إليه », mais dans le cas où « واو العطف » est introduit après « اسم مجرور معرف » on peut dire que « اسم مجرور » est satisfait, par contre si « واو العطف » est introduit après « اسم مجرور نكرة » et bien « اسم مجرور » devient « معطوف عليه » et le mot précédant « واو العطف » est un « معطوف », impliquant donc la deuxième partie à être reliée à son prédécesseur.

ألا أخبركم بما يمحو الله به الخطايا ويرفع به الدرجات إسباغ الوضوء **على المكاره** وكثرة الخطا إلى المساجد وانتظار الصلاة بعد الصلاة فذلكم الرباط فذلكم الرباط فذلكم الرباط

المسلم من سلم المسلمون **من لسانه ويده** والمهاجر من هجر ما نهى الله عنه

Par contre dans le cas de واو القسم nous l'avons traité en le mettant comme exception et donc passés directement au suivant.

3.1.3 Aw et Thoma (أو، ثم):

Deux autres حروف عطف qui sont ثم et أو que nous avons traité comme des particules qui peuvent commencer une phrase, et bien dans la linguistique aucun d'eux ne peut commencer une phrase. Mais vu le contexte, il représente une valeur à part dans les alfadh et comme واو العطف et bien ils sont tout à fait capables d'être suivis par une autre particule.

انتدب الله لمن خرج في سبيله لا يخرجه إلا إيمان بي وتصديق برسلي أن أرجعه بما نال من أجر أو غنيمة، أو أدخله الجنة، ولولا أن أشق على أمتي ما قعدت خلف سرية، ولوددت أني أقتل في سبيل الله **ثم** أحياء، **ثم** أقتل **ثم** أحياء، **ثم** أقتل

انتدب الله لمن خرج في سبيله لا يخرجه إلا إيمان بي وتصديق برسلي أن أرجعه بما نال من أجر **أو** غنيمة، **أو** أدخله الجنة، ولولا أن أشق على أمتي ما قعدت خلف سرية، ولوددت أني أقتل في سبيل الله **ثم** أحياء، **ثم** أقتل **ثم** أحياء، **ثم** أقتل

Chapitre 2 : Conception de l'outil d'extraction

Ceci était la représentation de **واو** dans notre algorithme d'extraction des **alfadh**, nous verrons par la suite les reste.

3.1.4 Illa (إلا):

Parmi les particules on trouve أداة إستثناء **إلا** qui est **إلا**, cette particule ne peut pas débiter une phrase et sera donc lié à ce qui le précède, donc dans notre vision on ne pouvait pas le mettre comme élément d'exportation d'un **lafdh**.

ما نقصت صدقة من مال وما زاد الله عبدا بعفو إلا عزا وما تواضع أحد لله إلا رفعه الله

3.1.5 Hatta (حتى) :

Autres particules prises en compte أداة شرطية, cette particule ne peut pas commencer une phrase elle aussi elle a besoin d'un début qu'on appelle الشرط et d'une réponse qu'on **جواب الشرط**.

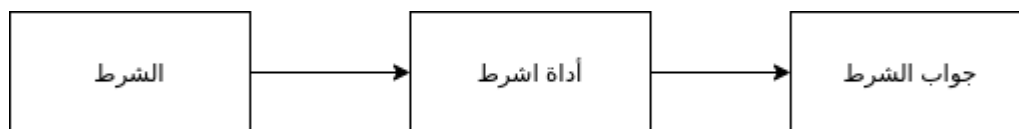


Figure 28: Aperçu des cas de charte

Mais dans notre cas, on ne pas traiter linguistiquement, on a préféré voir la valeur apporté pour un **lafdh**, et on a donc préféré les séparer et du coup chaque partie représentera un **lafdh** à part.

لا تقوم الساعة حتى تقاتلوا قوما صغار الأعين عراض الوجوه

جواب الشرط peut représenter une phrase, un paragraphe voir un texte entier.

3.1.6 Kaana (كأن):

Une autre entité que nous avons prise en compte c'est كأن. Ce mot est une combinaison de deux, le premier ك qui dans la langue est introduit au début d'une particule comme أن ou الذي. Il symbolise l'image ou la métaphore de ce qui le précède. Cette particule ne peut pas débiter une phrase, mais dans l'extraction des alfadh, il introduit une grande valeur sémantique au propos, donc nous l'avons traité comme s'il pouvait commencer une phrase et donc comme un lafdh à part.

