

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/338035783>

Extracting Word Synonyms from Text using Neural Approaches

Article in International Arab Journal of Information Technology · December 2019

DOI: 10.34028/iajit/17/1/6

CITATIONS

15

READS

2,688

1 author:



Nora A. Mohammed

University of Al-Qadisiyah

5 PUBLICATIONS 32 CITATIONS

SEE PROFILE

Extracting Word Synonyms from Text using Neural Approaches

Nora Mohammed

College of Engineering, Al-Qadisiyah University, Iraq

Abstract: *Extracting synonyms from textual corpora using computational techniques is an interesting research problem in the Natural Language Processing (NLP) domain. Neural techniques (such as Word2Vec) have been recently utilized to produce distributional word representations (also known as word embeddings) that capture semantic similarity/relatedness between words based on linear context. Nevertheless, using these techniques for synonyms extraction poses many challenges due to the fact that similarity between vector word representations does not indicate only synonymy between words, but also other sense relations as well as word association or relatedness. In this paper, we tackle this problem using a novel 2-step approach. We first build distributional word embeddings using Word2Vec then use the induced word embeddings as an input to train a feed-forward neural network using annotated dataset to distinguish between synonyms and other semantically related words.*

Keywords: *Neural networks, semantic similarity, word representations, natural language processing.*

Received April 17, 2017; accepted October 24, 2017

<https://doi.org/10.34028/iajit/17/1/6>

1. Introduction

Many attempts have been made in the literature to answer the question: what does it mean for two words to be synonyms? Jurafsky and Martin [5] state that the synonymy relation between two words holds if they can be substituted for each other in all situations and have the same propositional meaning. With this definition, there will only be a few examples that can be categorized as perfect synonyms since different words that have identical meanings in all contexts are rare in natural languages. For example, the pair big-large would be considered as synonyms by most native English speakers. However, in the sentence “He’s like a big brother to me”, “big” cannot be substituted by “large”. Leeuwenberg *et al.* [7] consider two words to be synonyms if they denote the same concept and are interchangeable in many contexts, with regard to one of their senses. The latter definition is quite convenient for computational applications.

Synonyms discovery has been a long-standing exercise for linguists and lexicographers. Furthermore, the accessibility of large repository of synonymy pairs can be crucial for many Natural Language Processing (NLP) applications such as information retrieval, question answering, and machine translation evaluation systems. The largest lexical database that has been manually built and maintained for English language is WordNet [10]. WordNet has been widely used as a synonyms repository in many NLP applications. However, manually building and maintaining such lexical resources is an expensive and labour-intensive task. Therefore, lexical databases such as WordNet for other languages are either small, compared to English

WordNet, or do not exist at all. Moreover, and despite the massive volume of WordNet entries, it does not include many contemporary compound adjectives (e.g., hard-wired, industry-leading, etc.). Computational synonymy discovery from textual sources can assist automating the process of building similar lexical resources for other languages.

From a linguistic point of view, synonymy can be categorized as one of the paradigmatic relations which relate two words that tend to occur in similar context. Besides synonymy, other paradigmatic relations have been recognized. These include antonymy (e.g., large-small), hypernymy (e.g., car-vehicle), co-hyponym (e.g., cat-dog), and metonymy (e.g., face-nose). On the other hand, two words exhibit a syntagmatic relation if they tend to co-occur in the same context [2]. For example, cup and coffee are syntagmatically related due to the popularity of the expression “cup of coffee”. In the psychological literature, this semantic relation is referred to as association [4]. Turney [13] referred to the two major relation categories as functional similarity (words that are paradigmatically related) and domain similarity (words that are syntagmatically related).

Recently, neural word embeddings techniques such as word2vec [9] and GLoVe [11] have been leveraged to build distributional semantic models that capture semantic similarity between word lexemes. However, it has been observed that these techniques cannot be directly used to discover synonyms in textual corpora because context-based distributional models as well as similarity measures are biased towards context similarity, which is not identical to semantic similarity most of the times. For example, in a pilot experiment

we have conducted, the most similar word to “large” was “small”, even though the two are actually exact opposites of each other (antonyms pair). These observations motivate the need for novel computational techniques that are tailored to capture synonymy between word pairs and not any other semantic relation.

The rest of this paper is organized as follows: section 2 presents the related work in the literature regarding computational synonyms extraction. Section 3 introduces the data and approach we have used to construct the word embeddings, or distributional word representations. Section 4 presents the neural network we have used to build a classifier for synonym identification.

2. Related Work

Word embeddings (also known as distributional word representations) are vector representations for words that are usually constructed from raw text based on linear context (words that occur in the neighbourhood of a target word). In these representations, each word is converted into a vector of numerical values or real values. However, in other cases some grammatical information are utilized in the process of constructing word vectors. An example of the latter approach is the work of Levy and Goldberg [8] which constructed vectors based on grammatical context (using dependency relations) instead of linear context.

