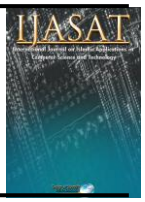
Design for Scientific RENAISSANCE



An LMF-based Normalization approach of Arabic Islamic dictionaries for Arabic Word Sense Disambiguation: application on hadith

Nadia Soudani[1,5,a], Ibrahim Bounhas[2,5,b], Bilel ElAyeb[3,6,c], Yahya Slimani [4, 5,d]

[1] Faculty of Sciences of Tunis (FST), University of Tunis El Manar, Tunisia
[2] Higher Institute of Documentation (ISD), University of Manouba, Tunisia
[3] Emirates College of Technology, P.O. Box: 41009. Abu Dhabi, United Arab Emirates.
[4] Higher Institute of Multimedia Arts of Manouba (ISAMM), University of Manouba, Tunisia
[5] LISI Laboratory of computer science for industrial systems, Carthage University, Tunisia
[6] RIADI Laboratory, National School of Computer Science (ENSI), Universiy of Manouba, Tunisia

[a]Nadia.soudani@gmail.com, [b]Bounhas.Ibrahim@gmail.com, [c]Bilel.Elayeb@riadi.rnu.tn, [d]Yahya.Slimani@fst.rnu.tn

ABSTRACT

In this paper, we propose an approach for normalizing Arabic Dictionaries. This approach is used to transform non structured Arabic dictionaries into LMF (Lexical Markup Framework) based-normalized ones. We are basically exploiting Arabic Islamic dictionaries of hadith. An ontology will be then constructed from these normalized dictionaries. This ontology will contain explicit and formal knowledge about information in hadith. It will be used later by an information retrieval system for Word Sense Disambiguation of Arabic terms of hadith either in the formulated user query or in the texts of hadith.

Keywords: Arabic language, Arabic Islamic Dictionary, hadith, LMF, Ontology, Word Sense Disambiguation, Information Retrieval System

1. Introduction

The corpus of hadith constitutes a rich set of knowledge which it is still ineffectively exploited. In addition, Arabic language has particular linguistic features at the morphologic, syntactic and semantic levels, which cause a lot of ambiguity with Arabic terms and particularly with the terms of hadith (Bounhas et al., 2011a). These specificities are challenging knowledge extraction in these collections. Natural Language Processing (NLP) of the Arabic Language suffers from the lack of linguistic resources as corpora, dictionaries, ontologies and standards test collections (Bounhas et al., 2011a; Bounhas, 2012; Jarrar, 2011). However, the existing resources as electronic dictionaries are neither exhaustive nor standardized. Then, their exploitation by Information Retrieval (IR) tools couldn't be effectively achieved. As a result, the performance of these tools in processing Arabic linguistic resources was declined in terms of relevance of search results. We essentially start dealing with these matters through the advent of the semantic web (Beseiso et al., 2010). In a context of IR, ambiguity is detected both in the text of the query and in the text of hadith itself. We

aim in our work to propose an approach for Arabic Word Sense Disambiguation (WSD) with experimentation taken on the Arabic terms of hadith.

But, WSD looks for identifying the real sense of the word, which is really meant by the user so as to be able to return to him effective and pertinent results. Then, we will need the use of dictionaries of hadith. When beginning study of these dictionaries, we notice that they aren't formalized and they couldn't be directly exploited by IR systems. Indeed, these dictionaries are incomplete and don't cover all the terms of the Arabic language and especially the terms of all the domains of hadith (eg. Prayer الصّلاة, Purification الطّهارة , Fasting الصّوم , Ablution الوضوء, etc.). We will so, for some words, turn into Arabic editorial dictionaries. Besides, the definitions of the terms of hadith are themselves led with ambiguity and the sense isn't clear. Then, different semantic relations are deduced between the different senses of the terms which lead to another problem for the identification of the searched sense.

To exploit these dictionaries, a formalization of both structure and sense is needed. Formalization of the structure consists in normalizing the dictionary into LMF format (Francopoulo, 2013) while that one of the sense will be fulfilled by efforts of transforming this new LMF dictionary into ontology. This ontology will contain hidden knowledge about information in hadith texts. After that, it will be exploited by an information retrieval tool for the purpose of user query reformulation by disambiguating the terms of the query.