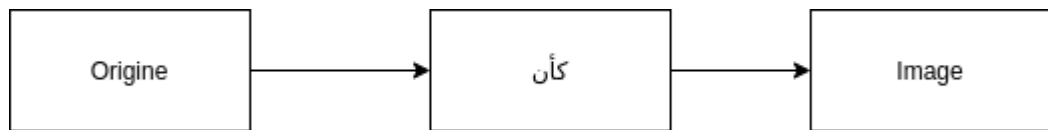


Figure 29: Aperçu des cas de ka-ana

لا تقوم الساعة حتى تقاتلوا قوما صغار الأعين عراض الوجوه كأن أعينهم حدق الجراد

3.1.7 Inna (إن):

L'un des principales particules dans les ahadith et qui est largement utilisée est أداة التأكيد. Cette particule assure que ce qui est dit est juste, il apporte une certaine assurance à l'auditeur que ce qu'il entend est bien vrai et que ce qu'il entend est un fait réel. Evidemment cette particule peut commencer une phrase et nous l'avons traitée bien sûr comme début d'un lafdh à part.

إن من أشراط الساعة أن تقاتلوا قوما عراض الوجوه

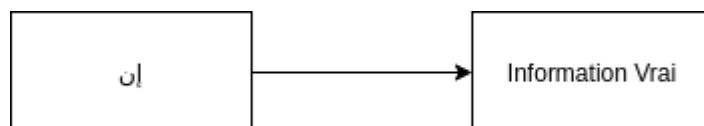


Figure 30: Aperçu des cas de inna

Chapitre 2 : Conception de l'outil d'extraction

3.1.8 Damir kasd (ضمير القصد):

ضمير قصد est une particule représentant une personne ou dans certain cas Dieu, donc cette particule a besoin d'un antécédent. Du coup, elle ne peut pas commencer une phrase mais nous l'avons aussi traitée comme si elle le pouvait.

Exemple :

ألا أخبركم بما يمحو الله به الخطايا ويرفع به الدرجات إسباغ الوضوء على المكاره وكثرة الخطا إلى المساجد وانتظار الصلاة بعد الصلاة فذلكم الرباط فذلكم الرباط فذلكم الرباط

حدثنا عبد الله بن يوسف، قال أخبرنا مالك، عن أبي الزناد، عن الأعرج، عن أبي هريرة، أن رسول الله صلى الله عليه وسلم قال " **والذي نفسي بيده** لقد هممت أن أمر بحطب فيحطب، ثم أمر بالصلاة فيؤذن لها، ثم أمر رجلا فيؤم الناس، ثم أخالف إلى رجال فأحرق عليهم بيوتهم، **والذي نفسي بيده** لو يعلم أحدهم أنه يجد عرقا سمينا أو مرماتين حسنتين لشهد العشاء ".

3.1.9 Kala (قال)

قال est aussi une autre entité que nous avons traitée. Cette entité ou plus exactement ce verbe introduit le discours d'un locuteur. Dans un matn, ceci représente en général une discussion entre le prophète et ses compagnons ou femmes. Dans ce cas, on peut s'attendre à un système question/réponse, à noter que le verbe lui-même n'apporte aucune valeur mais ses successeurs si.

ألا أدلكم على ما يكفر الله به الخطايا ويزيد به في الحسنات " **قالوا** بلى يا رسول الله . **قال** " إسباغ الوضوء عند المكاره

3.1.10 Résumé

On peut dire que notre algorithme était tiré de la linguistique et parfois de notre observation des matn qui représentent un contexte bien différent du reste de l'arabe académique.

