Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 5, 1315-1328 (2020)
www.astesj.com
Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Evaluating the Impact of Semantic Gaps on Estimating the Similarity Using Arabic Wordnet

Mamoun Abu Helou *

Management Information System Department, Al-Istiqlal University, Jericho, 11590, Palestine

ARTICLE INFO

Article history:
Received: 30 August, 2020
Accepted: 21 October, 2020
Online: 26 October, 2020

Keywords:
Lexical ontologies
Arabic wordnet
Semantic gaps
Semantic structure
Semantic similarity measure

ABSTRACT

Knowledge-based approach is wield used in various NLP applications. For example, to evaluate the semantic similarity between words, the semantic evidence in lexical ontologies (wordnets) is commonly used. The success of the English WordNet (EnWN) in this domain has inspired the creation of several wordnets in different languages, including the Arabic WordNet (ArWN). The English synsets have been extended to Arabic synsets through translation, which have introduced semantic gaps in ArWN structure. Therefore, compared to EnWN, ArWN has limited coverage in terms of lexical and semantic knowledge. This paper explores to what degree the richness of the wordnets' semantic structure influences the semantic evidence that can be used in wordnet-based applications, in particular the effect of filling the semantic gaps in ArWN. The paper studies the performance of applying English-based and Arabic-based similarity measures over ArWN. A set of experiments was performed by applying six path-based semantic similarity measures over Arabic benchmark dataset to investigate the usability and efficacy of the enriched structure of ArWN. The Performance measures, Person Correlation and Mean Square Error, are computed against and compared to human judgment benchmark. The obtained results demonstrate that the semantic similarity between words can be significantly improved when filling the semantic gaps. In addition, the experiment findings show that Arabic-based measures competitively perform well compared to the English-based measures. Further, ArWN enhanced structure is also available for public.

1 Introduction

In Natural Language Processing applications, a common task is to estimate the semantic similarity among words [1]. Lexical resources, such as, bilingual and multilingual dictionaries, thesauruses, lexical ontologies (wordnets), machine translation services among others, are widely used to estimate the similarity [2]. For instance, various tasks of natural language processing, knowledge engineering, and computational linguists have exploited the lexical and semantic knowledge encoded in the English WordNet (EnWN) [3, 4]; including sense disambiguation, information retrieval, text summarization, and question answering [5]–[6].

EnWN has been expanded to provide multilingual knowledge in many wordnet projects [7]–[8]. The Arabic WordNet (ArWN) [9] has extended EnWN by translating English synsets. However, English synsets that do not have translation in Arabic introduce *semantic gaps* in ArWN's semantic structure. For instance, synsets containing a single and polysemous word are difficult to determine their meaning by means of direct translation; in fact, more evidence is required to disambiguate their meaning [10]–[11]. Thus, similarity measures designed for English (i.e., English-based similarity measures) may not be effective in the same way when applied over resources in other languages; in this work we consider Arabic language.

Experiment findings in [12] showed that ArWN has limited coverage of lexical and semantic knowledge compared to EnWN. Further attempts have been made to improve the content of ArWN [9], [13]–[14]. However, resolving the semantic gaps was not considered. In [15, 16] they studied the performance of different similarity measures over ArWN. However, no explicit configuration was stated when calculating the similarity scores. Further, no explanation was given on how some semantic similarity scores were reported.

In [17], a preliminary study was conducted to examine the impact of the semantic gaps on estimating the semantic similarity scores using ArWN. They examined the impact of improving the semantic structure of ArWN on estimating the similarity between Ara-

*Corresponding Author: Mamoun Abu Helou, MIS Dept., Al-Istiqlal University, mabuhelou@pass.ps.

bic synsets. The semantic gaps were analyzed and identified. Then new synsets in Arabic were added to ArWN and mapped to their corresponding synsets in English, using interactive cross-lingual mapping approach [18]. The impact of the enriched ArWN was studied in semantic similarity experiment using only one English-based semantic similarity measure.

In this paper we extend previous work presented in [17]; a large scale experiment is conducted to further examine the degree to which wordnet-based applications can be influenced by improving their semantic structure, mainly considering ArWN. In particular, the main contributions of this work can be summarized as follow.

(i) Four settings are defined and applied over two variants of ArWN structure. In the experiment six path-based similarity measures are applied over ArWN and EnWN; including, four English-based similarity measures (Path [19], Li [2], WuP [20], and Lch [21]), and two Arabic-based similarity measures (AWSS [22], and Aldirey [16]).

(ii) Study to which extent the semantic similarity measures that are developed for Arabic-based applications can perform efficiently well compared to English-based similarity measures. A comprehensive comparison between the similarity measures over the different configurations is provided, for both EnWN and ArWN.

The similarity scores obtained from the different measures, in the different settings, are compared to a standard benchmark for Arabic word pairs obtained from the AWSS dataset [23]. Two measures, the Person Correlation and the Mean Square Error measures, are used to quantify the performance of the similarity measures. Reported values indicate the importance of the semantic evidence obtained from the enrichment process, and its significant effect on estimating the semantic similarity between words. In addition, the results show that Arabic-based measures performs competitively good compared to English-based measures.

The rest of this paper is organized as follows. Section 2 overviews related works on building wordnets, and the development of wordnet-based semantic similarity measures. Section 3 and describes the approach used to evaluate the impact of Semantic Gaps on estimating the Similarity over ArWN. Section 4 discusses experiments conducted: the benchmark dataset, the performance measures, and the obtained results. Finally Section 5 draws some conclusions and outlines future work.

2 Related works

This section provides an overview of the construction of wordnets and the ArWN contents; presents wordnet-based semantic similarity measures, which will be used in the experiment.

2.1 Wordnets overview

Wordnets, also known as lexical ontologies [24], are considered to be a resource of lexical and semantic knowledge, which organize natural language words (lexicons) into synsets. A synset is a collection of synonym words that express one meaning in a specific context (i.e., concept) [3, 25].

In wordnets, words are arranged in a lexical database. Words can have several senses, such that each sense of a given word is identified by a number and its part of speech type. For instance, the sense *village*#*n*#2 indicates the second (#2) nominal (#*n*) sense of the word "*village*". Words are linked through lexical relations, for example, *antonym* and *synonymy* relations. When a word can have more than one meaning, it is called *polysemous word*, which can be member of several synsets. Otherwise, it is called *monosemous word*, which is a member of a single synset. For example, the word "*village*" has three noun senses as defined in EnWN; which are indicated in the following set of synsets:{{village#n#1, small-town#n#1, settlement#n#2}, {village#n#2, hamle#n#3}, {Greenwich-village#n#1, village#n#3}}.

Synsets are related by semantic relations. The *Hypernymy* and *Hyponymy* relations are considered to be the key semantic relations that form the semantic structure in wordnets. Hypernymy is described as the inverse of Hyponymy. For instance, in Figure 1 the synset {village#n#2, hamle#n#3} is hypernymy of the synset {settlemt#n#6}, while the synset {settlemt#n#6} is hyponymy of the synset {village#n#2, hamle#n#3}. Further, definitions (glosses) are also attached to synsets to convey their meaning. For example, the word sense *village*#*n*#2 defined as "a settlement smaller than a town" [1].

The *HyperTree* of a given synset (i.e, word sense) is defined as the sequence of synsets that are linked with hypernymy relations, which connect a synset with its ancestor synsets up to the root node. The function $HyperTrees(word)$ produces the set of HyperTrees which a given word belongs. Figure 1 shows an excerpt of nominal HyperTrees in English and their correspondence in Arabic [2].

EnWN has been manually produced at Princeton University over the past three decades [3, 4]. EnWN 's success in many computational language domains has inspired the development of similarly structured lexicons, for both individual and multiple languages [26], such as EuroWordNet [7], BalkaNet [27], Polylingual WordNet [8], universal wordnet [28],MultiWordNet [29], WikiNet [30],and Arabic WordNet [9].

Computational linguistics has defined the Inter-Lingual Index [7], to establish links between different wordnets which is considered to be independent of language. For instance, near-equivalence and equivalence semantic relations are used to link synsets from the individual wordnets to the Inter-Lingual Index. Wordnets for several languages have been developed under the guidance of the Global WordNet Association [3], which seeks to organize the creation and linking of wordnets. Further, the Open Multilingual WordNet project [31] offers access to open wordnets in a number of languages, which are all connected to the latest version of EnWN (v3.0) [4].

[1] Definitions can be accessed at http://wordnetweb.princeton.edu/perl/webwn.
[2] In the following, for both Arabic and English senses, the *pos* (*n*) identification is removed for readability, as they all nominal senses.
[3] http://globalwordnet.org/
[4] http://compling.hss.ntu.edu.sg/omw/.

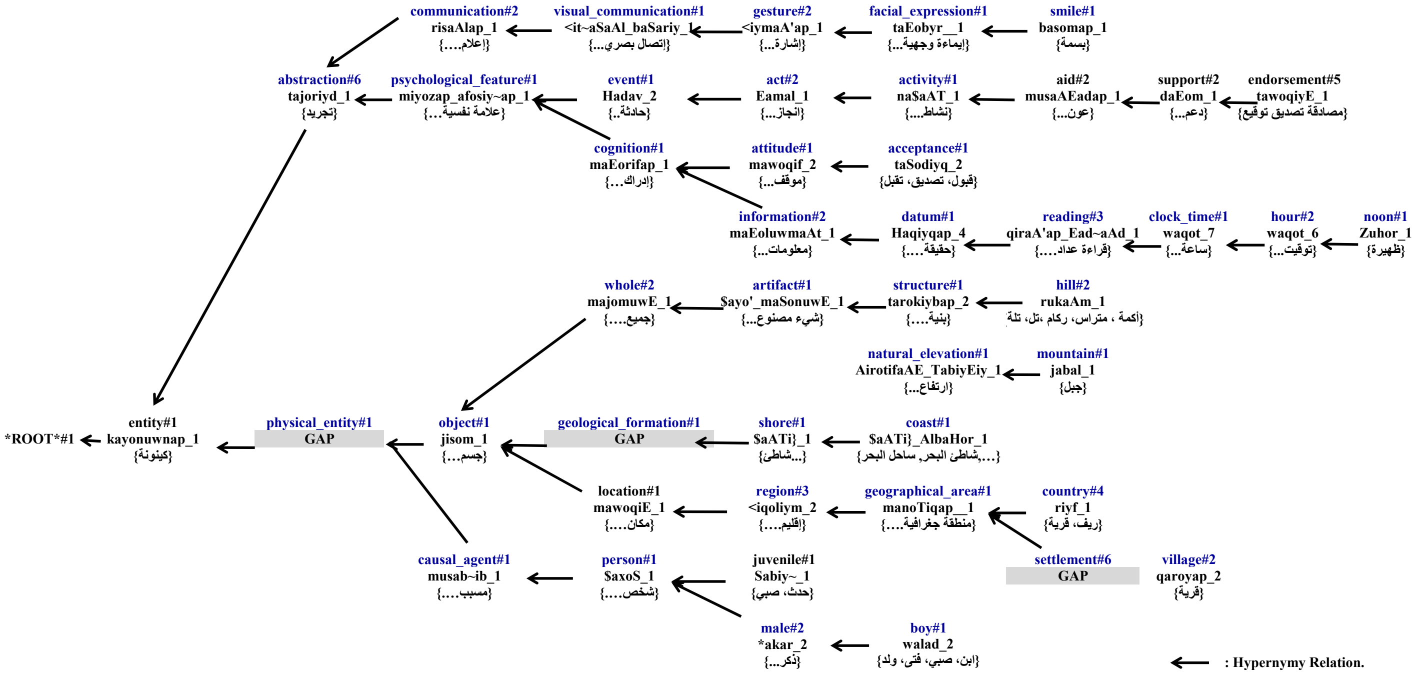

Figure 1: An excerpt of nominal HyperTrees from EnWN and its correspondences in ArWN.

2.2 *Arabic wordnet contents*

In the construction of ArWN [9], the extend method has been adopted. English Synsets have been translated into Arabic; and the structure of the EnWN (v2.0) has been inherited by ArWN . In the release of ArWN (v2.0) [5], 23,841 Arabic words, such as broken plurals, Named Entities, and roots have formed 11,296 synsets. Twenty-two types of semantic relationships have been used to connect synsets that formed 161,705 semantic links. Consequently, and in comparison with EnWN, which contains 147,306 words (117,659 synsets) [6]; one can observe that ArWN has a limited coverage in terms of semantic relations and lexicons [12].