At this level, we have to note that Arabic WSD is a complex problem and a difficult task (Navigli, 2009; Zouaghi et al., 2012). It essentially depends as mentioned before on the features of the Arabic language which is agglutinative, derivational and inflectional. It also depends on the ambiguity of the Arabic term and its capacity to have many significations changing when context changes. To automatically disambiguate a word, we need the combined use of different resources as corpora, dictionaries, ontologies (Bounhas et al., 2011b) and standard test collections. According to our readings, the only research project that aims to normalize Arabic dictionaries into LMF format is currently in MIRACL Laboratory (Ben Ammar et al., 2011 ; Khemakhem et al., 2009). However, the authors didn't consider Arabic Islamic dictionaries in their work. Besides, transformation of the normalized dictionary into ontology is just in its first primary steps and its use for Arabic IR is a future work (Ben Ammar et al., 2013) and it isn't currently handled. In our paper, we will present our generic proposed approach of normalization and some experiments.

2. Dictionaries of hadith

To determine the correspondent senses of the terms of hadith and then, in next steps concepts of ontology, it is necessary to use dictionaries of hadith. For this purpose, we investigate this class of dictionaries to identify them and to deeply study them. The following table (Table 1) describes the principle existent dictionaries of hadith, which are available in machine readable formats. A brief description of types of their lexical entries is given below.

Table 1: Dictionaries of hadith.

Type of lexical Entry	Format	Dictionary
Root then hadith then references to sources of hadiths	PDF	**المعجم المفهرس لألفاظ الحديث النبوى** (Al-Muâjam Al-Mufahres Li-Alfadh Al-hadith Al-Nabawi)
Root then hadith then explanations of words	Word	**الفائق في غريب الحديث للزمخشري** (Al-Faiq Fi-Gharib Al-hadith for Zamakhchari)
Root then hadith then explanations of words (It is more exhaustive than Al-Faiq)	Word	**النهاية في غريب الحديث والأثر** (Al-Nihaya Fi-Gharib Al-hadith wAl-Athar)
Lemma with "ال" (Ex : الأباء، الباب)	Word	**معجم لغة الفقهاء** (Muaâjam lughat Al-Fuqahaâ)
Name of narrator	Word	**إعراب ما يشكل من ألفاظ الحديث النبوي** (Iârab Ma Yuchkal Men Alfadh Al-hadith Al-Nabawi)

3. Steps of the normalization process

The following Figure 1 illustrates our generic proposed approach for normalization and for ontology construction inspired from the works of Mhiri et al. (2006) and Zaidi (2013). The proposed approach designs an abstract architecture delineating an interactive process. All steps interact by input and output flows. Nevertheless, in this paper, we only concentrate on the normalization steps. We have exploited the Arabic dictionary of hadith « النهاية في غريب الحديث و الأثر» (Al-Nihaya fi Gharib Al-hadith w Al-Athar) on which we are processing all our experimentations.