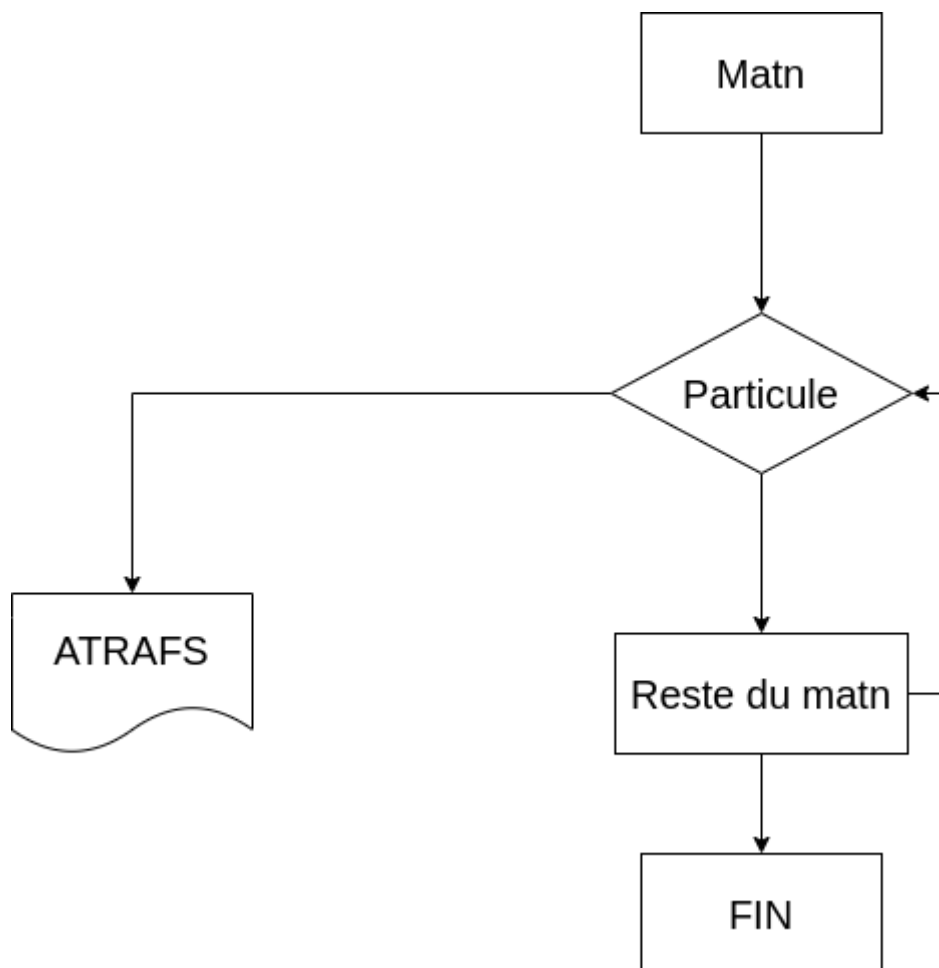


Figure 31: Aperçu générale de l'algorithme des alfadh

3.1.11 Conclusion

Nous avons pu observer dans cette deuxième partie d'extraction que notre inspiration est fortement liée à la linguistique arabe.

4 Conclusion

Dans ce chapitre nous avons présenté les deux méthodes d'extraction, la première du matin, la deuxième des al-fadh. Avec ça nous sommes parvenus à atteindre des résultats que nous démontrerons dans le chapitre Mise en œuvre et expérimentation [81].

Chapitre III : Mise en œuvre et expérimentation

1 Introduction

Dans les chapitres précédents nous avons parlé théoriquement des alfadh el hadith et de notre outil d'extraction des alfadh. Dans ce chapitre, nous allons présenter l'implémentation de notre système, de l'environnement de développement aux détails de chaque étape de développement.

2 Outils et bibliothèques

Cette section nous est dédiée aux bibliothèques et outils que nous avons utilisés pour développer notre solution.

2.1 Python

Python est un langage de programmation à usage général et de haut niveau qui peut être utilisé pour une grande variété d'applications. Il a été créé par Guido van Rossum et sorti pour la première fois le 20 février 1991.

Python est l'un des langages de programmation les plus utilisés par les développeurs aujourd'hui, et c'est de loin le langage le plus utilisé dans le domaine de l'intelligence artificielle. La plupart des bibliothèques NLP et Machine Learning sont disponibles en Python, c'est donc le meilleur choix pour notre cas.



Figure 32: Aperçu du logo de python

2.2 Flask

Flask est une infrastructure d'application Web WSGI (Web Server Gateway Interface) légère. Il est conçu pour faciliter et accélérer la mise en route, avec la possibilité de s'adapter

Chapitre 3 : Mise e œuvre et expérimentation

à des applications complexes. Il a commencé comme un simple wrapper autour de Werkzeug et Jinja et est devenu l'un des frameworks d'application Web Python les plus populaires.



Figure 33: Aperçu du logo de Flask

2.3 Tashaphyne

Tashaphyne est un Arabic light stemmer and segmentor. Il prend principalement en charge la dérivation légère (suppression des préfixes et des suffixes) et donne toutes les segmentations possibles. Il utilise un automate à états finis modifiés qui permettent de générer toutes les segmentations.

Il offre à la fois le stemming et l'extraction des racines contrairement aux souches Khoja, ISRI semmer, Assem, Farasa.

Tashaphyne est livré avec des préfixes et des suffixes par défaut, et accepte l'utilisation de listes de préfixes et de suffixes personnalisés, qui lui permettent de gérer plus d'aspect et de créer des souches personnalisées sans changer de code.

Tashaphyne a été développé par le professeur Taha Zerrouki, professeur à l'université de Buira.

2.4 Black-Box

Black-Box ou la boîte noire est un outil d'analyse morphologique développé par nos compatriotes Anes Fassih et Aymen Kerchi, deux étudiants qui ont passé aussi leur master avec le thème *الإعراب* el i'rab.