To this end, many attempts have been made to enhance the quality of ArWN by expanding its lexical coverage [13, 9] or semantic relationships [32, 14] by different approaches. In [32] they released their work under the Lexical Markup Framework. However, the public release of ArWN ignores the synsets that are not linked to EnWN [31]. Nevertheless, synsets (semantic gaps) which are resolved in this work will be made for public [7] . In future work we plan to compile an *xml* format of ArWN enhanced structure, to enable researcher to utilize the ArWN in different applications.

2.3 *Wordnet-based similarity measures*

In linguistics, philosophy and information theory, estimating the semantic similarity between concepts is extensively studied [2, 15], which is a common and crucial task in many NLP applications, text summarization, word sense disambiguation, entailment, machine translation, among many others [33]–[6], [34, 35].

Estimating the semantic similarity between words is calculated by measuring the similarity between concepts (synsets) associated with the words [2]. Given two words, one can calculate the semantic similarity by exploiting wordnet (i.e, a lexical knowledge base). The lexical and semantic knowledge in wordnet have been used in many semantic similarity measures, which are originally designed and evaluated over EnWN (English-based measures) [36, 2].

In [15] they defined four broad categories of the similarity measures; Path-based similarity measures [2, 16, 19, 20, 21, 22]; information content similarity measures [37, 38]; feature-based similarity measures [39]; and hybrid similarity measures [40, 41]. There have been few works concerned with the similarity of Arabic; AWSS measure [22] and Aldiery measure [16]. These have mainly adapted measures from those constructed for English. In particular, Li measure [2] was adapted, which is a path-based measure that consider the depth of concepts in the HyperTrees; the distance between two compared concepts; and the depth of the least common concept (*lsc*) that subsumed two compared concepts. Noting that, these measures needs to tune weighting parameters to find the optimal values [22, 16]. In this regards, several preliminary experiments are necessary to find the best weights that provide the optimal values.

An attempt to investigate the performance of the similarity measures over ArWN was conducted in [15]. They studied the performance of seven measures; including AWSS measure [22]. All measures were applied over 40 word pairs that are selected from AWSS dataset [23], which are also considered as the benchmark dataset in this work. The experiments findings [15] showed that

[5]ArWN obtained from the AWN browser available at: AWNBrowser; or as XML file available at: Open Multilingual Wordnet
[6]EnWN-3.0 available at: https://wordnet.princeton.edu/download
[7]Full list of the Semantic Gaps available at:https://fas.alistiqlal.edu.ps/cv-2-ar.html

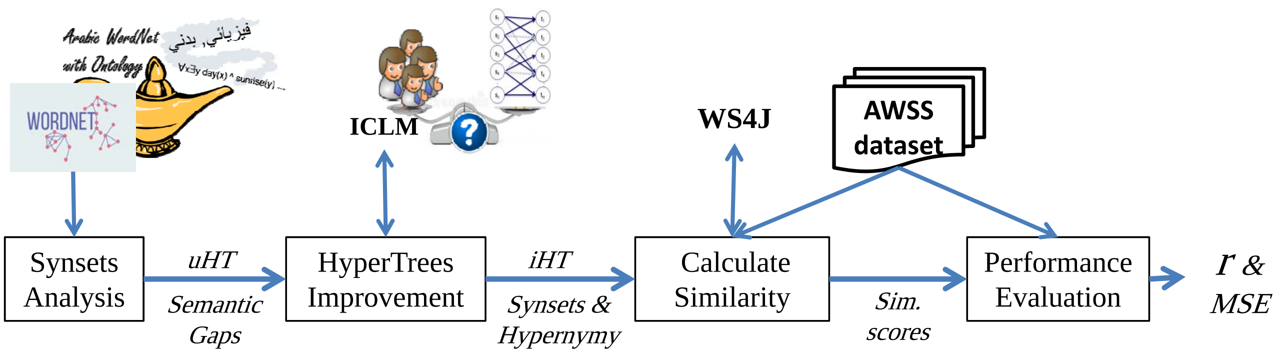

Figure 2: The adopted approach overview

WuP measure [20] has the best performance in estimating the semantic similarity between Arabic word pairs. The experiments in [16] also introduced a competitive Arabic-based similarity measures (Aldiery measure) in comparison to WuP measure.

In [17] they further studied the impact of enhancing the HyperTree over the Wup measure. This work adopted and extend their experimental configurations and examine further the impact of the enhanced semantic structure of ArWN over Six measures including English and Arabic path-based measure, further details are provided in Section 4.

Recall that, for a given concepts c_i and c_j, the function $Sim_m(c_i, c_j)$ calculates the semantic similarity between c_i and c_j, where m indicates the name of the measure. Next the description of the measures used in the experiment is given.

1. *Path* measure [19] finds the shortest path between the two concepts, by counting the number of edge (hypernymy relation) between the concepts, in order to compute the semantic similarity. Path measure which is considered as the pioneer similarity measure is defined in equation (1).

$$sim_{path}(c_i, c_j) = \frac{1}{len(c_i, c_j)} \quad (1)$$

Where the length function, $len(c_i, c_j)$, returns the length of the shortest path between c_i and c_j in the wordnet semantic hierarchy. For example, in Figure 1, $len(hill\#2, mountain\#1) = 3$, and $Sim_{path}(hill\#2, mountain\#1) = 0.333$.

2. *Wup* measure [20] calculates the similarity by computing the distance between the two concepts and the maximum depth of the least common concept (lsc) that subsumed the two concepts under evaluation. WuP measure is defined in equation (2).

$$sim_{WuP}(c_i, c_j) = \frac{2 * d(lcs(c_i, c_j))}{d(c_i) + d(c_j)} \quad (2)$$

Where $d(c_i)$ is the depth of the concept c_i using edge counting in the semantic hierarchy, $lcs(c_i, c_j)$ is the least common subsumer of c_i and cj, $d(lcs(c_i, c_j))$ is the maximum length between lcs of c_i and c_j and the *root* of the hierarchy, where $d(entity) = 1$. For example in Figure 1, $d(hill\#2) = 7$, $d(mountain\#1) = 7$, $d(lcs(hill\#2, mountain\#1) = 6$, and $Sim_{Wup}(hill\#2, mountain\#1) = 0.857$.

3. *Lch* measure [21] uses the length of the shortest path between the two concepts, and also the maximum depth of the semantic hierarchy of a given part of speech type. Lch measure is defined in equation (3).

$$sim_{Lch}(c_i, c_j) = -\log \frac{len(c_i, c_j)}{2 * maxDepth_{pos}} \quad (3)$$

Where, $maxDepth_{pos}$ is the maximum depth of the hypernymy structure for a given part of speech. For instance, $maxDepth_n$ is 20 and 15 in EnWN and ArWN, respectively. For example in Figure 1, $Sim_{Lch}(hill\#2, mountain\#1) = -log(3/2 * 20) = 2.590$.

Noting that the Lch scores reported in Section 4 are normalized into the range 0 to 1 by dividing Lch scors over 3,688, Hence, $Sim_{Lch}(hill\#2, mountain\#1) = 0.702$.

4. *Li* measure [2] computes the similarity using non-linear function, which consumes the shortest length between concepts and the minimum depth of the concepts in the semantic hierarchy. Li measure is defined in equation (4).

$$sim_{Li}(c_i, c_j) = e^{\alpha * len(c_i, c_j)} \frac{e^{\beta * d(lcs(c_i, c_j))} - e^{-\beta * d(lcs(c_i, c_j))}}{e^{\beta * d(lcs(c_i, c_j))} + e^{-\beta * d(lcs(c_i, c_j))}} \quad (4)$$

Noting that the parameters α and β need to be calculated manually for good performance. The optimal parameters are $\alpha = 0.2$ and $\beta = 0.6$ as reported in [2]. For example, $Sim_{Li}(hill\#2, mountain\#1) = 0.548$.

5. *AWSS* measure [22] is an Arabic-based measure that adapted Li measure to compute semantic similarity with modification on the depth and length computation to be proper for ArWN [23]. AWSS measure is defined in equation (5).

$$sim_{AWSS}(c_i, c_j) = e^{-\alpha * d(lcs(c_i, c_j))} * \tanh(\beta * len(c_i, c_j)) \quad (5)$$

Where the parameters α and β are the length and depth factors respectively. The optimal performance was obtained at $\alpha = 0.162$ and $\beta = 0.234$ as reported in [22]. For example in Figure 1, $len(rukaAm\_1, Jabal\_1) = 4$ and $d(lcs(rukaAm\_1, Jabal\_1)) = 8$, then $Sim_{AWSS}(rukaAm\_1, Jabal\_1) = 0.201$.

Table 1: Semantic gaps frequency distribution in ArWN for nominal synsets

Semantic Gaps	1	1	1	1	1	1	1	2	3	1	1	2	1	6	3	3	2	11	1	7	12	11	15	**88**
Freq	4525	187	100	50	48	46	36	30	24	17	15	14	12	10	9	8	7	6	5	4	3	2	1	**5493**

6. *Aldiery* measure [16] is an Arabic-based measure also adapted Li measure to compute semantic similarity with modification on the depth and length computation to be proper for ArWN. Aldiery measure is defined in equation (6).

$$sim_{Aldiery}(c_i, c_j) = \tanh(\frac{2 * d(lcs(c_i, c_j))}{d(c_i) + d(c_j)} + W(\frac{1}{len(c_i, c_j)} + \frac{\log(len(c_i, c_j))}{\log(maxDepth_{pos})})) \tag{6}$$

Where [16] defines $W = 0.5$. For example in Figure 1, $d(rukaAm\_1) = 8$, $d(Jabal\_1) = 7$, $len(rukaAm\_1, Jabal\_1) = 4$, and $d(lsc(rukaAm\_1, Jabal\_1)) = 8$, and $maxDepth_n = 15$, then $Sim_{Aldiery}(rukaAm\_1, Jabal\_1) = 0.692$.

Noting that, the similarity functions defined above consume either words, or word senses as parameters. In the first case, the similarity function returns the highest similarity score for all the possible combination of word senses for the two given words. In the second case, it returns the similarity score between the two defined senses.

In addition, the six measures defined in the equations (1,2,3,4,5, and 6) are path-based measures, this study focus on the impact of the structure without interference of other semantic evidence such as features extracted from corpuses, which depend on the quality of the used cuprous, as well as the availability of resources in Arabic.

On the other hand, Path, WuP, and Lch measures are considered as linear path-based measures, while Li measure is a non-linear path based measure. AWSS and Aldiery are also non-linear path based measures, which are derived from Li and purposely developed for Arabic.

Observe that, for the Path, Wup, and Lch measures no weights are required to be tuned. While the other measures need to find optimal value of the defined weights. The four English-based measures, as well as the two Arabic-based measures are selected because they achieved good performance against other measures [22, 16], and to compare the performance between the measures using Arabic benchmark dataset.

3 Evaluating the impact of semantic gaps on estimating the similarity

This section presents the approach that is used to evaluate the impact of enhancing the structure of ArWN on estimating the semantic similarity. Figure 2 illustrates the main phases of the approach, which are explained as follow.

1. **Synset Analysis**. In this phase the semantic gaps are identified through a comparison between the structures of ArWN (v2.0) and EnWN (v3.0). In particular, for each nominal synset in ArWN, Hypernymy relations are compared with their EnWN correspondences. The HyperTrees for each synset in ArWN is compared with its correspondence HyperTrees in EnWN. For example, Figure 1 indicates two semantic gaps in the ArWN *HyperTree*(*$aATi}_AlbaHor_1*, ساحل) ={*ROOT*#1 , kayonuwnap_1, *GAP*, jisom_1, *GAP*, $aATi}_1, $aATi}_AlbaHor_1} [8], where the correspondence HyperTree in EnWN is, HyperTree(coast#1) = {*ROOT*#1, entity#1, physical entity#1, object#1, geological formation#1, shore#1, coast#1}.