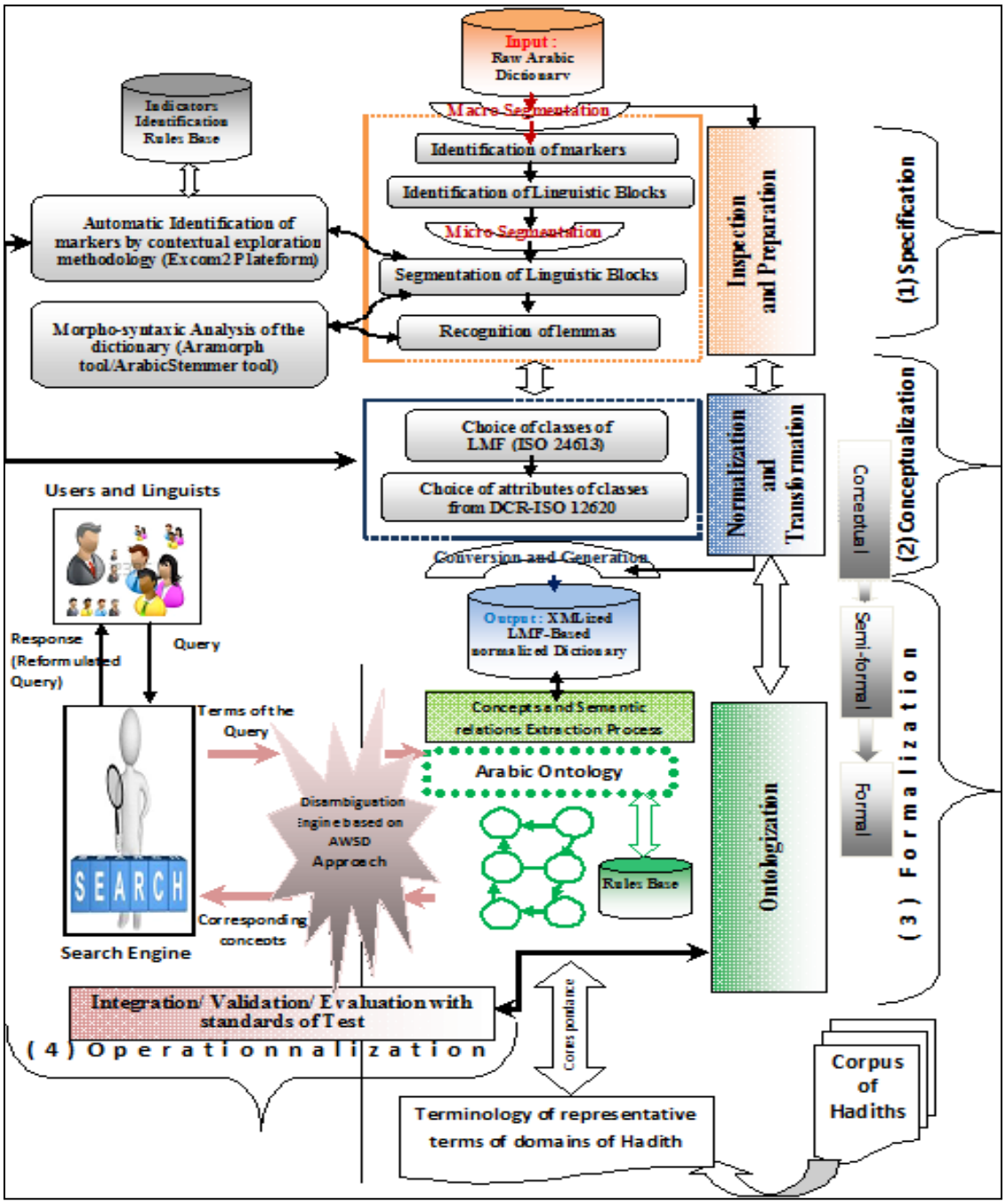


Figure 1: Generic Approach for normalization of Arabic Dictionaries of hadith

This approach is based on two principle phases: the Inspection and Preparation phase that consists in a whole work of specification and the second phase of Normalization and Transformation, which consists in a whole work of conceptualization.

3.1 Phase of Specification: Inspection and Preparation

This phase consists in exploring the dictionary to be normalized by a macro-segmentation. It begins with the segmentation of the raw dictionary to cross after four essential and fundamental steps which are: identification of markers, identification of linguistic blocks, their segmentation and recognition of lemmas with an Arabic stemmer tool. The linguistic Analysis of the dictionary will use a morphological disambiguation tool based on Aramorph, developed by Ayed et al. (2012). To automatically identify markers and all kinds of information, we are using the Excom2 platform based on contextual exploration (Alrahabi, 2010). We opt for this choice due that this method allows us to locate with linguistic markers corresponding linguistic information as specified in indicators.

3.1.1 Macro Segmentation

Taking an Arabic raw dictionary, a primordial step consists in segmenting this dictionary by identifying all the lexical entries of the dictionary and bordering all their relative blocks.
Figure 2 shows an example of a raw lexical entry أبد from Al-Nihaya Dictionary.