Chapitre 3 : Mise e œuvre et expérimentation

La boîte noire fait pratiquement la même chose que Tashaphyne, elle représente le déclic de beaucoup de problème constater dans Tashaphyne, aussi elle offre quelque chose de beaucoup mieux qu'est la distinction d'un mot c'est-à-dire qu'elle peut dire si c'est un nom ou un verbe, et sa ça ouvre des portes en traitement de la langue arabe qui n'ont jamais été considérés.

2.5 ReportLab

ReportLab est un outil de création rapport PDF, il propose multiple outil pour l'esthétique des rapports, mais ne prend pas en charge la langue arabe.

2.6 Arabic_Reshaper

Arabic_Reshaper est un outil qui permet d'introduire un nouvel encodage prenant compte de la langue arabe, c'est en effet l'outil utilisé pour l'introduction de la langue arabe dans la génération du PDF.

2.7 Kacstqurn

kacstqurn représente la font que nous avons utilisé pour matérialiser les caractère de la langue arabe dans la génération du PDF vue que les fonts standard ne la prennent pas en charge.

3 Présentation des résultats

Tout d’abord, il faut faire une sélection de ahadith de la sunna kawlia du dataset, puis importer le fichier dans le système, et le travail commence.

3.1Présentation du dataset

Comme nous l’avons précisé en haut, nous avons été chanceux de trouver un dataset regroupant non seulement tous les ahadiths des six livres référence mais en plus il est très bien structuré.

Dataset Ahadith
+ ID: Int
+ hadith_id: Int
+ source: Text
+ champter_no: Int
+ chapter: Text
+ chain_index: Int
+ text_ar: Text
+ text_en: Text

Figure 34: Aperçu de l’architecture du dataset

Dans la **Figure 34** nous avons vu le dataset ou chaque hadith avait un **ID** présent dans la première colonne qui correspond a sont id dans le livre où il est apparue, dans la deuxième colonne **hadith_id** ceci correspond à l’**id** du hadith dans le dataset complet c’est-à-dire si le dataset contient mille hadith on obtiendra mille id (de un à mille) ceci va nous être primordiale dans la suite de notre travaille, la colonne **source** représente le nom du Rawi en l’occurrence Albukhari, Muslim,..., la colonne **chapter_no** représente le **id** de chaque livre, la colonne **hadith_no** représente le numéro du hadith dans l’ensemble des livre d’un Rawi, la colonne **chapter** représente les livres de chaque Rawi, puis dans les deux dernière colonnes on retrouve le hadith en arabe dans la colonne **text_ar** puis en anglais dans la colonne **text_en**.

3.2Présentation des données de teste

Le fichier que nous allons utiliser contient trois ahadith sélectionnés à partir du fichier principal des ahadith (dataset), pour notre test nous avons choisi des ahadith considérés comme des variantes :

id	hadith_id	source	chapter_no	hadith_no	chapter	chain_indx	text_ar	text_en
183	30676	Sunan Ibn	1	462	The Book of	30201, 203	حدثنا أبو بكر بن أبي شيبة، حدثنا يحيى بن أبي بكر، حدثنا ز	It was narrated from Abu Sa'eed Khudri tha
184	30677	Sunan Ibn	1	463	The Book of	32048, 221	حدثنا يعقوب بن حميد بن كاسب، حدثنا سفيان بن حمزة، عن	It was narrated from Abu Hurairah that:
143	24584	Sunan an-	1	144	The Book of	30367, 200	أخبرنا قتيبة، عن مالك، عن العلاء بن عبد الرحمن، عن أبيه،	It was narrated from Abu Hurairah that the

Figure 35: fichier d'importation pour test

Tableau 10: Ahadith utilisé pour le 1^{er} teste

<p>حدثنا أبو بكر بن أبي شيبة، حدثنا يحيى بن أبي بكر، حدثنا زهير بن محمد، عن عبد الله بن محمد بن عقيل، عن سعيد بن المسيب، عن أبي سعيد الخدري، أنه سمع رسول الله صلى الله عليه وسلم قال " ألا أدلكم على ما يكفر الله به الخطايا ويزيد به في الحسنات " . قالوا بلى يا رسول الله . قال " إسباغ الوضوء على المكاره وكثرة الخطا إلى المساجد وانتظار الصلاة بعد الصلاة " .</p>
<p>حدثنا يعقوب بن حميد بن كاسب، حدثنا سفيان بن حمزة، عن كثير بن زيد، عن الوليد بن رباح، عن أبي هريرة، أن النبي صلى الله عليه وسلم قال " كفارات الخطايا إسباغ الوضوء على المكاره وإعمال الأقدام إلى المساجد وانتظار الصلاة بعد الصلاة " .</p>
<p>أخبرنا قتيبة، عن مالك، عن العلاء بن عبد الرحمن، عن أبيه، عن أبي هريرة، أن رسول الله صلى الله عليه وسلم قال " ألا أخبركم بما يمحو الله به الخطايا ويرفع به الدرجات إسباغ الوضوء على المكاره وكثرة الخطا إلى المساجد وانتظار الصلاة بعد الصلاة فذلكم الرباط فذلكم الرباط فذلكم الرباط " .</p>