In the last few years, several neural network-based models for constructing word embeddings have been proposed. Among these, the most popular approach is the Continuous Bag-of-Words model (CBow) and the Skip-Gram model (SG) developed by Mikolov *et al.* [9]. The popularity of these two models is due to the fact that they require less time for constructing the vectors compared to other neural-based approaches. It has been shown that embeddings constructed by these two approaches encode many valuable information about the relation between a pair of words (e.g., similarity, relatedness, morphological information, etc.). Word embeddings constructed by the Skip-Gram model, for instance, have shown to perform very well for the analogy task. A typical example of the analogy task is figure out the question: “man” is to “woman” as “king” is to “?”, which is not a trivial task for computers to perform. Using word embeddings, it has been reported that the distance between the vector representations of the words “king” and “queen” is very close to the distance between the words “man” and “woman”. The Skip-Gram model of word2vec is illustrated in Figure 1.

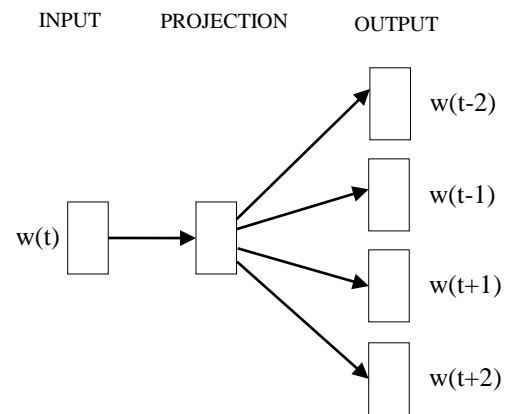


Figure 1. Architecture of the Skip-Gram model for learning word representations.

Computational techniques to extract synonyms from raw text have been inspired by the classical distributional hypothesis “if two words have almost identical environments, we say that they are synonyms” [3]. Since word embeddings that have been constructed using linear contexts encode information about the environments of a target word, they can be used to the synonyms extraction task. A notable work in this direction is that of Van der Plas and Tiedemann [14] which used distributional word similarity to find synonyms in Dutch. Their measure of distributional similarity was constructed using a monolingual Dutch corpus. The authors observed that using this approach other word relation categories (e.g., antonyms, hyponyms, etc.) would occur more than synonyms. To tackle this problem, the authors have used different parallel corpora and word alignment to obtain translation context. The idea is to harness the lexical gap that exists in some languages or the fact that words which have similar meaning translate to only word in another target language. Using both distributional similarity and translational similarity, the authors were able to find more true synonyms than using distributional similarity in isolation. Nevertheless, using translational similarity to extract synonyms for one language requires the availability of several translations of that language to other languages, which is not always available specially for least resourced languages.

Perhaps the most similar work to the research conducted in this paper in the recent literature for synonymy discovery is the work of Leeuwenberg *et al.* [7]. The authors presented a minimally supervised approach to extract synonyms using distributional word vectors for English and German. Their aim was to improve evaluation systems for machine translation by adding more synonymous words to the lexical repository. The main contribution of their work is that they proposed a custom similarity measure tailored to the synonymy discovery task which they referred to as relative cosine similarity. It has been shown in their work that the conventional cosine similarity is a bad measure for the synonymy identification task. To

achieve their goal, the authors conducted an intrinsic and extrinsic evaluation on their approach. For intrinsic evaluation, two human evaluators accessed whether or not each word pair is synonymous. For extrinsic evaluation, the authors used their system in machine translation evaluation task and observe an improvement on the evaluation metric of the machine translation. It worth mentioning that Leeuwenberg *et al.* [7] favoured minimally supervised system so their approach can be extended to other least resourced languages which may not supported by rich lexical databases or sophisticated NLP tools. The only source of annotation that was used in that research is part-of-speech tagged corpus.

In this paper, we use a similar approach to the work proposed by Leeuwenberg *et al.* [7], but we add another supervised step that requires a small set of annotated data. We show that synonymy identification can be tackled as a supervised machine learning task where the features are actually the word embeddings constructed using a generic technique such as Word2Vec. Contrary to previous works in NLP and script recognition which relies on manually engineered features (for example, the work of Bahashwan *et al.* [1] and Khan *et al.* [6]), our work relies on hidden representative features discovered by a neural network.

3. Constructing Word Embeddings

An intuitive simple solution to this problem is to convert each lexical item in the text corpus into a vector representation, then computing the similarity between the vectors using a similarity measure (e.g., cosine similarity). Given vector representations for two words, namely w_1 and w_2 , the cosine similarity can be expressed as inner product space which measures the cosine of the angle between them. The higher the similarity between two vectors, the higher the chance that the lexical items they represent are synonyms. This might be followed by a classification step to identify whether the similarity is actually due to synonymy (positive instance) or other type of similarity/association (negative instance). In this section, the approach that was used to build the word embeddings is explained in details.