أَبَدَ
قَالَ رَافِعُ بْنُ خَدِيجٍ: أَصَبْنَا نَهْبَ إِبِلٍ فَنَدَّ مِنْهَا بَعِيرٍ فَرَمَاهُ رَجُلٌ بِسَهْمٍ فَحَبَسَهُ، فَقَالَ رَسُولُ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ:
«إِنَّ لِهَذِهِ الْإِبِلِ [G البهائم] أَوَابِدَ كَأَوَابِدِ الْوَحْشِ، فَإِذَا غَلَبَكُمْ مِنْهَا شَيْءٌ فَافْعَلُوا بِهِ هَكَذَا» الْأَوَابِدُ: جَمْعُ آبِدَةٍ
وَهِيَ الَّتِي قَدْ تَأَبَّدَتْ أَيْ تَوَحَّشَتْ ونَفَرَتْ مِنَ الْإِنْسِ. وَقَدْ أَبَدَتْ تَأْبِدُ وتَأْبُدُ.
وَمِنْهُ حَدِيثُ أُمِّ زَرْعٍ «فَأَرَاحَ عَلَيَّ مِنْ كُلِّ سَائِمَةٍ زَوْجَيْنِ، وَمِنْ كُلِّ آبِدَةٍ اثْنَتَيْنِ» تُرِيدُ أَنْوَاعًا مِنْ ضُرُوبِ الْوَحْشِ.
وَمِنْهُ قَوْلُهُمْ: جَاءَ بِآبِدَةٍ: أَيْ بِأَمْرٍ عَظِيمٍ يُنْفَرُ مِنْهُ ويُسْتَوْحَشُ.

Figure 2 : Reproduced extract of dictionary of hadith «النهاية في غريب الحديث و الأثر» for the entry أبد showing some occurrences of the term أبد

3.1.1.1 Identification of markers

This step consists in analyzing the dictionary and precisely each lexical entry to identify the different indicators that serve to distinguish each type of information. Table 2 shows examples of types of information being identified in the dictionary of hadith.

Table 2: Examples of existent kinds of information in the dictionary «Al-Nihaya».

Type of information	Examples of markers
Root	أبد، أبب، أبر، أثر
Lemma	أبد A derived word or inflectional word of the entry, it'll be detailed in next steps of Micro-segmentation
Implicit morphological attributes	(to be determined with Ayed's tool) POS=verb, schema=فعل
Expression introducing a definition	تريد، أراد، يعني، أي، قال الرياشي أي، وقيل
Definition	أبب and the entry أبّ for the lemma المرعى المتهيّء للرعي و القطع
Expressions that indicate place of words in other dictionaries	for the entry of dictionary أبر هَكَذَا أَخْرَجَهُ الْحَافِظُ أَبُو مُوسَى الْأَصْفَهَانِيُّ فِي حَرْفِ الْهَمْزَةِ، وَعَادَ أَخْرَجَهُ فِي حَرْفِ الْبَاءِ
Alternative for a given word	الإبل for[G البهائم] ,المنافق for[G الكافر]
Beginning of a new sense	وفي حديث...، ومنه قوله...
Introduction of verses	قال الله تعالى، قوله تعالى
Verses (Quranic Context)	فاكهة و أبّا
Hadith (hadithian context of the lexical entry)	في حديث (الراوي) ، قال (الراوي) ، في وصف (الراوي) ، في كلام (الراوي) ، وفي حديث : بعضهم

Explanation of hadith	نَحْنُ وَإِيَّاكُمْ فِي الْحُكْمِ سَوَاءٌ، لَا فَضْلَ لِلأَمِيرِ عَلَى مَأْمُورٍ، كَالْخُوصَةِ إِذَا شُقَّت بِاثْنَتَيْنِ مُتَسَاوِيَتَيْنِ For the hadith (« فِي حَدِيثِ السَّقِيفَةِ «الْأَمْرُ بَيْنَنَا وَبَيْنَكُمْ كَقَدِّ الأُبْلُمَة) for the entry أبلم
Expressions that introduce names of ALLAH	في أسماء الله تعالى
Chain of narrators	الحسن، أَمّ زرع، عائشة، معاوية، أبي هريرة،...
Indicator of poesy	و منه قول الشاعر، كقوله
Poesy (Poetic context)	وَالْمَرْء مَا عَاشَ مَمْدُودٌ لَهُ أَمَلٌ ... لَا يَنْتَهِي الْعُمْرُ حتَّى يَنْتَهِي الأَثَرُ for the entry أثر

3.1.1.2 Identification of linguistic blocks

When identifying markers, we discern for each lexical entry of the dictionary of hadith, the existence only of semantic blocks and the absence of morpho-syntaxic attributes. These attributes will be determined by Ayed's tool.

3.1.2 Micro Segmentation

3.1.2.1 Segmentation of semantic blocks

This step consists in marking and annotating each linguistic entity. Given the big semantic richness of the lexical entries of the dictionaries of hadith and even their complex structure, this step is not simple. Intervention of the expert (linguist) is necessary to identify kinds of information of each semantic sub-block and to validate with us identified types.