3.3Présentation des interfaces

Après sélection du fichier et importation nous passons à la deuxième interface de téléchargement.



Figure 36: Interface d'importation du fichier

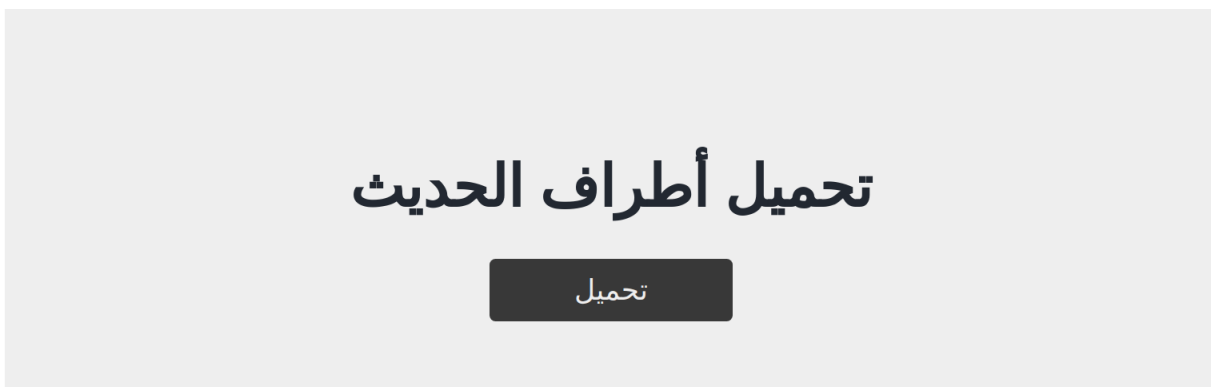


Figure 37: Interface de téléchargement du fichier

3.4 Résultats

Le résultat sera un fichier PDF contenant les alfadh avec leurs index :

وإعمال الأقدام إلى المساجد	ألا أخبركم بما يمحو الله به الخطايا
سنن ابن ماجه كتاب الطهارة وسننها حديث رقم 463	سنن النسائي كتاب الطهارة حديث رقم 144
والتظار الصلاة بعد الصلاة	ألا أدلكم على ما يكفر الله به الخطايا
سنن ابن ماجه كتاب الطهارة وسننها حديث رقم 462	سنن ابن ماجه كتاب الطهارة وسننها حديث رقم 462
سنن ابن ماجه كتاب الطهارة وسننها حديث رقم 463	إسباغ الوضوء على المكاره
سنن النسائي كتاب الطهارة حديث رقم 144	سنن ابن ماجه كتاب الطهارة وسننها حديث رقم 462
وكثرة الخطا إلى المساجد	بلى يا رسول الله
سنن ابن ماجه كتاب الطهارة وسننها حديث رقم 462	سنن ابن ماجه كتاب الطهارة وسننها حديث رقم 462
سنن النسائي كتاب الطهارة حديث رقم 144	فذلكم الرباط
ويرفع به الدرجات إسباغ الوضوء على المكاره	سنن النسائي كتاب الطهارة حديث رقم 144
سنن النسائي كتاب الطهارة حديث رقم 144	سنن النسائي كتاب الطهارة حديث رقم 144
ويزيد به في الحسنات	كفارات الخطايا إسباغ الوضوء على المكاره
سنن ابن ماجه كتاب الطهارة وسننها حديث رقم 462	سنن ابن ماجه كتاب الطهارة وسننها حديث رقم 463

Figure 38: résultats de l'échantillon

Les alfadh étant en rouge et l'indexation en noir nous arrivons avec cet échantillon à voir les premiers résultats, et nous voyons clairement que le résultat est très prometteur.

Nous sommes donc passés à un test plus grand, plus varié, plus de ahadith pour voir le comportement de notre programme.

Nous avons testé sur une cinquantaine de ahadith ayant comme critère la variation entre eux c'est-à-dire leurs apparitions dans plusieurs livres pour plusieurs auteurs.