   In total, [17] reported that 5,493 (69%) of the 7,960 nominal synsets in ArWN have at least one semantic gap. In particular, compared to the structure of EnWN, the semantic gaps have been resulted from the missing of 88 synsets in ArWN.

   The distribution frequency of the semantic gaps in ArWN is reported in Table 1, "Semantic Gaps" refers to the number of synsets that have the reported freq, and "Freq" indicates the number of HyperTrees that have at least one semantic gap. For instance, the first column reports an English synset ({"physical-entity#1"}) that has no correspondence in Arabic, introduces 4,525 semantic gaps in ArWN. While the 8^{th} column indicates two synsets ({"armed-service#1",...}, and {"health-care-provider#1",...}), each introduces 30 semantic gaps in ArWN. Last column reports the totals.

2. **HyperTrees Improvement**. In this phase ICLM Web application [18] is used to fill the identified semantic gaps. ICLM is a semi-automatic matching approach that supports feedback provided by multiple users. In ICLM the number of users that are asked to perform each mapping task is estimated based on the lexical characterization of concepts under evaluation, i.e., on the estimation of the ambiguity conveyed by the concepts involved in mappings [42], with the assumption that as the selection tasks difficulties increase, more users agreement is required.

   The candidate matching of the source concepts in Arabic are automatically computed to the English target concepts using a lexical based disambiguation algorithm [43]. The study [42] recommended that combining lexical resources improves the quality of translations and provide a valuable support for candidate match retrieval in cross-lingual ontology matching problems. Accordingly, translations of the missing synsets are collected by combining lexical knowledge from different external resources. English synset translation was

[8] Represents the HyperTree of the first nominal sense for the word ساحل.

[9] https://translate.google.com/

[10] https://babelnet.org/

[11] https://www.almaany.com/ar/dict/ar-en/

Table 2: Top ten Frequent Semantic Gaps in ArWN with EnWN correspondence sysnets

	EnWN synset	Freq	Semantic gaps in ArWN
1	{physical entity}	4,525	{كيان مادي}
2	{substance}	187	{جوهر، مادة}
3	{defender, guardian, protector, shielder}	100	{مدافع ، وصي، حارس، حامي، مدافع، ولي}
4	{variety, assortment, miscellanea,...}	50	{تشكيلة،تنوع، مجموعة منوعة،منوعات}
5	{aristocrat,blue blood, patrician}	48	{رفيع الجناب، نبالة، نسب أصيل، النبيل، الأرستوقراطي}
6	{formataion, geological formation}	46	{تشكيل ، التكوين الجيولوجي ، التشكلات الجيولوجية}
7	{deceiver, beguiler, cheat, cheater, ...}	36	{مخادع، مضل ، غشاش ، محتال ، دجال}
8	{armed service, service, military service}	30	{القوة المسلحة ، الخدمة العسكرية}
9	{health care provider, health professional,...}	30	{مقدم عناية صحية ، مزود الرعاية الصحية الأولية، الصحة المهنية}
10	{wrongdoer, offender}	24	{مذنب، مجرم ، ظالم ، معتد، آثم}

collected from; Google Translate[9], BabelNet[10], and Almaany dictionary[11].

The difficulties of the mapping selection tasks, that is determining the number of user which are asked to perform the task, are estimated using lexical characteristics of concepts under evaluation: Ambiguity of lexicalization, Synonym-richness, and Uncertainty in the selection Step. The mapping tasks are validated by some users based on a CAUTIOUS strategy. The task difficulty level is estimated as Low, Mid, and High level. One, three, or five users are asked to perform the Low, Mid, or High tasks, respectively.

In [17] ten users (bilingual speakers) are asked to validate the mapping tasks, that is, to fill a semantic gap in ArWN, and accordingly define new link with EnWN, hence, import the semantic relations among the concepts. The top ten frequent semantic gap are listed in Table 2. As a result 94% of the identified gaps are resolved, that is more than 98% of HyperTrees are filled in.

Observe that, some concepts are hard to resolve, and more evidences are needed. For Example, {mechanism#3}, {attache#1}, and {climber#1} synsets, which contain a single and polysemous word, are hard to determine their meaning with direct translation and no context [42], for this reason in the validation task users did not reach an agreement. Noting that, the semantic gaps for every word sense in the benchmark dataset used in the experiment are resolved.

3. **Calculate Similarity**. In this phase similarity measures defined in Section 2.3 are applied over the ArWN and EnWN using Arabic benchmark dataset (AWSS dataset [23]). Different configuration explained in Section 4.4 are applied to calculate the semantic similarity using the WS4J online application (see Section 4.1). Similarity scores are reported and passed to the next phase.

4. **Performance Evaluation**. In this phase the obtained similarity scores are compared with Human Rating benchmark [23] using two performance measures; The Person Correlation measure (r) and the Mean Squared Error (MSE). Further details are provided in the experiment Section 4.3

4 Experiment

The conducted experiment aims at studying the efficacy of the semantic evidence in ArWN. In particular, the experiment focuses on the improvement of hypernymy relations in the semantic structure of ArWN. The experiment studies the extent to which the semantic structure of ArWN affects measuring the semantic similarity between concepts. This section reports and discusses the results obtained from running a set of configurations for measuring the semantic similarity scores over ArWN and EnWN.

Next sections present the tool which is used to calculate the semantic similarity scores, the benchmark dataset, the measures used to evaluate the performance of the structure improvement, and discuss obtained results.

4.1 Similarity Measure Tools

Significant efforts are being made in developing similarity measures to consume ArWN content. For example, the Java ArWN API [12]. The application consumes Arabic words with diacritics (vocalized), whereas the benchmark dataset in this experiment contains unvocalized (without diacritics) word pairs. If Arabic words are vocalized, similar to the work done in [16, 15], then their senses will be defined in advance. The experiment's configuration DS (see Section 4.4) studies the performance of determining the word senses on the similarity scores.

To avoid predefined senses, in this experiment the similarity scores are obtained using the $WS4J$ online application [13]. In computing the scores, $WS4J$ uses EnWN's semantic structure (v3.0), which is used to measure the similarity scores between Arabic words. Noting that, in this experiment Arabic senses under evaluation have the same structure of their correspondence senses in English, as the semantic gaps in ArWN has been improved and linked to EnWN(v3.0). The similarity scores between the Arabic concepts are then measured using their correspondence concepts in EnWN. In addition, $WS4J$ provides the description of all HyperTree of words under evaluation. The HyperTrees which returned for EnWN are validated to obtain Arabic words' HyperTrees with semantic gaps as depicted in Figure 1. For instance, this information is necessary to measure the similarity scores in uHT configuration,

[12]https://sourceforge.net/projects/javasourcecodeapiarabicwordnet
[13]http://ws4jdemo.appspot.com/

details are provided in Section 4.4.

4.2 Benchmark dataset

Similar to the work performed in [15, 16], the AWSS benchmark [22] will be used in this experiment. The obtained similarity scores will be compared with Human Judgments obtained from the dataset of AWSS [23]. The AWSS dataset contains 70 nominal word pairs of Arabic, divided into three similarity levels, Low, Medium, and High; 40 word pairs are selected and used in this experiment, listed in Table 3,which are also used in [15, 16].

Table 3: Arabic word pairs benchmark dataset

NO.	Sim. level	En Word Pairs		Ar Word Pairs	HR
1	Low	Coast	Endorsement	تصديق ساحل	0.01
2	low	Noon	String	خيط ظهر	0.01
3	low	Stove	Walk	مشي موقد	0.01
4	low	Cord	midday	حبل ظهيرة	0.02
5	low	Signature	String	خيط توقيع	0.02
6	low	Boy	Endorsement	تصديق صبي	0.03
7	low	Boy	Midday	ظهيرة صبي	0.04
8	low	Smile	Village	قرية ابتسامة	0.05
9	low	Noon	Fasting	صيام ظهر	0.07
10	low	Glass	Diamond	الماس كاس	0.09
11	low	Sepulcher	Sheikh	شيخ ضريح	0.22
12	low	Countryside	Vegetable	ريف خضار	0.31
13	mid	Tumbler	Tool	اداة قدح	0.33
14	mid	Laugh	Feast	ضحك عيد	0.34
15	mid	Girl	Odalisque	جارية فتاة	0.49
16	mid	Feast	Fasting	صيام عيد	0.49
17	mid	Coach	Means	وسيلة حافلة	0.52
18	mid	Sage	Sheikh	شيخ حكيم	0.56
19	mid	Girl	Sister	اخت فتاة	0.6
20	mid	Hen	Pigeon	حمامة دجاجة	0.65
21	mid	Hill	Mountain	جبل تل	0.65
22	mid	Master	Sheikh	شيخ سيد	0.67
23	mid	Food	Vegetable	خضار طعام	0.69
24	mid	Slave	Odalisque	جارية عبد	0.71
25	mid	Run	Walk	مشي جري	0.75
26	high	Cord	String	خيط حبل	0.77
27	high	Forest	Woodland	احراش غابة	0.79
28	high	Sage	Thinker	مفكر حكيم	0.82
29	high	Journey	Travel	سفر رحلة	0.84
30	high	Gem	Diamond	الماس جوهرة	0.84
31	high	Countryside	Village	قرية ريف	0.85
32	high	Cushion	Pillow	مخدة مسند	0.85
33	high	Smile	Laugh	ضحك ابتسامة	0.87
34	high	Signature	Endorsement	توقيع تصديق	0.89
35	high	Tools	Means	وسيلة اداة	0.92
36	high	Sepulcher	Grave	قبر ضريح	0.93
37	high	Boy	Lad	فتى صبي	0.93
38	high	Wizard	Magician	مشعوذ ساحر	0.94
39	high	Coach	Bus	حافلة باص	0.95
40	high	Glass	Tumbler	قدح كاس	0.95

Noting that, some words in the dataset benchmark are not covered in ArWN. For instance, the words "موقد" stove, "ساحر" wizard, and "مشعوذ" magician are not covered in ArWN, hence, the 3^{rd} and 38^{th} word pairs are not covered in the experiment. While, the words "ابتسامة" smile and "جوهرة" Gem, which are also not covered in ArWN, instead the words "بسمة" and "جوهر" are used to measure the similarity scores, respectively.

4.3 Performance Measures

The obtained similarity scores are evaluated against human ratings benchmark (*HR*), which is a human judgment similarity scores of Arabic nominal word pairs obtained from the dataset of AWSS.

Two measures are used to quantify the performance of the obtained similarity scores. The Person Correlation measure (r) defines the strength of the linear relationship between the obtained similarity scores and *HR*; the Mean Squared Error (MSE) calculates the average squared difference between the similarity scores and *HR*. The best performance is indicated by a similarity measure with the smallest MSE value and r value is close to 1. While the negative r value means that the obtained scores are increase as the *HR* ratings decrease. In addition, the similarity scores are compared to the performance results reported in [15, 16], which are listed in Table 4.

Table 4: Performance measures reported in [16, 15]

#	**Measure**	r	MSE
1	WuP	0.94	0.01648
2	LCH	0.89	0.03708
3	Path	0.75	0.16038
4	LI	0.85	0.10205
5	AWSS	0.88	0.04424
6	Aldiery	0.96	0.01893

4.4 Experimental settings

Six path-based semantic similarity measures, which are defined in equations (1,2,3,4,5, and 6), will be applied over the Arabic word pairs benchmark dataset, which is described in Section 4.2. Using the following configurations, the similarity measures are applied over ArWN and EnWN to quantify the efficiency of ArWN structure enrichment:

1. *UnDefined Senses* (uDS): calculates the semantic similarity between given words without determining their senses. In this setting, which is considered as the default setting of the similarity measures, the similarity measure returns the maximum score obtained from the all possible combination of the senses of the given words.

2. *Defined Senses* (DS): calculates the semantic similarity between given words senses (i.e, sense are determined in advance). By extending the work in [17], the sense of each word pairs under evaluation is determined based on a majority vote (consensus) approach. Similar to the tasks of filling the semantic gaps [17, 18] (see Section 3), the CAUTIOUS strategy is adopted, where users are avoided to decide among word pairs that share the same words.