3.1.2.2 Automatic identification of markers by Excom2 platform

Given the huge size of information and the voluminous number of lexical entries and also the great variety of kinds of information being detected in the experimented dictionary « النهاية في غريب الحديث و الأثر», we are thinking to the Contextual exploration method of Desclès (Alrahabi, 2010) to automatically annotate the dictionary with semantic categories. These categories refer to the different kinds of information. After defining markers and generating rules of automatic annotation for each semantic category, the annotation can be processed. Figure 3 shows examples of corresponding semantic categories for some kinds of information. We can annotate the text of the input file of the dictionary of hadith with only one semantic category or with all of them. Figure 4 illustrates the annotation of text for only the definitions by automatically applying the correspondent rules on the markers of the definition. We only take a portion of the dictionary composed of the first 15 lexical entries. We are applying tests incrementally and iteratively until performing results and increasing the recognition rates.

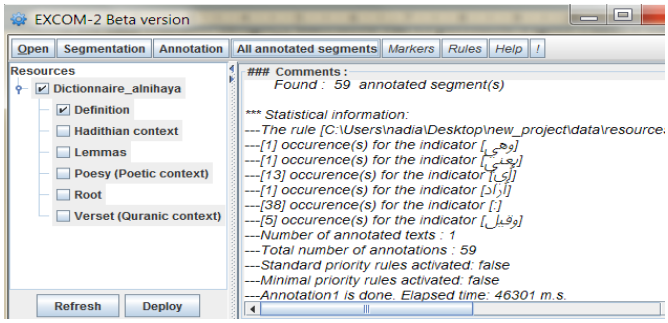


Figure 3 : Examples of semantic categories defined in Excom2 and Annotation of Definitions of the dictionary: application to a portion of the dictionary of hadith

As shown in figure 4 below, the definitions are delimited by their correspondent markers.

Source	Sentence	Annotation Name	Rule Type	Rule Scope
...esults\annotated1\text	أوابد [البهائم G] قال رافع بن خديج: أصبنا نهب إبل فند منها بعير فرماه رجل بسهم فحبسه، فقال رسول الله صلى الله عليه وسلم: «إن لهذه الإبل أوابد كأوابد الوحش، فإذا غلبكم منها شيء فافعلوا به هكذا» الأوابد: جمع آبدة وهي التي قد تأبدت أي توحشت ونفرت من الإنس.	Dictionnaire_alnihaya.Definition	1	sentence
...esults\annotated1\text	أوابد [البهائم G] قال رافع بن خديج: أصبنا نهب إبل فند منها بعير فرماه رجل بسهم فحبسه، فقال رسول الله صلى الله عليه وسلم: «إن لهذه الإبل أوابد كأوابد الوحش، فإذا غلبكم منها شيء فافعلوا به هكذا» الأوابد: جمع آبدة وهي التي قد تأبدت أي توحشت ونفرت من الإنس.	Dictionnaire_alnihaya.Definition	1	sentence
...esults\annotated1\text	ومنه قولهم: جاء بأبدة: أي بأمر عظيم ينفر منه ويستوحش.	Dictionnaire_alnihaya.Definition	1	sentence

Figure 4 : Example of results of annotated segments for Definition for the entry

3.1.2.3 Morpho-syntaxic Analysis with Aramorph

Morphologic and syntactic attributes are almost completely absent for each entry except rarely, we meet some ones as the grammatical number like plural (أوابد جمع آبدة). In the normalized format of the dictionary as we will see in the section 3.2, we need to characterize all the occurrences of a lexical entry: the root, the lemma and all the different derived and flectional forms of lemmas. So, we envisage the use of Ayed's tool to identify the different attributes. Nevertheless, the intervention of the expert is usually necessary to validate results.

3.1.2.4 Recognition of lemmas

We consider recognition of lemmas as an obligatory task to be able to deduce all possible flexional and derived forms of the root and all possible corresponding senses so as not to lose the original structure of both each entry of dictionary and the entire dictionary. All of these modalities of information are so important and they will enormously facilitate the step of normalization. At this level, we have used automatic linguistic tools as an Arabic Stemmer. When knowing all the derived forms both verbal and nominal ones for a given term, we will be then able to determine all the contextual possibilities of use of this word and its probable appearance in Arabic language. From this task of lemmatization for example for a portion of the entry أبد , we deduce that the root أبد have three lemmas which are especially أبدas a noun and أبد as a verb and آبدas a noun and which figures in the dictionary in its feminine gender آبدة .