Pour connaître la précision de notre algorithme nous avons utilisé موسوعة أطراف الحديث comme modèle de comparaison.

TP

TP + FP

Figure 39: Formule de précision

Dans la formule, on présente les cas de (Positif et Négatif) True positif et False positif, ou le premier représente le nombre de cas déclarés vrai et considéré vrai par l’algorithme, quant à False positif c’est le cas où c’est négatif et considéré positif par l’algorithme.

Dans notre cas on n’a pas une telle classification, nous avons donc tiré une autre solution pour pouvoir faire un calcul de précision :

Nombre de atrafs extrait et parue dans l'encyclopédie

Nombre de atrafs extrait total

Figure 40: Formule de précision des alfadh

Les résultats de correspondance (lafdh et indexation) étaient les suivants :

Tableau 11: résultats finaux

Nbr de alfadh de l’algorithme	120
Nbr de alfadh de l’encyclopédie	101
Nbr de alfadh de l’algorithme qui sont dans l’encyclopédie	41

L’encyclopédie est notre référence dans le calcul mais reste un outil présentant certains défauts que nous avons constaté lors de nos recherches pour effectuer le calcul de précision, à savoir que certains alfadh n’ont pas était pris dans l’encyclopédie, c’est-à-dire qu’en les cherchant, on les trouve nulle part est ça c’est paralysant dans notre cas, vue que nous n’avons pas été en mesure de retrouver plus d’une dizaine de alfadh qui devrait y être.

Chapitre 3 : Mise e œuvre et expérimentation

Malgré ça nous avons effectué le calcul de précision et a était évalué à 0.34 soit 34%.

4 Conclusion

Le travail de contrôle des résultats est de détection du taux de réussite dans notre travail est difficile à réaliser car il y a un besoin accru d'avoir un expert sous la main à qui on donnerait une liste de hadith et qui se chargerait d'élaborer la liste des alfadh et les références des hadiths desquels ont été extraits ces alfadh. Malheureusement, nous n'avons pas pu trouver pareil expert.

Conclusion générale

Conclusion générale

1 Sommaire

Depuis des siècles, le besoin de alfadh el hadith est devenu un besoin de grande valeur, et il a augmenté encore plus dans ces dernières années. Et c'est parce que les alfadh el hadith fournissent un moyen de recherche rapide des ahadiths, en plus de leur indexation qui fait d'eux une source très fiable.

En effet, on aurait du mal à retrouver tous les ahadiths dans lesquels figure une certaine expression du fait que cela obligerai à balayer une très grande quantité de données. Ajouter à cela que les variantes d'un hadith ne sont pas toujours « calculables ». On peut avoir des placements différents des mots, voire des différences notables.

Dans notre projet, l'objectif était de créer un algorithme faisant le traitement des ahadiths et fournissant à la fin une encyclopédie des alfadh automatique, équivalente à موسوعة أطراف الحديث. Pour cela, nous étions obligés déjà d'apprendre à connaître le monde des ahadith, d'apprendre ce que sont les alfadh el hadith, leur traits discriminants, leurs intérêts, leurs histoires...

Pendant la recherche nous avons été choqués par l'opposition d'idée de départ que nous avions sur la langue arabe et pire le domaine du hadith qui paraissaient épanoui, mais au contraire les travaux se font rare et basique voir parfois classique surtout quand on voit ce que le traitement automatique du langage peut apporter à ces domaines-là, on peut dire que ce sont des terrains vierges qui n'attendent qu'à être exploités.

Les objectifs que nous nous étions fixés initialement ont été partiellement atteints, nous avons réussi à créer un outil d'extraction des alfadh el hadith fonctionnel qui peut avoir des ahadiths en entrée puis les transforme en alfadh en les identifiants, bien que nous ayons initialement espéré quelque chose de plus que cela, on peut dire que ce que nous avons réalisé est acceptable compte tenu de la situation que le monde a traversée lors de du démarrage du projet.

Nous avons rencontré beaucoup de problèmes, certains nous les avons traités, d'autres non et nous avons été obligés d'adapter et réadapter nos choix. Le plus grand problème rencontré était l'absence de travaux sur l'extraction automatique des alfadh, le manque de travaux dans

Conclusion générale

le domaine du hadith, l'absence des données qui au départ nous a même fait douter de la possibilité de poursuivre notre projet. On peut dire que c'est impossible de les couvrir tous.