3. *wordnets Translation* ($wnTrans$): calculates the semantic similarity over ArWN by selecting the senses that match the

Table 5: uDs configuration over ArWN

					iHT						uHT					
NO.	Ar Word Pairs Senses		En Word Pairs Senses		WuP	LCH	Path	LI	AWSS	Aldiery	WuP	LCH	Path	LI	AWSS	Aldiery
1	$aATi}_1	taSodiyq_2	shore#1	acceptance#1	0.308	0.298	0.100	0.113	0.086	0.451	0.364	0.358	0.125	0.168	0.119	0.504
2	mu&ax~irap_1	xayoT_1	back#2	cord#4	0.706	0.436	0.167	0.301	0.335	0.808	0.667	0.436	0.167	0.300	0.312	0.781
3																
4	Hamol_1	Zuhor_1	gestation#2	midday#1	0.316	0.207	0.071	0.058	0.063	0.504	0.316	0.207	0.071	0.058	0.063	0.504
5	tawoqiyE_1	daliyl_2	endorsement#5	lead#3	0.444	0.272	0.091	0.109	0.123	0.633	0.444	0.272	0.091	0.109	0.123	0.633
6	Sabiy~_1	taSodiyq_2	juvenile#1	acceptance#1	0.308	0.298	0.100	0.113	0.086	0.451	0.333	0.326	0.111	0.138	0.102	0.475
7	Sabiy~_1	Zuhor_1	juvenile#1	midday#1	0.235	0.207	0.071	0.051	0.045	0.379	0.250	0.227	0.077	0.062	0.053	0.394
8	basomap_1	qaroyap_1	smile#1	village#1	0.375	0.272	0.091	0.105	0.102	0.553	0.375	0.272	0.091	0.105	0.102	0.553
9	Zuhor_1	Sawom_1	noon#1	fasting#1	0.364	0.188	0.067	0.049	0.065	0.574	0.364	0.188	0.067	0.049	0.065	0.574
10	kuwb_1	AlomAs_1	glass#2	diamond#2	0.353	0.248	0.083	0.086	0.087	0.535	0.267	0.248	0.083	0.076	0.062	0.411
11	maqaAm_1	ra}iyos_1	shrine#1	head#4	0.500	0.272	0.091	0.110	0.139	0.687	0.556	0.326	0.111	0.164	0.192	0.720
12	riyf_1	xuDaAr_1	country#4	vegetable#1	0.375	0.272	0.091	0.105	0.102	0.553	0.286	0.272	0.091	0.092	0.073	0.429
13	sahom_3	adaAp_2	arrow#2	instrument#1	0.857	0.922	1.000	0.819	0.826	0.943	0.842	0.546	0.250	0.449	0.499	0.874
14	<iHotifaAl_1	DaHik_2	laughter#2	celebration#2	0.824	0.546	0.250	0.449	0.485	0.864	0.824	0.546	0.250	0.449	0.485	0.864
15	fataAp_1	xaAdim_1	girl#1	retainer#2	0.762	0.436	0.167	0.301	0.361	0.842	0.737	0.436	0.167	0.301	0.351	0.827
16	Eiyod_1	Sawom_1	celebration#2	fasting#1	0.700	0.395	0.143	0.246	0.298	0.811	0.700	0.395	0.143	0.246	0.298	0.811
17	HaAfilap_1	wasiylap_1	coach#5	means#2	0.778	0.486	0.200	0.368	0.412	0.845	0.750	0.486	0.200	0.367	0.394	0.828
18	fayolasuwf_1	ra}iyos_1	philosopher#1	head#4	0.762	0.436	0.167	0.301	0.361	0.842	0.737	0.436	0.167	0.301	0.351	0.827
19	fataAp_1	>xot_1	girl#1	sister#1	0.696	0.358	0.125	0.202	0.261	0.815	0.667	0.358	0.125	0.202	0.254	0.796
20	dajaAjap_1	HamaAm_1	hen#1	pigeon#1	0.828	0.436	0.167	0.301	0.376	0.879	0.815	0.436	0.167	0.301	0.374	0.872
21	rukaAm_1	jabal_1	hill#2	mountain#1	0.533	0.358	0.125	0.199	0.201	0.692	0.500	0.395	0.143	0.233	0.195	0.649
22	say~id_1	ra}iyos_1	sir#1	head#4	0.762	0.436	0.167	0.301	0.361	0.842	0.737	0.436	0.167	0.301	0.351	0.827
23	TaEaAm_3	xuDaAr_1	food#2	vegetable#1	0.857	0.624	0.333	0.548	0.545	0.875	0.833	0.624	0.333	0.546	0.507	0.861
24	xaAdim_1	xaAdim_1	retainer#2	retainer#2	1.000	0.922	1.000	0.819	0.835	0.958	1.000	0.922	1.000	0.819	0.826	0.957
25	jaroy_1	ma$oy_1	run#7	walk#1	0.909	0.624	0.333	0.549	0.604	0.905	0.909	0.624	0.333	0.549	0.604	0.905
26	Habol_1	gazol_1	cord#1	thread#1	0.941	0.734	0.500	0.670	0.690	0.918	0.933	0.734	0.500	0.670	0.671	0.913
27	dagol_1	dagol_1	jungle#1	jungle#1	1.000	0.922	1.000	0.818	0.754	0.950	1.000	0.922	1.000	0.815	0.701	0.947
28	fayolasuwf_1	mufak~ir_1	philosopher#1	intellect#3	0.900	0.624	0.333	0.549	0.597	0.900	0.889	0.624	0.333	0.549	0.587	0.894
29	riHolap_1	safar_1	journey#1	travel#1	0.952	0.734	0.500	0.670	0.710	0.926	0.952	0.734	0.500	0.670	0.710	0.926
30	HajarN_kariym_1	AlomAs_1	gem#2	diamond#2	0.875	0.624	0.333	0.549	0.570	0.886	0.857	0.624	0.333	0.548	0.545	0.875
31	riyf_1	riyf_1	country#4	country#4	1.000	0.922	1.000	0.819	0.811	0.955	1.000	0.922	1.000	0.818	0.789	0.953
32	wisaAdap_1	wisaAdap_1	cushion#3	cushion#3	1.000	0.922	1.000	0.819	0.811	0.955	1.000	0.922	1.000	0.818	0.789	0.953
33	basomap_1	DaHik_2	smile#1	laugh#1	0.533	0.358	0.125	0.199	0.201	0.692	0.533	0.358	0.125	0.199	0.201	0.692
34	tawoqiyE_1	tawoqiyE_1	endorsement#5	endorsement#5	1.000	0.922	1.000	0.819	0.835	0.958	1.000	0.922	1.000	0.819	0.826	0.957
35	>adaAp_1	wasiyolap_1	tool#2	means#1	0.941	0.734	0.500	0.670	0.690	0.918	0.941	0.734	0.500	0.670	0.690	0.918
36	qabor_1	qabor_1	grave#2	grave#2	1.000	0.922	1.000	0.819	0.826	0.957	1.000	0.922	1.000	0.819	0.811	0.955
37	Sabiy~_2	Sabiy~_2	spring chicken#1	spring chicken#1	1.000	0.922	1.000	0.819	0.835	0.958	1.000	0.922	1.000	0.819	0.826	0.957
38																
39	HaAfilap_1	HaAfilap_1	coach#5	coach#5	1.000	0.922	1.000	0.819	0.835	0.958	1.000	0.922	1.000	0.819	0.826	0.957
40	kuwb_1	kuwb_1	glass#2	glass#2	1.000	0.922	1.000	0.819	0.826	0.957	1.000	0.922	1.000	0.819	0.811	0.955

	Performance Measures											
Sim. level	Correlation r						Correlation r					
all	0.858	0.774	0.658	0.787	0.806	0.831	0.850	0.814	0.712	0.825	0.840	0.826
low	0.060	-0.115	-0.162	-0.131	-0.074	0.155	-0.075	-0.088	-0.139	-0.135	-0.090	-0.089
mid	0.122	-0.095	-0.127	-0.077	-0.075	-0.052	0.103	0.269	0.324	0.239	0.188	0.050
high	0.152	0.314	0.393	0.265	0.300	0.177	0.171	0.314	0.393	0.267	0.345	0.197
Sim. level	MSE						MSE					
all	0.066	0.045	0.104	0.055	0.047	0.109	0.064	0.038	0.092	0.048	0.044	0.104
low	0.118	0.050	0.010	0.015	0.016	0.247	0.118	0.057	0.010	0.017	0.016	0.240
mid	0.072	0.056	0.178	0.087	0.071	0.101	0.067	0.031	0.142	0.067	0.057	0.091
high	0.020	0.030	0.109	0.056	0.050	0.009	0.020	0.030	0.109	0.056	0.053	0.008

translations defined in the benchmark dataset. In $wnTrans$ the maximum similarity score is selected, such that the ArWN and the EnWN cover the Arabic word and its translation in English, respectively. Otherwise, the default setting uDS is applied.

4. *Upper Bound* (UB): calculates the semantic similarity between given words senses, such that, UB selects the sense pair that maximize correlation r values and minimize MSE values w.r.t the HR ratings (benchmark dataset). UB indicates the optimal scores for the considered experiment settings.

5. *Unimproved HyperTrees* (uHT): calculates the semantic similarity using ArWN while ignoring the structure enhancement. That is, the semantic gaps are considered in calculating the similarly scores.

6. *Improved HyperTrees* (iHT): calculates the semantic similarity using the enhanced structure of ArWN.

4.5 Results & Discussion

Tables 5, 6, 7, and 8 report the semantic similarity scores using six similarity measures, which resulted from applying uDS, DS, $wnTrans$ and UB configurations over ArWN; respectively. Such that two variants, uHT and iHT, are considered. The tables also list the Arabic senses and their correspondences senses in English, which are used to provide the obtained similarity scores. Table 9 reports the semantic similarity scores that are obtained from applying uDs, DS, and UB configurations over EnWN. English-based