All of these preparatory steps will after facilitate normalization of dictionaries to better profit from their semantic richness and their implicit knowledge.

3.2 Normalization and Transformation

3.2.1 Choice of LMF classes (ISO 24613)

This step mainly consists in identifying the corresponding classes from the LMF standard for the outputs of the preceding preparatory steps. These classes form simply the correspondent LMF model for Arabic dictionaries of hadith. They are reproduced from the extension MRD (Machine Readable Dictionary) of LMF combined with classes taken from the morphological extension and the semantic one of LMF. We choose LMF [1] (baptized ISO 24613 [2]) (Francopoulo, 2013) due that it is mostly an ISO standard that supports NLP and it defines structures of linguistic resources. It is applicable to Arabic language and to dictionaries.

3.2.2 Choice of Attributes from DCR_ISO (ISO 12620)

For each class of the LMF model corresponding to the dictionary of hadith, their relative attributes are associated, and which are defined in the ISO DCR-ISO (Data Category Registry). Association between classes and attributes will be more clarified and shown in the

[1] http://www.lexicalmarkupframework.org/
[2] http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=37327

XML file which is an XMLized image of the LMF-based normalized dictionary (Fig.5). Tags represent classes (elements having the same names of classes) and attributes are their correspondent ones.

3.2.3 Conversion and generation of LMF-based XMLized dictionary

Given such types of information of the dictionary which are « representative and semantic components » of sense of terms of hadith, are identified and the correspondent LMF model is built, conversion of the conceptual normalized dictionary into LMF-based XML format will be automatic and direct by generation and correspondence. A portion of this new format of the dictionary is shown in the following Fig.5.

```
<?xml version="1.0" encoding="UTF-8"?>
<Lexicon>
<feat att="language" val="arab" />
<LexicalEntry id="أبد">
<feat att="type" val="root" />
<Lemma>
<feat att="WittenForm" val=" أبد " />
 </Lemma>
<RelatedForm targets="1أبد">
 <feat att="type" val="derivedForm" />
</RelatedForm>
<RelatedForm targets="2أبد"> <feat att="type"
val="derivedForm"/>
</RelatedForm>
<RelatedForm targets="3أبد"> <feat att="type"
val="derivedForm"/>
</RelatedForm>
</ LexicalEntry>
<LexicalEntry id="1أبد">
<feat att="PartOfSpeech" val="CommonNoun" />
<Lemma>
<feat att="WittenForm" val="آبد" />
<feat att="type" val="Active Participle"/>
<feat att="scheme" val="فاعل"/>
<feat att="GrammaticalNumber" val="singular"/>
</Lemma>
< WordForm >
<feat att="WittenForm" val=" آبدة" />
<feat att="scheme" val="فاعلة"/>
<feat att="GrammaticalGender" val="feminine"/>
<feat att="GrammaticalNumber" val="singular"/>
</ WordForm >
<WordForm>
<feat att="WittenForm" val=" أوابد" />
<feat att="scheme" val="فواعل"/>
<feat att="Grammatical number" val="plural"/>
</WordForm>
<Sense id="1أبدة">
<SenseRelation targets="2أبدة">
<feat att="label" val="homonym">
</SenseRelation>
<SenseRelation targets="3أبدة">
<feat att="label" val="homonym">
</SenseRelation>
<Context>
<feat att="text" val=" قَالَ رَافِعُ بْنُ خَدِيجٍ: أَصَبْنَا نَهْبَ إِبِلٍ فَنَدَّ مِنْهَا
بَعِيرٍ فَرَمَاهُ رَجُلٌ بِسَهْمٍ فَحَبَسَهُ، فَقَالَ رَسُولُ اللهِ صَلَّى اللهُ