2 Perspective

- Nous avons aussi testé un moteur de recherche et que bien que donnant des résultats, ne réponds pas exactement aux besoins.
- Nous avons même essayé de travailler avec des techniques d'apprentissage automatique en vain.
- Ce projet pourrait aller vers une nouvelle dimension, si dans certaine mesure les travaux sur la langue arabe se développent en l'occurrence la sémantique.
- Il est clair que pour mesurer la qualité de notre travail, il faudrait un échantillon plus grand. Toutefois, la difficulté réside dans le fait que pour faire la comparaison avec un traitement manuel en vue de valider les résultats -ou les invalider-, il faudrait parcourir des quantités énormes de volumes encyclopédiques pour lesquels nous ne possédons que des versions PDF. Vérifier un seul hadith pourrait prendre des heures. Aussi, avoir un échantillon d'un millier de hadith nous mènerait à un travail de plusieurs milliers d'heures.
- Nous avons essayé de nous faire aider par des connaisseurs à qui nous avons confié une cinquantaine de hadith et à qui nous avons demandé d'en extraire les alfadh. Malheureusement, et malgré des promesses, nous n'avons pas eu de retour.
- Nous estimons avoir déblayé un petit peu ce travail qui pourrait faire l'objet d'études plus poussées où les aspects grammaticaux voire sémantiques seraient pris en charge.
- Ce travail pourrait prendre un niveau meilleur si nous arrivons à prendre des objectifs basés sur le traitement de la sémantique.
- Une approche basé sur l'extraction des concepts pourrait aussi faire un bon objet de recherche dans un tel projet et serait une grande aide aux chercheurs.

Références bibliographiques

Bibliographie

1. Ahsan Mahmood, Fawaz K. Alarfaj, Hikmat Ullah Khan*, Muhammad Ramzan, Mahwish Ilyas. (2018). A Multilingual Datasets Repository of the Hadith Content. *International Journal of Advanced Computer Science and Applications*,, 165.
2. Ahsan Mahmood, Hikmat Ullah Khan, Zahoor-ur-Rehman, Wahab Khan. (2017). Query based information retrieval and knowledge extraction using Hadith datasets. *Islamabad*. 13th International Conference on Emerging Technologies (ICET),.
3. Aldhaln, Kawther A., Akram M. Zeki, Ahmed M. Zeki. (2012). *Knowledge extraction in Hadith using data mining technique*.
4. Aldhlan, K.A, Zeki, AM. (2010). *Datamining and Islamic knowledge extraction: alhadith as a knowledge resource*.
5. Alfi, M. Y. (2004). An Applied Linguistics Approach to Improving the Memorization of the Holy Quran: Suggestions for Designing Practice Activities for Learning and Teaching. *Journal King Saud Univ*, 1-32.
6. Ali, Imran. (2012). Application of a Mining Algorithm to Finding Frequent Patterns in a Text Corpus: A Case Study of the Arabic. *International Journal of Software Engineering and Its Applications*, 3.
7. Altaher, Altyeb. (2017). Hybrid approach for sentiment analysis of Arabic tweets based on deep learning model and features weighting. *International Journal of Advanced and Applied Sciences*, 43-49.
8. Ammar Alsalka, Eric Atwell, Shatha Altammami. (22 July 2019). *Text Segmentation Using N-grams to Annotate Hadith Corpus*. Leeds: Association for Computational Linguistics.
9. Azmi AM,Alkhalifah F, Alsaheed A, Barnawi Y. (2014). *Using non-conventional search schemes to retrieve Hadiths*.

Bibliographie

10. Azmi, A. (2010). *iTree – Automating the Construction of the Narration Tree of Hadiths (Prophetic Traditions)*. King saud University: Natural Language Processing and Knowledge Engineering (NLP-KE), 2010 International Conference on.
11. Boella. (2011). *Regular expressions for interpreting and cross-referencing Hadith texts*.
12. Bounhas., I. (2019). *On the usage of a classical arabic corpus as a language resource: related*. ACM Transactions on Asian and Low-Resource Language Information Processing.
13. Brodie, Benjamin C., David E. Taylor, Ron K. Cytron. (2006). A scalable architecture for high-throughput regular-expression pattern matching. *ACM SIGARCH Computer Architecture News*.
14. Cardoza, C. (2017). *Text analysis framework for understanding cyber-crimes*. Bangalor: Christ University.
15. Chee Yong Chan, Minos Garofalakis, Rajeev Rastogi. (2014). Indexed Regular Expression Matching. *Springer US*, 1-6.
16. Crescenzi, Valter, Giansalvatore Mecca, Paolo Merialdo. (2001). Roadrunner: Towards automatic data extraction from large web sites. *VLDB*, 1.
17. F. Harrag,. (2014). Text mining approach for knowledge extraction in Sahîh Al-Bukhari. *Computers in Human Behavior archive*, 558-566.
18. Fang Yu, Zhifeng Chen, Yanlei Diao, T.V. Lakhsman, Randy H.Katz. (2006). *Fast and memory-efficient regular expression matching for deep packet inspection*.
19. Fouzi Harrag, Aboubekour Hamdi-Cherif, Eyas El-Qawasmeh. (2008). *"Vector space model for Arabic information retrieval—application to "Hadith" indexing*.
20. Fouzi Harrag, Eyas El-Qawasmeh, Abdul Malik Salman Al-Salman. (2011). *Extracting named entities from prophetic narration texts (Hadith)*. Berlin.
21. Garrett, Jesse James. (2015). *Ajax: A new approach to web applications*.