Table 6: DS configuration over ArWN

					iHT						uHT					
NO.	Ar Word Pairs Senses		En Word Pairs Senses		WuP	LCH	Path	LI	AWSS	Aldiery	WuP	LCH	Path	LI	AWSS	Aldiery
1	$aATi}_AlbaHor_1	tawoqiyE_1	coast#1	endorsement#5	0.235	0.207	0.071	0.051	0.045	0.379	0.267	0.248	0.083	0.076	0.062	0.411
2	Zuhor_1	gazol_1	noon#1	thread#1	0.200	0.154	0.059	0.028	0.028	0.343	0.211	0.170	0.063	0.034	0.033	0.354
3																
4	Habol_1	Zuhor_1	cord#1	midday#1	0.211	0.170	0.063	0.034	0.033	0.354	0.222	0.188	0.067	0.042	0.038	0.366
5	tawoqiyE_1	gazol_1	endorsement#5	thread#1	0.211	0.170	0.063	0.034	0.033	0.354	0.222	0.188	0.067	0.042	0.038	0.366
6	walad_1	tawoqiyE_1	boy#1	endorsement#5	0.235	0.207	0.071	0.051	0.045	0.379	0.250	0.227	0.077	0.062	0.053	0.394
7	walad_1	Zuhor_1	boy#1	midday#1	0.222	0.188	0.067	0.042	0.038	0.366	0.235	0.207	0.071	0.051	0.045	0.379
8	basomap_1	qaroyap_2	smile#1	village#2	0.235	0.207	0.071	0.051	0.045	0.379	0.267	0.248	0.083	0.076	0.062	0.411
9	Zuhor_1	Sawom_1	noon#1	fasting#1	0.364	0.188	0.067	0.049	0.065	0.574	0.364	0.188	0.067	0.049	0.065	0.574
10	kuwb_1	AlomAs_1	glass#2	diamond#2	0.353	0.248	0.083	0.086	0.087	0.535	0.267	0.248	0.083	0.076	0.062	0.411
11	qabor_1	ra}iyos_1	grave#2	head#4	0.444	0.272	0.091	0.109	0.123	0.633	0.375	0.272	0.091	0.105	0.102	0.553
12	riyf_1	xuDaAr_1	country#4	vegetable#1	0.375	0.272	0.091	0.105	0.102	0.553	0.286	0.272	0.091	0.092	0.073	0.429
13	kuwb_1	>adaAp_1	glass#2	tool#2	0.222	0.922	1.000	0.683	0.371	0.691	0.235	0.207	0.071	0.051	0.045	0.379
14	DaHik_2	<iHotifAl_1	laugh#1	celebration#1	0.400	0.298	0.100	0.128	0.120	0.574	0.400	0.298	0.100	0.128	0.120	0.574
15	fataAp_1	xaAdim_1	girl#1	retainer#2	0.762	0.436	0.167	0.301	0.361	0.842	0.737	0.436	0.167	0.301	0.351	0.827
16	<iHotifAl_1	Sawom_1	celebration#1	fasting#1	0.526	0.298	0.100	0.135	0.163	0.703	0.526	0.298	0.100	0.135	0.163	0.703
17	HaAfilap_1	wasiylap_1	coach#5	means#2	0.778	0.486	0.200	0.368	0.412	0.845	0.750	0.486	0.200	0.367	0.394	0.828
18	fayolasuwf_1	ra}iyos_1	philosopher#1	head#4	0.762	0.436	0.167	0.301	0.361	0.842	0.737	0.436	0.167	0.301	0.351	0.827
19	fataAp_1	>xot_1	girl#1	sister#1	0.696	0.358	0.125	0.202	0.261	0.815	0.667	0.358	0.125	0.202	0.254	0.796
20	dajaAjap_1	HamaAm_1	hen#1	pigeon#1	0.828	0.436	0.167	0.301	0.376	0.879	0.815	0.436	0.167	0.301	0.374	0.872
21	rukaAm_1	jabal_1	hill#2	mountain#1	0.533	0.358	0.125	0.199	0.201	0.692	0.500	0.395	0.143	0.233	0.195	0.649
22	say~id_1	ra}iyos_1	sir#1	head#4	0.762	0.436	0.167	0.301	0.361	0.842	0.737	0.436	0.167	0.301	0.351	0.827
23	TaEaAm_1	xuDaAr_1	food#1	vegetable#1	0.571	0.395	0.143	0.243	0.236	0.716	0.500	0.395	0.143	0.233	0.195	0.649
24	Eabod_1	xaAdim_1	slave#1	retainer#2	0.842	0.546	0.250	0.449	0.499	0.874	0.824	0.546	0.250	0.449	0.485	0.864
25	jaroy_1	ma$oy_1	run#7	walk#1	0.909	0.624	0.333	0.549	0.604	0.905	0.909	0.624	0.333	0.549	0.604	0.905
26	Habol_1	gazol_1	cord#1	thread#1	0.941	0.734	0.500	0.670	0.690	0.918	0.933	0.734	0.500	0.670	0.671	0.913
27	dagol_1	dagol_1	jungle#1	jungle#1	1.000	0.922	1.000	0.818	0.754	0.950	1.000	0.922	1.000	0.815	0.701	0.947
28	fayolasuwf_1	mufak~ir_1	philosopher#1	intellect#3	0.900	0.624	0.333	0.549	0.597	0.900	0.889	0.624	0.333	0.549	0.587	0.894
29	riHolap_1	safar_1	journey#1	travel#1	0.952	0.734	0.500	0.670	0.710	0.926	0.952	0.734	0.500	0.670	0.710	0.926
30	HajarN_kariym_1	AlomAs_1	gem#2	diamond#2	0.875	0.624	0.333	0.549	0.570	0.886	0.857	0.624	0.333	0.548	0.545	0.875
31	riyf_1	qaroyap_2	country#4	village#2	0.824	0.546	0.250	0.449	0.485	0.864	0.857	0.624	0.333	0.548	0.545	0.875
32	wisaAdap_1	wisaAdap_1	cushion#3	cushion#3	1.000	0.922	1.000	0.819	0.811	0.955	1.000	0.922	1.000	0.818	0.789	0.953
33	basomap_1	DaHik_2	smile#1	laugh#1	0.533	0.358	0.125	0.199	0.201	0.692	0.533	0.358	0.125	0.199	0.201	0.692
34	tawoqiyE_1	tawoqiyE_1	Endorsement#5	Endorsement#5	1.000	0.922	1.000	0.819	0.835	0.958	1.000	0.922	1.000	0.819	0.826	0.957
35	>adaAp_1	wasiyolap_1	tool#2	means#1	0.941	0.734	0.500	0.670	0.690	0.918	0.941	0.734	0.500	0.670	0.690	0.918
36	qabor_1	qabor_1	grave#2	grave#2	1.000	0.922	1.000	0.819	0.826	0.957	1.000	0.922	1.000	0.819	0.811	0.955
37	muraAhiq_1	Tifol_1	adolescent#1	child#1	0.900	0.624	0.333	0.549	0.597	0.900	0.889	0.624	0.333	0.549	0.587	0.894
38																
39	HaAfilap_1	HaAfilap_1	coach#5	bus#1	1.000	0.922	1.000	0.819	0.835	0.958	1.000	0.922	1.000	0.819	0.826	0.957
40	kuwb_1	kuwb_1	glass#2	glass#2	1.000	0.922	1.000	0.819	0.826	0.957	1.000	0.922	1.000	0.819	0.811	0.955

	Performance Measures											
Sim. level	Correlation r						Correlation r					
all	0.925	0.799	0.607	0.824	0.877	0.933	0.920	0.859	0.688	0.864	0.878	0.925
low	0.810	0.853	0.875	0.901	0.882	0.776	0.564	0.687	0.712	0.756	0.738	0.478
mid	0.736	-0.156	-0.389	0.018	0.463	0.636	0.698	0.770	0.702	0.739	0.695	0.686
high	0.130	0.234	0.290	0.199	0.242	0.150	0.136	0.229	0.288	0.193	0.279	0.161
Sim. level	MSE						MSE					
all	0.028	0.042	0.131	0.065	0.052	0.061	0.026	0.034	0.118	0.062	0.057	0.053
low	0.044	0.020	0.007	0.005	0.005	0.135	0.042	0.026	0.007	0.006	0.007	0.125
mid	0.024	0.060	0.207	0.104	0.076	0.055	0.022	0.033	0.176	0.099	0.088	0.042
high	0.018	0.043	0.158	0.076	0.067	0.008	0.018	0.040	0.152	0.071	0.067	0.008

similarity (Path, lch, WuP, and Li measures) scores are reported for each configuration. English senses that are used to compute the scores are also reported.

Observe that the word senses are defined differently based on the applied configuration. For example, the word boy "صبي" is selected differently w.r.t the applied configuration; in Table 5, in the uDS setting the selected sense is (Sabiy _1, juvenile#1) [14], in DS (Table 6) and $wnTrans$ (Table 7) settings the selected sense is (walad_1, boy#1), and in UB (Table 8) setting the sense is (walad_2, boy#2). Moreover, considering $wnTrans$ setting, the translation which are defined in the AWSS benchmark for 13 Arabic words that exist in 17 word pairs, does not exist in the mapping between ArWN and EnWN; the words and their translation are {signature:توقيع ; sepulcher: ضريح; sheikh:شيخ; countryside:ريف; tumbler:قدح; feast:عيد; odalisque:جارية; sage: حكيم; thinker:مفكر; pillow:مخدة; signature:توقيع; lad:فتى}. For example, the word "توقيع" has one sense in ArWN "*tawoqiyE_1*", which is mapped into the "*endorsement#5*" in EnWN, while none of the five senses for the word "*signature*" in EnWN is mapped into ArWN. Noting that 28 word pairs out of the 40 word pairs has at least one missing correspondence sense in EnWN when considering uHT setting, For example; similarity scores of the 21^{st} word pairs (Hill "تل"; mountain "جبل"); which is also illustrated in

[14] Represents Arabic sense in ArWN and its correspondence in EnWN

Table 7: wnTrans configuration over ArWN

					iHT						uHT					
NO.	Ar Word Pairs Senses		En Word Pairs Senses		WuP	LCH	Path	LI	AWSS	Aldiery	WuP	LCH	Path	LI	AWSS	Aldiery
1	$aATi}_AlbaHor_1	tawoqiyE_1	coast#1	endorsement#5	0.235	0.207	0.071	0.051	0.045	0.379	0.267	0.248	0.083	0.076	0.062	0.411
2	Zuhor_1	xayoT_2	noon#1	string#9	0.353	0.248	0.083	0.086	0.087	0.535	0.353	0.248	0.083	0.086	0.087	0.535
3																
4	Habol_1	Zuhor_1	cord#1	midday#1	0.211	0.170	0.063	0.034	0.033	0.354	0.222	0.188	0.067	0.042	0.038	0.366
5	tawoqiyE_1	xayoT_2	endorsement#5	string#9	0.375	0.272	0.091	0.105	0.102	0.553	0.375	0.272	0.091	0.105	0.102	0.553
6	walad_1	tawoqiyE_1	boy#1	endorsement#5	0.235	0.207	0.071	0.051	0.045	0.379	0.250	0.227	0.077	0.062	0.053	0.394
7	walad_1	Zuhor_1	boy#1	midday#1	0.222	0.188	0.067	0.042	0.038	0.366	0.235	0.207	0.071	0.051	0.045	0.379
8	basomap_1	qaroyap_1	smile#1	village#2	0.235	0.207	0.071	0.051	0.045	0.379	0.267	0.248	0.083	0.076	0.062	0.411
9	Zuhor_1	Sawom_1	noon#1	fasting#1	0.364	0.188	0.067	0.049	0.065	0.574	0.364	0.188	0.067	0.049	0.065	0.574
10	kuwb_1	AlomAs_1	glass#2	diamond#2	0.353	0.248	0.083	0.086	0.087	0.535	0.267	0.248	0.083	0.076	0.062	0.411
11	maqaAm_1	ra}iyos_1	shrine#1	head#4	0.500	0.272	0.091	0.110	0.139	0.687	0.556	0.326	0.111	0.164	0.192	0.720
12	riyf_1	xuDaAr_1	country#4	vegetable#1	0.375	0.272	0.091	0.105	0.102	0.553	0.286	0.272	0.091	0.092	0.073	0.429
13	kuwb_1	>aadaAp_1	tool#1	glass#2	0.778	0.922	1.000	0.818	0.789	0.927	0.750	0.486	0.200	0.367	0.394	0.828
14	DaHik_2	<iHotifaAl_1	laugh#1	celebration#2	0.375	0.272	0.091	0.105	0.102	0.553	0.375	0.272	0.091	0.105	0.102	0.553
15	fataAp_1	xaAdim_1	girl#1	retainer#2	0.762	0.436	0.167	0.301	0.361	0.842	0.737	0.436	0.167	0.301	0.351	0.827
16	Eiyod_1	Sawom_1	day#3	fasting#1	0.316	0.207	0.071	0.058	0.063	0.504	0.316	0.207	0.071	0.058	0.063	0.504
17	HaAfilap_1	wasiylap_1	coach#5	means#2	0.778	0.486	0.200	0.368	0.412	0.845	0.750	0.486	0.200	0.367	0.394	0.828
18	fayolasuwf_1	ra}iyos_1	philosopher#1	head#4	0.762	0.436	0.167	0.301	0.361	0.842	0.737	0.436	0.167	0.301	0.351	0.827
19	fataAp_1	>xot_1	girl#1	sister#1	0.696	0.358	0.125	0.202	0.261	0.815	0.667	0.358	0.125	0.202	0.254	0.796
20	dajaAjap_1	HamaAm_1	hen#1	pigeon#1	0.828	0.436	0.167	0.301	0.376	0.879	0.815	0.436	0.167	0.301	0.374	0.872
21	rukaAm_1	jabal_1	hill#2	mountain#1	0.533	0.358	0.125	0.199	0.201	0.692	0.500	0.395	0.143	0.233	0.195	0.649
22	say~id_1	ra}iyos_1	sir#1	head#4	0.762	0.436	0.167	0.301	0.361	0.842	0.737	0.436	0.167	0.301	0.351	0.827
23	TaEaAm_1	xuDaAr_1	food#1	vegetable#1	0.571	0.395	0.143	0.243	0.236	0.716	0.500	0.395	0.143	0.233	0.195	0.649
24	xaAdim_1	xaAdim_1	retainer#2	retainer#2	1.000	0.922	1.000	0.819	0.835	0.958	1.000	0.922	1.000	0.819	0.826	0.957
25	jaroy_1	ma$oy_1	run#7	walk#1	0.909	0.624	0.333	0.549	0.604	0.905	0.909	0.624	0.333	0.549	0.604	0.905
26	Habol_1	xayoT_2	cord#1	string#9	0.286	0.272	0.091	0.092	0.073	0.429	0.308	0.298	0.100	0.113	0.086	0.451
27	gaAbap_1	dagol_1	forest#1	jungle#1	0.308	0.298	0.100	0.113	0.086	0.451	0.333	0.326	0.111	0.138	0.102	0.475
28	fayolasuwf_1	mufak~ir_1	philosopher#1	intellect#3	0.900	0.624	0.333	0.549	0.597	0.900	0.889	0.624	0.333	0.549	0.587	0.894
29	riHolap_1	safar_1	journey#1	travel#1	0.952	0.734	0.500	0.670	0.710	0.926	0.952	0.734	0.500	0.670	0.710	0.926
30	HajarN_kariym_1	AlomAs_1	gem#2	diamond#2	0.875	0.624	0.333	0.549	0.570	0.886	0.857	0.624	0.333	0.548	0.545	0.875
31	riyf_1	qaroyap_2	country#4	village#2	0.824	0.546	0.250	0.449	0.485	0.864	0.857	0.624	0.333	0.548	0.545	0.875
32	wisaAdap_1	wisaAdap_1	cushion#3	cushion#3	1.000	0.922	1.000	0.819	0.811	0.955	1.000	0.922	1.000	0.818	0.789	0.953
33	basomap_1	DaHik_2	smile#1	laugh#1	0.533	0.358	0.125	0.199	0.201	0.692	0.533	0.358	0.125	0.199	0.201	0.692
34	tawoqiyE_1	tawoqiyE_1	Endorsement#5	Endorsement#5	1.000	0.922	1.000	0.819	0.835	0.958	1.000	0.922	1.000	0.819	0.826	0.957
35	>adaAp_1	wasiyolap_1	tool#2	means#1	0.941	0.734	0.500	0.670	0.690	0.918	0.941	0.734	0.500	0.670	0.690	0.918
36	qabor_1	qabor_1	grave#2	grave#2	1.000	0.922	1.000	0.819	0.826	0.957	1.000	0.922	1.000	0.819	0.811	0.955
37	walad_1	Sabiy~_2	boy#1	spring chicken#1	0.800	0.486	0.200	0.368	0.424	0.858	0.778	0.486	0.200	0.368	0.412	0.845
38																
39	HaAfilap_1	HaAfilap_1	coach#5	bus#1	1.000	0.922	1.000	0.819	0.835	0.958	1.000	0.922	1.000	0.819	0.826	0.957
40	kuwb_1	kuwb_1	glass#2	glass#2	1.000	0.922	1.000	0.819	0.826	0.957	1.000	0.922	1.000	0.819	0.811	0.955