عَلَيْهِ وَسَلَّمَ: إِنَّ لِهَذِهِ الْإِبِلِ أَوَابِدَ كَأَوَابِدِ الْوَحْشِ، فَإِذَا غَلَبَكُمْ مِنْهَا شَيْءٌ
فَافْعَلُوا بِهِ هَكَذَا " />
<feat att="source" val="حديث نبويّ شريف" />
</Context>
<Definition>
<feat att="text" val="  هي التي قد تأبّدت" />
</Definition>
< /Sense">
<Sense id="2أبدة">
<Context>
<feat att="text" val=" مِنْهُ حَدِيثُ أُمِّ زَرْعٍ فَأَرَاحَ عَلَيَّ مِنْ
كُلِّ سَائِمَةٍ زَوْجَيْنِ، وَمِنْ كُلِّ آبِدَةٍ اثْنَتَيْنِ " />
<feat att="source" val="حديث التابعين" />
</Context>
<Definition>
<feat att="text" val=" أَنْوَاعًا مِنْ ضُرُوبِ الْوَحْشِ " />
</Definition>
< /Sense">
<Sense id="3أبدة">
<Context>
<feat att="text" val=" قَوْلُهُمْ: جَاءَ بِآبِدَةٍ " />
<feat att="source" val="حديث التابعين" />
</Context>
<Definition>
<feat att="text" val=" أَمْرٍ عَظِيمٍ يُنْفَرِ مِنْهُ ويُسْتَوْحَشُ
" />
</Definition>
< /Sense">
<RelatedForm targets="أبد">
<feat att="type" val="root" />
</RelatedForm>
<RelatedForm targets="2أبد">
<feat att="type" val="tronc"/>
</RelatedForm>
</LexicalEntry>
<LexicalEntry id="2أبد">
<feat att="PartOfSpeech" val="Verb" />
<Lemma>
<feat att="WittenForm" val="أبد" />
<feat att="scheme" val="فعل"/>
</Lemma>
<WordForm>
<feat att="WittenForm" val=" تَأَبَّدَتْ" />
<feat att="scheme" val="تَفَعَّلَت"/>
<feat att="GrammaticalNumber"
val="singular"/>
<feat att="GrammaticalGender"
val="feminine"/>
<feat att="GrammaticalVoice" val="active"/>
<feat att="Person" val="3" />
</WordForm>
<Sense id=" 1تَأَبَّدَتْ ">
<SenseRelation targets="2تَأَبَّدَتْ ">
<feat att="label" val="synonym">
</SenseRelation>
<Definition>
<feat att="text" val="  تَوَحَّشَتْ" />
</Definition>
< /Sense">
<Sense id=" 2تَأَبَّدَتْ">
<Definition>
<feat att="text" val=" نَفَرَتْ مِنَ الإِنْسِ " />
</Definition>
< /Sense">
<WordForm>
<feat att="WittenForm" val=" أَبَدَتْ " />
<feat att="scheme" val="فعلَتْ"/>
<feat att="GrammaticalNumber"
val="singular"/>
<feat att="GrammaticalGender"
val="feminine"/>
<feat att="Person" val="3" />
</WordForm>
<WordForm>
<feat att="WittenForm" val=" تَأْبِدُ " />
<feat att="scheme" val="تَفْعِلْ"/>
<feat att="GrammaticalNumber"
val="singular"/>
<feat att="GrammaticalGender"
val="feminine"/>
<feat att=Person" val="3"/>
<feat att="GrammaticalNumber"
val="singular"/>
<feat att="VerbFormMood" val="Indicative"/>
</WordForm>
<WordForm>
<feat att="WittenForm" val=" تَأْبُدُ " />
<feat att="scheme" val="تَفْعُلْ"/>
<feat att="GrammaticalNumber"
val="singular"/>
<feat att="GrammaticalGender"
val="feminine"/>
<feat att=Person" val="3"/>
<feat att="VerbFormMood" val="Indicative"/>
</WordForm>
</LexicalEntry> </Lexicon>
```

Figure 5 : Extract of LMF based normalized dictionary in XML.

As seen in the XMLized LMF format of the dictionary, it is now easy to enumerate for a term all its forms and senses. Even more, it is possible to recognize relations between senses of the same lexical entry or between senses of different lexical entries. This will make the construction of the ontology easier. We must note that the process of ontologization is iterative

and incremental. Also, we will not take terms arbitrarily, but from the terminologies of domains of hadith as for example terminology of prayer (الصلاة) or terminology of purification (الطهارة) which means that we will construct ontology of hadith truly composed of different modular domain ontologies of hadith.