Bibliographie

22. Jbara, K. (2010). Knowledge discovery in Al-Hadith using text classification algorithm. *Journal of American Science*, 19-409.
23. Kais Dukes, Eric Atwell. (2012). *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul: European Language Resources Association (ELRA).
24. Kais Dukes, Tim Buckwalter. (2010). *A Dependency Treebank of the Quran using Traditional Arabic Grammar*.
25. Khan HU, Saqlain SM, Shoaib M, Sher M. (2013). Ontology Based Semantic Search in Holy Quran. *International Journal of Future Computer and Communication*, 6.
26. Maheen Akhter Ayesha, Sahar Noor, Muhammad Ramzan, Hikmat Ullah Khan, Muhammad Shoaib. (2017). Evaluating Urdu to Arabic Machine Translation Tools,. *International Journal of Advanced Computer Science and Applications*, 90-96.
27. Michela Becchi ,Anat Bremler-Barr ,David Hay ,Omer Kochba ,Yaron Koral,. (2015). *Accelerating regular expression matching over compressed HTTP*.
28. Mohammad Alqahtani, Eric Atwell. (2016). Arabic Quranic Search Tool Based on Ontology. *Natural Language Processing and Information Systems*, 478-485.
29. Mohanad Jasim Jaber, Saidah Saad. (2016). NER in english translation of hadith documents using classifiers combination. *Journal of Theoretical and Applied Information Technology*, 348-354.
30. Muhammad Shoaib M, Yasin MN, Khan HU, Saeed MI, Khiyal MS. (2009). *Relational WordNet Model for Semantic Search in Holy Quran*.
31. Nursyahidah Alias, Zulhilmi Mohamed Nor, Nurazzah Abdul Rahman. (2016). *Searching Algorithm of Authentic Chain of Narrators" in Shahih Bukhari Book*. MALAYSIA,.
32. Puteri N. E. Nohuddin, Zuraini Zainol , Angela S. H. Lee , A. Imran Nordin , Zaharin Yusoff. (2018). A case study in knowledge acquisition for logistic cargo distribution data Mining framework. *International Journal of Advanced and Applied Sciences*, 8-14.

Bibliographie

33. Rehan Khan, Hikmat Ullah Khan, Muhammad Shehzad Faisal, Khalid Iqbal, Muhammad Shahid Iqbal Malik,. (2016). An Analysis of Twitter users of Pakistan. *International Journal of Computer Science and Information Security*, 8.
34. Ruqaiya J. Kadhim, Dr. Norita M. Norwawi, Dr. Adel M. Abdulaaziz, Ashraf Al_Omouh. (2015). Extraction of Hadith Based on Semantic Annotation. *International Journal of Computer Science and Network*,, 273.
35. Saloot, M. A., Idris, N., Mahmud, R., Ja'afar, S., Thorleuchter, D., & Gani, A. (2016). Hadith data mining and classification: a comparative analysis. *Artificial Intelligence Review*, 113-118.
36. Shatnawi MQ, Abuein QQ, Darwish O. (2011). *Verification Hadith Correctness in Islamic Web Pages Using Information Retrieval Techniques*.
37. Siddiqui, Muazzam Ahmed, Mostafa El-Sayed Saleh, Abobakr Ahmed Bagais. (2014). *Extraction and Visualization of the Chain of Narrators from Hadiths using Named Entity Recognition and Classification*.
38. Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth,. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 18.
39. Zulfiqar Ali, Waseem Shahzad , Syed Khuram Shahzad. (2017). A review on comparative performance analysis of associative classifiers. *International Journal of Advanced and Applied Sciences*, 96-103.
40. دار الكتب العلمية موسوعة أطراف الحديث. م. ا. زغلول (1996).
41. دار الكتب العلمية موسوعة أطراف الحديث. م. ا. زغلول (1996).
42. المعجم المفهرس لألفاظ الحديث النبوي. (2002, 07 01). إ. ويب: <https://www.islamweb.net/>
43. كتب الأطراف. (2002, 06 30). إ. ويب: <https://www.islamweb.net/>

Bibliographie