	Performance Measures											
Sim. level	Correlation r						Correlation r					
all	0.787	0.703	0.540	0.703	0.719	0.757	0.787	0.763	0.616	0.763	0.762	0.774
low	0.615	0.569	0.583	0.610	0.645	0.586	0.391	0.555	0.576	0.538	0.470	0.318
mid	0.453	0.067	-0.082	0.088	0.152	0.345	0.434	0.473	0.397	0.466	0.452	0.407
high	0.672	0.656	0.627	0.647	0.658	0.693	0.665	0.643	0.624	0.635	0.653	0.689
Sim. level	MSE						MSE					
all	0.051	0.061	0.155	0.097	0.089	0.082	0.050	0.051	0.140	0.087	0.084	0.076
low	0.062	0.027	0.008	0.006	0.006	0.170	0.065	0.033	0.007	0.006	0.007	0.167
mid	0.046	0.065	0.199	0.114	0.096	0.071	0.043	0.039	0.165	0.094	0.085	0.059
high	0.047	0.084	0.230	0.153	0.147	0.021	0.044	0.078	0.221	0.144	0.144	0.019

Figure 1, before the enhancement of the ArWN structure, the $HyperTree$ has two semantic gaps $\{geological\,formation\#1\}$ and $\{physical\,entity\#1\}$; while The $HyperTree$(جبل) has one semantic gap, which is $\{physical\,entity\#1\}$.

The performance measures r and MSE are reported for every configuration in the bottom ofTables 6, 5, 7, and 8; including the performance for each similarity level. Observe that, r values show that iHT achieves better performance compared to uHT. While; MSE values indicate that uHT has less difference in similarity scores than iHT, compared to HR rates. In fact, the values of MSE are strongly influenced by uHT, the semantic gaps. Noting that when $HyperTrees$ of two senses have the same semantic gaps; the lcs is reduced which decreases the similarity scores. This gives less difference in similarity scores compared to HR rates. In particular, this happens for MSE values at mid similarity level. For examples, in row 10, the $HyperTrees$ of the word pairs (Glass; Diamond) has the $\{physical\,entity\#1\}$ as a semantic gap. That is, $d(glass\#2) = 9$; $d(diamond\#2) = 8$ and $d(lcs(glass\#2, diamond\#2)) = 3$; while $d(kuwb\_1) = 8$; $d(AlomAs\_1) = 7$; $d(lcs(kuwb\_1, AlomAs\_1)) = 2$.

Furthermore, $wnTrans$ configuration scored the worst performance; this is due to the low Arabic word coverage. A significant finding is that, the richness of ArWN content has a high effect on the evaluation the semantic similarity between the concepts, in terms of the coverage of lexical and semantic relations.

Performance measures in [15, 16]; presented in Table 4; showed that WuP measure scored the best MSE value 0.0165 with 0.94 for r; and comparatively $Aldiery$ measure has obtained the values 0.96

Table 8: UB configuration over ArWN

					iHT						uHT					
NO.	Ar Word Pairs Senses		En Word Pairs Senses		WuP	LCH	Path	LI	AWSS	Aldiery	WuP	LCH	Path	LI	AWSS	Aldiery
1	$aATi}_AlbaHor_1	tawoqiyE_1	coast#1	endorsement#5	0.235	0.207	0.071	0.051	0.045	0.379	0.267	0.248	0.083	0.076	0.062	0.411
2	Zuhor_1	gazol_1	noon#1	thread#1	0.200	0.154	0.059	0.028	0.028	0.343	0.211	0.170	0.063	0.034	0.033	0.354
3																
4	Habol_1	Zuhor_1	cord#1	midday#1	0.211	0.170	0.063	0.034	0.033	0.354	0.222	0.188	0.067	0.042	0.038	0.366
5	tawoqiyE_1	gazol_1	endorsement#5	thread#1	0.211	0.170	0.063	0.034	0.033	0.354	0.222	0.188	0.067	0.042	0.038	0.366
6	walad_2	tawoqiyE_1	boy#2	endorsement#5	0.222	0.188	0.067	0.042	0.038	0.366	0.235	0.207	0.071	0.051	0.045	0.379
7	walad_2	Zuhor_1	boy#2	midday#1	0.211	0.170	0.063	0.034	0.033	0.354	0.222	0.188	0.067	0.042	0.038	0.366
8	basomap_1	qaroyap_2	smile#1	village#2	0.235	0.207	0.071	0.051	0.045	0.379	0.267	0.248	0.083	0.076	0.062	0.411
9	mu&ax~irap_1	Sawom_1	back#2	fasting#1	0.200	0.154	0.059	0.028	0.028	0.343	0.211	0.170	0.063	0.034	0.033	0.354
10	kuwb_1	AlomAs_1	glass#2	diamond#2	0.353	0.248	0.083	0.086	0.087	0.535	0.267	0.248	0.083	0.076	0.062	0.411
11	qabor_1	$ayox_2	grave#2	senator#1	0.400	0.227	0.077	0.073	0.089	0.601	0.333	0.227	0.077	0.070	0.074	0.519
12	riyf_1	xuDar_1	country#4	green#7	0.353	0.248	0.083	0.086	0.087	0.535	0.267	0.248	0.083	0.076	0.062	0.411
13	kuwb_1	>adaAp_1	glass#2	tool#2	0.222	0.922	1.000	0.683	0.371	0.691	0.235	0.207	0.071	0.051	0.045	0.379
14	DaHik_1	Eiyod_1	laughter#2	day#3	0.375	0.272	0.091	0.105	0.102	0.553	0.375	0.272	0.091	0.105	0.102	0.553
15	fataAp_1	xaAdim_1	girl#1	retainer#2	0.762	0.436	0.167	0.301	0.361	0.842	0.737	0.436	0.167	0.301	0.351	0.827
16	<iHotifAl_1	Sawom_1	celebration#1	fasting#1	0.526	0.298	0.100	0.135	0.163	0.703	0.526	0.298	0.100	0.135	0.163	0.703
17	HaAfilap_1	wasiylap_1	coach#5	means#2	0.778	0.486	0.200	0.368	0.412	0.845	0.750	0.486	0.200	0.367	0.394	0.828
18	fayolasuwf_1	$ayox_2	philosopher#1	senator#1	0.696	0.358	0.125	0.202	0.261	0.815	0.667	0.358	0.125	0.202	0.254	0.796
19	fataAp_1	>xot_1	girl#1	sister#1	0.696	0.358	0.125	0.202	0.261	0.815	0.667	0.358	0.125	0.202	0.254	0.796
20	dajaAjap_1	HamaAm_1	hen#1	pigeon#1	0.828	0.436	0.167	0.301	0.376	0.879	0.815	0.436	0.167	0.301	0.374	0.872
21	rukaAm_1	jabal_1	hill#2	mountain#1	0.533	0.358	0.125	0.199	0.201	0.692	0.500	0.395	0.143	0.233	0.195	0.649
22	say~id_1	$ayox_1	lord#3	graybeard#1	0.696	0.358	0.125	0.202	0.261	0.815	0.667	0.358	0.125	0.202	0.254	0.796
23	TaEaAm_3	xuDar_1	food#2	green#7	0.800	0.546	0.250	0.449	0.464	0.850	0.769	0.546	0.250	0.447	0.431	0.831
24	Eabod_1	xaAdim_1	slave#1	retainer#2	0.842	0.546	0.250	0.449	0.499	0.874	0.824	0.546	0.250	0.449	0.485	0.864
25	jaroy_1	ma$oy_1	run#7	walk#1	0.909	0.624	0.333	0.549	0.604	0.905	0.909	0.624	0.333	0.549	0.604	0.905
26	Habol_1	xayoT_1	cord#1	cord#4	0.750	0.486	0.200	0.367	0.394	0.828	0.714	0.486	0.200	0.366	0.367	0.805
27	dagol_1	dagol_1	jungle#1	jungle#1	1.000	0.922	1.000	0.818	0.754	0.950	1.000	0.922	1.000	0.815	0.701	0.947
28	fayolasuwf_1	mufak~ir_1	philosopher#1	intellect#3	0.900	0.624	0.333	0.549	0.597	0.900	0.889	0.624	0.333	0.549	0.587	0.894
29	riHolap_1	safar_1	journey#1	travel#3	0.857	0.546	0.250	0.449	0.508	0.883	0.857	0.546	0.250	0.449	0.508	0.883
30	HajarN_kariym_1	AlomAs_1	gem#2	diamond#2	0.875	0.624	0.333	0.549	0.570	0.886	0.857	0.624	0.333	0.548	0.545	0.875
31	riyf_1	qaroyap_2	country#4	village#2	0.824	0.546	0.250	0.449	0.485	0.864	0.857	0.624	0.333	0.548	0.545	0.875
32	wisaAdap_1	wisaAdap_1	cushion#3	cushion#3	1.000	0.922	1.000	0.819	0.811	0.955	1.000	0.922	1.000	0.818	0.789	0.953
33	basomap_1	DaHik_2	smile#1	laugh#1	0.533	0.358	0.125	0.199	0.201	0.692	0.533	0.358	0.125	0.199	0.201	0.692
34	tawoqiyE_1	tawoqiyE_1	endorsement#5	endorsement#5	1.000	0.922	1.000	0.819	0.835	0.958	1.000	0.922	1.000	0.819	0.826	0.957
35	>adaAp_1	wasiyolap_1	tool#2	means#1	0.941	0.734	0.500	0.670	0.690	0.918	0.941	0.734	0.500	0.670	0.690	0.918
36	qabor_1	qabor_1	grave#2	grave#2	1.000	0.922	1.000	0.819	0.826	0.957	1.000	0.922	1.000	0.819	0.811	0.955
37	muraAhiq_1	Tifol_1	adolescent#1	child#1	0.900	0.624	0.333	0.549	0.597	0.900	0.889	0.624	0.333	0.549	0.587	0.894
38																
39	HaAfilap_1	HaAfilap_1	coach#5	bus#1	1.000	0.922	1.000	0.819	0.835	0.958	1.000	0.922	1.000	0.819	0.826	0.957
40	kuwb_1	kuwb_1	glass#2	glass#2	1.000	0.922	1.000	0.819	0.826	0.957	1.000	0.922	1.000	0.819	0.811	0.955