4. Conclusion

We presented in this paper a work in progress, which aims to structure and formalize hadith Dictionaries. The main goal is to propose a generic normalization approach for Arabic Dictionaries and especially for Dictionaries of hadith as a first step before passing to the construction of the ontology and its integration into an IR system which is our semantic portal of hadith to be used for AWSD. This paper allowed us to present the first experiments in analyzing hadith dictionaries through contextual explorations and to identify the main tools, steps and classes of the LMF standard to be used. This effort will lead to a reusable knowledge of rich linguistic resource, which will be extended and used in several contexts.

References

Alrahabi, M. (2010). EXCOM-2: plateforme d'annotation automatique de catégories sémantiques. Applications à la catégorisation des citations en Français et en Arabe. Thèse de doctorat, sous la direction du Prof. Jean Pierre Desclès, Université Paris-Sorbonne.

Ayed R., Bounhas, I., Elayeb, B., Evard, F. and Bellamine Ben Saoud, N. (2012). A Possibilistic Approach for the Automatic Morphological Disambiguation of Arabic Texts. In Proceedings of 13th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), IEEE Computer Society, August 08-10, Kyoto, Japan, 187-194.

Ben Amar Baccar, F., Gargouri, B. and Ben Hamadou, A. (2011). Domain ontology generation using LMF standardized dictionary structure. The 6th international conference on software and Data Technologies (ICSOFT2011), July 18-21, Seville, Spain.

Ben Amar Baccar, F., Gargouri, B. and Ben Hamadou, A. (2013). Domain Ontology Enrichment Based on the Semantic Component of LMF-Standardized Dictionaries. The 6th International Conference on KSEM, August 10-12, Dalian, China.

Beseiso, M., Ahmad, A.R. and Ismail, R. (2010). A Survey of Arabic language Support in Semantic web. International Journal of Computer Applications 9(1), 35–40.

Bounhas, I., Elayeb, B., Evrard, F.and Slimani, Y. (2011a).Organizing contextual knowledge for arabic text disambiguation and terminology extraction. Knowledge Organization Journal. Ergon Verlag of Würzburg. Vol. 38 N. 6, 473-490.

Bounhas, I., Elayeb, B., Evrard, F.and Slimani, Y. (2011b). ArabOnto: Experimenting a new distributional approach for Building Arabic Ontological Resources. International Journal of Metadata, Semantics and Ontologies (IJMSO). Volume 6 Issue 2, May 2011, Inderscience Publishers, Geneva, Switzerland, 81-95.

Bounhas, I. (2012). Construction et intégration d'ontologies pour la cartographie socio-sémantique de fonds documentaires arabes guidée par la fiabilité de l'information. Thèse de doctorat en informatique, Faculté des Sciences de Tunis, Tunisie.

Francopoulo, G. (2013). LMF Lexical Markup Framework. Etats Unis: Wiley-ISTE

Jarrar, M. (2011). Building a Formal Arabic Ontology Methodology and Progress. In Proceedings of the Experts meeting on Arabic Ontologies and Semantic Networks. Alecso.Arab League, April 26-28, Tunis, Tunisia.

Khemakhem, A., Elleuch, I., Gargouri, B. and Ben Hamadou, A. (2009). Towards an automatic conversion approach of editorial Arabic dictionaries into LMF-ISO 24613

standardized mode. The 2nd International Conference on Arabic Language Resources and Tools (MEDAR), April 22-23, Cairo, Egypt.

Mhiri, M., Gargouri, F. and Benslimane, D. (2006). Détermination automatique des relations sémantiques entre les concepts d'une ontologie, In Proceedings of INFORSID'2006. pp.627-642, June 01-03, Hammamet, Tunisia.

Navigli, R. (2009). Word sense disambiguation: A survey. ACM Computing Surveys (CSUR) 41(2), 1-69.

Zaidi Ayad, S.(2013). Une plate-forme pour la construction d'ontologie en arabe : Extraction des termes et des relations à partir de textes (Application sur le Saint Coran). Thèse de doctorat. Université Badgi Mokhtar-Annaba, Algérie.

Zouaghi, A., Merhbene, L.and Zrigui, M. (2012). Combination of information retrieval methods with LESK algorithm for Arabic word sense disambiguation. Artificial Intelligence Review 38 (4), 257-269.