	Performance Measures											
Sim. level	Correlation r						Correlation r					
all	0.935	0.787	0.580	0.813	0.874	0.945	0.934	0.849	0.662	0.856	0.879	0.943
low	0.825	0.696	0.710	0.765	0.815	0.828	0.621	0.458	0.456	0.509	0.591	0.613
mid	0.815	-0.087	-0.358	0.093	0.537	0.739	0.806	0.792	0.734	0.766	0.773	0.788
high	0.345	0.415	0.402	0.419	0.465	0.333	0.372	0.413	0.401	0.417	0.509	0.369
Sim. level	MSE						MSE					
all	0.023	0.046	0.144	0.073	0.058	0.054	0.022	0.037	0.131	0.070	0.062	0.047
low	0.035	0.019	0.008	0.007	0.006	0.114	0.034	0.025	0.008	0.008	0.008	0.105
mid	0.022	0.060	0.205	0.105	0.074	0.055	0.018	0.033	0.175	0.099	0.084	0.041
high	0.015	0.054	0.193	0.096	0.083	0.006	0.015	0.051	0.186	0.091	0.085	0.006

and 0.0189 for r and MSE; respectively. Nevertheless, [15, 16] did not explicitly state which configuration was considered in calculating the similarity scores. For instance. in Table 8; where the UB scores indicate the best value for r is 0.945 (with 0.542 for MSE); which is obtained by *Aldiery* measure; and the best MSE value is 0.0203 (with 0.935 for r); which is obtained by *WuP* measure. Further, in [15, 16] semantic similarity scores were reported to be equal to zero for the word pairs in rows $1-9$, which are at the low similarity level, and the word pair in row 21 was considered as not covered ArWN, hence, this increased the r values and reduced MSE values. However, no explanation is provided.

Overall, the reported performance values show that the enhancement of the semantic structure has a strong effect on estimating the semantic similarity between the concepts. Observe that, word pairs at low and mid similarity levels gives better r values than high similarity level. While words pairs in high similarity level gives better MSE values. in other words, similarity measures obtained best coloration values when the concepts are not similar. Both ArWN and EnWN, r and MSE measures indicate that best performance is achieved when word senses are determined in advance, i.e., DS configuration. However, it is important to distinguish the approach which is used to define the sense, in this work consensus based approach is used.

In other hand; the user feedback based approach, ICLM application that adopted to fill the semantic gaps, shows its effectiveness in selecting the senses, such that scores obtained in DS are close to optimal scores achieved with upper bound setting UB. Further, Arabic-based measure *Aldiery* performs better than $AWSS$, also

Table 9: uDS, DS, and UB configuration over EnWN

	uDS						DS						UB					
NO.	En Word Pairs senses		WuP	LCH	Path	LI	En Word Pairs senses		WuP	LCH	Path	LI	En Word Pairs senses		WuP	LCH	Path	LI
1	coast#4	endorsement#2	0.632	0.436	0.125	0.202	coast#1	endorsement#1	0.286	0.350	0.091	0.092	coast#1	endorsement#5	0.235	0.285	0.071	0.051
2	noon#1	string#9	0.353	0.326	0.083	0.086	noon#1	string#9	0.353	0.326	0.083	0.086	noon#1	string#2	0.182	0.202	0.053	0.019
3	stove#2	walk#5	0.632	0.436	0.125	0.202	stove#1	walk#1	0.167	0.175	0.048	0.013	stove#1	walk#6	0.160	0.162	0.045	0.010
4	cord#2	midday#1	0.316	0.285	0.071	0.058	cord#1	midday#1	0.211	0.248	0.063	0.034	cord#3	midday#1	0.190	0.216	0.056	0.023
5	signature#5	string#7	0.737	0.514	0.167	0.301	signature#1	string#1	0.235	0.285	0.071	0.051	signature#4	string#2	0.200	0.232	0.059	0.028
6	boy#1	endorsement#1	0.286	0.350	0.091	0.092	boy#1	endorsement#5	0.235	0.285	0.071	0.051	boy#2	endorsement#5	0.222	0.266	0.067	0.042
7	boy#1	midday#1	0.222	0.266	0.067	0.042	boy#1	midday#1	0.222	0.266	0.067	0.042	boy#2	midday#1	0.211	0.248	0.063	0.034
8	smile#1	village#1	0.375	0.350	0.091	0.105	smile#1	village#1	0.375	0.350	0.091	0.105	smile#1	village#2	0.235	0.285	0.071	0.051
9	noon#1	fasting#1	0.364	0.266	0.067	0.049	noon#1	fasting#1	0.364	0.266	0.067	0.049	noon#1	fasting#1	0.364	0.266	0.067	0.049
10	glass#1	diamond#2	0.667	0.514	0.167	0.300	glass#1	diamond#1	0.353	0.326	0.083	0.086	glass#4	diamond#3	0.148	0.138	0.042	0.007
11	sepulcher#1	sheikh#1	0.476	0.326	0.083	0.090	sepulcher#1	sheikh#1	0.476	0.326	0.083	0.090	sepulcher#1	sheikh#1	0.476	0.326	0.083	0.090
12	countryside#1	vegetable#2	0.400	0.305	0.077	0.073	countryside#1	vegetable#2	0.400	0.305	0.077	0.073	countryside#1	vegetable#1	0.353	0.326	0.083	0.086
13	tumbler#2	tool#1	0.737	0.514	0.167	0.301	tumbler#2	tool#1	0.737	1.000	1.000	0.818	tumbler#1	tool#4	0.316	1.000	1.000	0.775
14	laugh#1	feast#2	0.400	0.376	0.100	0.128	laugh#1	feast#2	0.400	0.376	0.100	0.128	laugh#2	feast#1	0.333	0.305	0.077	0.070
15	girl#1	odalisque#1	0.833	0.564	0.200	0.368	girl#1	odalisque#1	0.833	0.564	0.200	0.368	girl#3	odalisque#1	0.750	0.472	0.143	0.247
16	feast#2	fasting#1	0.526	0.376	0.100	0.135	feast#2	fasting#1	0.526	0.376	0.100	0.135	feast#4	fasting#1	0.500	0.350	0.091	0.110
17	coach#5	means#2	0.778	0.564	0.200	0.368	coach#5	means#2	0.778	0.564	0.200	0.368	coach#1	means#2	0.526	0.376	0.100	0.135
18	Sage#1	Sheikh#1	0.762	0.514	0.167	0.301	Sage#1	Sheikh#1	0.762	0.514	0.167	0.301	Sage#3	Sheikh#1	0.636	0.404	0.111	0.165
19	girl#1	sister#4	0.957	0.812	0.500	0.670	girl#1	sister#1	0.696	0.436	0.125	0.202	girl#1	sister#1	0.696	0.436	0.125	0.202
20	Hen#2	pigeon#1	0.846	0.564	0.200	0.368	hen#2	pigeon#1	0.846	0.564	0.200	0.368	hen#4	pigeon#1	0.828	0.514	0.167	0.301
21	hill#1	mountain#1	0.857	0.702	0.333	0.548	hill#1	mountain#1	0.857	0.702	0.333	0.548	hill#2	mountain#1	0.667	0.404	0.111	0.165
22	master#2	Sheikh#1	0.900	0.702	0.333	0.549	master#2	Sheikh#1	0.900	0.702	0.333	0.549	master#7	Sheikh#1	0.667	0.404	0.111	0.165
23	food#2	vegetable#1	0.857	0.702	0.333	0.548	food#2	vegetable#1	0.857	0.702	0.333	0.548	food#1	vegetable#1	0.571	0.472	0.143	0.243
24	slave#1	odalisque#1	0.727	0.472	0.143	0.247	slave#1	odalisque#1	0.727	0.472	0.143	0.247	slave#2	odalisque#1	0.696	0.436	0.125	0.202
25	run#7	walk#1	0.909	0.702	0.333	0.549	run#7	walk#1	0.909	0.702	0.333	0.549	run#6	walk#1	0.750	0.472	0.143	0.247
26	cord#1	string#1	0.941	0.812	0.500	0.670	cord#1	string#1	0.941	0.812	0.500	0.670	cord#3	string#2	0.762	0.514	0.167	0.301
27	forest#2	woodland#1	1.000	1.000	1.000	0.818	forest#2	woodland#1	1.000	1.000	1.000	0.818	forest#2	woodland#1	1.000	1.000	1.000	0.818
28	Sage#1	thinker#1	0.857	0.624	0.250	0.449	Sage#1	thinker#1	0.857	0.624	0.250	0.449	Sage#1	thinker#1	0.857	0.624	0.250	0.449
29	journey#1	travel#1	0.952	0.812	0.500	0.670	journey#1	travel#1	0.952	0.812	0.500	0.670	journey#1	travel#3	0.857	0.624	0.250	0.449
30	Gem#5	diamond#1	0.952	0.812	0.500	0.670	Gem#5	diamond#1	0.952	0.812	0.500	0.670	gem#2	diamond#2	0.875	0.702	0.333	0.549
31	countryside#1	village#2	0.778	0.564	0.200	0.368	countryside#1	village#2	0.778	0.564	0.200	0.368	countryside#1	village#2	0.778	0.564	0.200	0.368
32	cushion#3	pillow#1	0.941	0.812	0.500	0.670	cushion#3	pillow#1	0.941	0.812	0.500	0.670	cushion#3	pillow#1	0.941	0.812	0.500	0.670
33	smile#1	laugh#2	0.875	0.702	0.333	0.549	smile#1	laugh#2	0.875	0.702	0.333	0.549	smile#1	laugh#2	0.875	0.702	0.333	0.549
34	signature#1	endorsement#4	0.941	0.812	0.500	0.670	signature#1	endorsement#4	0.941	0.812	0.500	0.670	signature#1	endorsement#4	0.941	0.812	0.500	0.670
35	tool#2	means#1	0.941	0.812	0.500	0.670	tool#1	means#2	0.824	0.624	0.250	0.449	tool#2	means#1	0.941	0.812	0.500	0.670
36	sepulcher#1	grave#2	0.941	0.812	0.500	0.670	sepulcher#1	grave#2	0.941	0.812	0.500	0.670	sepulcher#1	grave#2	0.941	0.812	0.500	0.670
37	boy#1	lad#2	0.952	0.812	0.500	0.670	boy#1	lad#2	0.952	0.812	0.500	0.670	boy#1	lad#2	0.952	0.812	0.500	0.670
38	wizard#2	magician#2	1.000	1.000	1.000	0.819	wizard#2	magician#2	1.000	1.000	1.000	0.819	wizard#2	magician#2	1.000	1.000	1.000	0.819
39	coach#5	bus#1	1.000	1.000	1.000	0.819	coach#5	bus#1	1.000	1.000	1.000	0.819	coach#5	bus#1	1.000	1.000	1.000	0.819
40	glass#2	tumbler#2	0.947	0.812	0.500	0.670	glass#2	tumbler#2	0.947	0.812	0.500	0.670	glass#2	tumbler#2	0.947	0.812	0.500	0.670

Performance Measures												
sim. level	correlation r				correlation r				correlation r			
all	0.856	0.851	0.726	0.865	0.949	0.832	0.623	0.845	0.965	0.801	0.565	0.796
low	-0.071	-0.247	-0.241	-0.231	0.697	0.246	0.223	0.310	0.731	0.570	0.645	0.776
mid	0.666	0.604	0.542	0.612	0.675	0.010	-0.289	0.093	0.791	-0.319	-0.488	-0.309
high	0.261	0.268	0.234	0.279	0.139	0.174	0.173	0.170	0.611	0.550	0.422	0.601
sim. level	MSE				MSE				MSE			
all	0.066	0.038	0.104	0.046	0.036	0.040	0.126	0.056	0.016	0.043	0.160	0.091
low	0.150	0.089	0.011	0.021	0.059	0.056	0.008	0.008	0.035	0.035	0.007	0.006
mid	0.055	0.013	0.125	0.051	0.046	0.044	0.174	0.081	0.010	0.068	0.250	0.174
high	0.008	0.018	0.161	0.062	0.009	0.023	0.179	0.073	0.005	0.026	0.205	0.088

Aldiery measure provided a competitive performance in comparison to *WuP* measures.

5 Conclusion & Future Work

Six path-based similarity measures including English and Arabic based measures are applied over ArWN and EnWN to examine the effect of the improvement of the lexical and semantic coverage on wordnet-based semantic similarity measures. wo variants uHT and iHT of ArWN structure are considered in the experiment to evaluate the impact of filling the semantic gaps on estimating the semantic similarity. The efficacy of the improved structure is examined by experiments in the context of semantic similarity. The semantic similarity scores for a benchmark dataset, human rating for 40 Arabic nominal word pairs, are calculated over ArWN and EnWN in different configurations (uDs, DS, $wnTrans$, and UB). The obtained performance values indicate the importance of the semantic evidence gained with the enrichment process; and its signification effect on estimating the semantic similarity between concepts. Moreover, when considering Arabic-based measures the experiment results showed that *Aldiery* measure performs better than $AWSS$ measure. Beside that, *Aldiery* measure has provided a competitive performance in comparison to the English-based WuP measures. Finally, the resolved semantic gaps of the new structure are made for public.

As a future direction, we plan to compile *xml* format of the new structure, and to integrate it with available ArWN resources (i,e., ArWN release available at Open Multilingual WordNet [31]). It is also interesting is to study the effect of the semantic gaps over NLP applications; for instances Question Answering similar to the work presented in [44], and word sense disambiguation [33, 35] in the context of Arabic.

References

[1] D. Jurafsky, J. H. Martin, Speech and Language Processing (2Nd Edition), Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2009.

[2] Y. Li, Z. A. Bandar, D. Mclean, "An Approach for Measuring Semantic Similarity between Words Using Multiple Information Sources," IEEE Trans. on Knowl. and Data Eng., **15**(4), 871–882, 2003, doi:10.1109/TKDE.2003.1209005.

[3] G. A. Miller, "WordNet: A Lexical Database for English," Commun. ACM, **38**(11), 39–41, 1995, doi:10.1145/219717.219748.

[4] C. Fellbaum, editor, WordNet An Electronic Lexical Database, The MIT Press, Cambridge, MA ; London, 1998.

[5] F. Christiane, H. Amanda, "When WordNet Met Ontology," in Ontology Makes Sense - Essays in honor of Nicola Guarino, 136–151, 2019, doi: 10.3233/978-1-61499-955-3-136.

[6] N. Bouhriz, F. Benabbou, E. H. B. Lahmar, "Word Sense Disambiguation Approach for Arabic Text," (IJACSA) International Journal of Advanced Computer Science and Applications, **7**(4), 2016, doi:10.14569/IJACSA.2016.070451.

[7] P. Vossen, "Eurowordnet: A Multilingual Database Of Autonomous And Language-Specific Wordnets Connected Via An Inter-Lingualindex," International Journal of Lexicography, **17**(2), 161–173, 2004, doi:10.1093/ijl/17.2.161.

[8] M. Arcan, J. P. McCrae, P. Buitelaar, "Polylingual Wordnet," CoRR, **abs/1903.01411**, 2019.

[9] H. Rodríguez, D. Farwell, J. Farreres, M. Bertran, M. A. Martí, W. Black, S. Elkateb, J. Kirk, P. Vossen, C. Fellbaum, "Arabic WordNet: Current State and Future Extensions," in Proceedings of the Forth International Conference on Global WordNet, 2008.

[10] G. A. Miller, W. G. Charles, "Contextual correlates of semantic similarity," Language & Cognitive Processes, **6**(1), 1–28, 1991.

[11] F. Bond, L. Morgado da Costa, M. W. Goodman, J. P. McCrae, A. Lohk, "Some Issues with Building a Multilingual Wordnet," in Proceedings of the 12th Language Resources and Evaluation Conference, 3189–3197, European Language Resources Association, Marseille, France, 2020.

[12] L. Abouenour, K. Bouzoubaa, P. Rosso, "On the evaluation and improvement of Arabic WordNet coverage and usability," Language Resources and Evaluation, **47**, 2013, doi:10.1007/s10579-013-9254-z.

[13] H. Rodríguez, D. Farwell, J. Ferreres, M. Bertran, M. Alkhalifa, A. Martí, "Arabic WordNet: Semi-automatic Extensions using Bayesian Inference," 2008.

[14] M. A. Batita, M. Zrigui, "The Enrichment of Arabic WordNet Antonym Relations," in A. Gelbukh, editor, Computational Linguistics and Intelligent Text Processing, 342–353, Springer International Publishing, Cham, 2018, doi:10.1007/978-3-319-77113-7_27.

[15] N. Mohammed, D. Mohammed, "Experimental Study of Semantic Similarity Measures on Arabic WordNet," International Journal of Computer Science and Network Security, **17**(2), 2017.

[16] M. G. Aldayri, The semantic similarity measures using Arabic ontology, (Master's theses Theses and Dissertations Master). Middle East University, Jordan, 2017.

[17] M. A. Helou, "Effects of Semantic Gaps on Arabic WordNet-Based Similarity Measures," in 2019 International Conference on Innovative Computing (ICIC), 1–10, 2019, doi:10.1109/ICIC48496.2019.8966672.

[18] M. A. Helou, M. Palmonari, "Multi-user Feedback for Large-scale Cross-lingual Ontology Matching," in Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, Funchal, Madeira, Portugal, November 1-3, 57–66, 2017, doi:10.5220/0006503200570066.

[19] R. Rada, H. Mili, E. Bicknell, M. Blettner, "Development and application of a metric on semantic nets," in IEEE Transactions on Systems, Man and Cybernetics, 17–30, 1989, doi:10.1109/21.24528.

[20] Z. Wu, M. Palmer, "Verbs semantics and lexical selection," in Proceedings of the 32nd annual meeting on Association for Computational Linguistics, 133–138, Association for Computational Linguistics, Morristown, NJ, USA, 1994.

[21] C. Leacock, M. Chodorow, "Combining local context and WordNet similarity for word sense identification," WordNet: An electronic lexical database, **49**(2), 265–283, 1998.

[22] F. A. Almarsoomi, J. D. OShea, Z. Bandar, K. Crockett, "AWSS: An Algorithm for Measuring Arabic Word Semantic Similarity," in 2013 IEEE International Conference on Systems, Man, and Cybernetics, 504–509, 2013, doi:10.1109/SMC.2013.92.

[23] F. A, Almarsoomi, J. D, O'Shea, Z. A, Bandar, K. A, Crockett, "Arabic Word Semantic Similarity," International Journal of Cognitive and Language Sciences, **6**(10), 2497 – 2505, 2012, doi:10.5281/zenodo.1080052.

[24] G. Hirst, "Ontology and the Lexicon," in eds. S. Staab, R. Studer, editors, Handbook on Ontologies and Information Systems, Heidelberg, Springer, 2004, doi:https://doi.org/10.1007/978-3-540-24750-0_11.

[25] G. A. Miller, C. Leacock, R. Tengi, R. T. Bunker, "A semantic concordance," in Proceedings of the workshop on Human Language Technology, HLT '93, 303–308, Association for Computational Linguistics, Stroudsburg, PA, USA, 1993.

[26] H. Graeme, "Overcoming Linguistic Barriers to the Multilingual Semantic Web," in P. Buitelaar, P. Cimiano, editors, Towards the Multilingual Semantic Web, 3–14, Springer Berlin Heidelberg, 2014, doi:10.1007/978-3-662-43585-4.

[27] D. Tufis, D. Cristea, S. Stamou, "BalkaNet: Aims, Methods, Results and Perspectives. A General Overview," In: D. Tufiş (ed): Special Issue on BalkaNet. Romanian JSTI, 2004.

[28] G. de Melo, G. Weikum, "Towards a universal wordnet by learning from combined evidence," in D. W.-L. Cheung, I.-Y. Song, W. W. Chu, X. Hu, J. J. Lin, editors, CIKM, 513–522, ACM, 2009, doi:10.1145/1645953.1646020.

[29] E. Pianta, L. Bentivogli, C. Girardi, "MultiWordNet: developing an aligned multilingual database," in Proceedings of the 1st Inter. Global Wordnet Conference, 2002.

[30] V. Nastase, M. Strube, B. Boerschinger, C. Zirn, A. Elghafari, "WikiNet: A Very Large Scale Multi-Lingual Concept Network," in N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, D. Tapias, editors, LREC, European Language Resources Association, 2010.

[31] F. Bond, R. Foster, "Linking and Extending an Open Multilingual Wordnet," in ACL (1), 1352–1362, The Association for Computer Linguistics, 2013.

[32] M. M. Boudabous, N. Chaâben Kammoun, N. Khedher, L. H. Belguith, F. Sadat, "Arabic WordNet semantic relations enrichment through morpho-lexical patterns," in 2013 1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA), 1–6, 2013.

[33] E. Bilel, "Arabic word sense disambiguation: a review," Artif. Intell. Rev., **52**(4), 2475–2532, 2019, doi:10.1007/s10462-018-9622-6.

[34] A. Budanitsky, G. Hirst, "Evaluating WordNet-based Measures of Lexical Semantic Relatedness," Comput. Linguist., **32**(1), 13–47, 2006, doi: 10.1162/coli.2006.32.1.13.

[35] N. Bouhriz, F. Benabbou, E. H. Benlahmar, Y. Zidoun, "Arabic Word Semantic similarity: a survey," in 1st International Conference On Research In Applied Mathematics And Computer, Casablanca, Morroco, 2019.

[36] H. Rubenstein, J. B. Goodenough, "Contextual correlates of synonymy," Commun. ACM, **8**(10), 627–633, 1965, doi:10.1145/365628.365657.

[37] P. Resnik, "Using information content to evaluate semantic similarity in a taxonomy," in IJCAI'95: Proceedings of the 14th international joint conference on Artificial intelligence, 448–453, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1995.

[38] D. Lin, "An Information-Theoretic Definition of Similarity," in J. W. Shavlik, J. W. Shavlik, editors, ICML, 296–304, Morgan Kaufmann, 1998.

[39] M. Lesk, "Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone," in SIGDOC '86: Proceedings of the 5th annual international conference on Systems documentation, 24–26, ACM, New York, NY, USA, 1986, doi:https://doi.org/10.1145/318723.318728.

[40] Z. Zhou, Y. Wang, J. Gu, "New model of semantic similarity measuring in wordnet," in 3rd International Conference on Intelligent System and Knowledge Engineering, volume 1, 256–261, 2008, doi:10.1109/ISKE.2008.4730937.

[41] M. A. Helou, A. Abid, "Semantic Measures based on Wordnet using Multiple Information Sources," in KDIR 2010 - Proceedings of the International Conference on Knowledge Discovery and Information Retrieval, Valencia, Spain, October 25-28, 2010, 500–503, 2010.

[42] M. A. Helou, M. Palmonari, M. Jarrar, "Effectiveness of Automatic Translations for Cross-Lingual Ontology Mapping," J. Artif. Intell. Res., **55**, 165–208, 2016, doi:10.1613/jair.4789.

[43] M. A. Helou, M. Palmonari, "Cross-lingual lexical matching with word translation and local similarity optimization," in Proceedings of the 11th International Conference on Semantic Systems, SEMANTICS 2015, Vienna, Austria, September 15-17, 97–104, 2015, doi:10.1145/2814864.2814888.

[44] Y. Regragui, L. Abouenour, F. Krieche, K. Bouzoubaa, P. Rosso, "Arabic WordNet: New Content and New Applications," in In Proceeding of the 8th Global Wordnet Conference (GWN 2016), 2016.