3.1. Data and Approach

In this paper, we used NewsCrawl 2014 corpus from WMT workshop. We removed the headlines from the corpus by removing sentences that do not end with a punctuation mark. After that, the corpus was tagged using Stanford Part-Of-Speech (POS) tagger developed by Toutanova *et al.* [12]. The advantage of using POS tagged data is that homographs with different POS tags can be distinguished if the POS tag is appended to the word. For example, the two main senses of the word “book” would be represented as two different words, that is, the noun “book/NN” which refers the instrument that is used to convey information to the

general public in written form, and the verb “book/VB” which is synonymous to the verb “reserve”. To ensure a good quality of the constructed word embeddings, we only considered the words with frequency higher than 25. This was followed by a normalization step in which words were lower-cased (e.g., “The/DT” to “the/DT”) and digits were replaced by a wildcard (e.g., “546/CD” to “DIGIT/CD”). The final vocabulary size was 197,361 word types. Our pre-processed corpus contained 24.7 million sentences and around 602 million word tokens.

In order to construct vectors from the POS-tagged corpus, we used the neural word embedding model (word2vec). Using word2vec, vectors of similar lexical items would be grouped together in the semantic vector-space. The Skip-Gram model was used, number of dimensions was set to 300, and context window size was set to 5. Then, we used the cosine similarity measure to obtain a cluster of similar words for nouns, verbs, adjectives, and adverbs.

3.2. Qualitative Analysis

To perform a simple (qualitative) evaluation of the aforementioned approach, we manually inspect the 5 most similar words to a given set of target words (Table 1).

Many interesting observations can be made from Table 1. For the word “police”, none of the top 5 words can be recognized as a direct synonym. However, these words exhibit domain similarity to the target word. Since the corpus used in this experiment is a news corpus crawled from the open web, one can conclude that these words occur in crime-related reports. In these contexts, the phrase “police investigators” can be regarded as a synonym to the phrase “police detectives”. More interestingly, other paradigmatic relations can be recognized within the top 5 words themselves. For example, “police investigators” and “police officers” are co-hyponyms of the hypernym “employees at the police department”.

Table 1. A few target words (with their POS tag) and their 5 most similar words.

Target word (POS)	Top 5 similar words
police (NN)	authorities, officers, investigators, detectives, eyewitnesses
large (JJ)	small, huge, sizable, massive, big
social-networking (JJ)	photo-sharing, video-sharing, micro-blogging, on-demand, twitter-like
funded (JJ)	state-funded, government-funded, non-federal, non-university, bursary
scientific (JJ)	peer-reviewed, empirical, neuroscientific, laboratory-based, scholarly
inexpensive (JJ)	cheap, off-the-shelf, affordable, cost-effective, nontoxic
murdered (VBD)	raped, abducted, hanged, stabbed, killed
treated (VBN)	cared, discharged, drugged, hospitalized, readmitted
quickly (RB)	soon, swiftly, fast, easily, slowly

The second target word is the adjective “large”. The top 5 similar words are all adjectives that can be used to describe size. Among these, three words are synonymous to the word large (huge, massive, and big). However, the most similar word is “small”, which actually an antonym to the target word. The top 5 words seem to exhibit functional similarity to the target word.

We have observed the presence of many contemporary compound adjectives in the corpus that are made of multi-words separated by a hyphen. An example of these adjectives is social-networking as in “social-networking/JJ website/NN”. The top 5 similar words to this token were also compound adjectives, and interestingly, all of them are within the web 2.0 domain. However, these adjectives exhibit functional similarity as well. For example, if we replace “social-networking” by “micro-blogging” in the sentence “Twitter is a popular social-networking platform”, the meaning of the sentence would not change. In addition, “photo-sharing”, “video-sharing”, and “micro-blogging” are different ways of creating user-generated content on the web, which manifests the co-hyponymy relation. This observation motivates an appealing research direction that aims to discover multi-word adjectives and their domain of usage (e.g., technology, bio-sciences, politics. etc.). A similar trend can be observed with the adjective “funded”. The two most similar lexical items are actually a specification of the adjective, namely “state-funded” and “government-funded”.

Because the technique used in this paper is based on linear context (that is, context words that are present in the neighbourhood of a target word), adjectives that tend to modify the same set of words would have high cosine similarity. For example, the most similar word to the target word “scientific” is “peer-reviewed”, which can be justified by the fact that these two adjectives modify similar word (e.g., “scientific journal” vs. “peer-reviewed” journal or “scientific evidence” vs. “empirical evidence”). Perhaps the only direct example of the feasibility of distributional similarity to discover synonyms is present in the target word “inexpensive”, where the most similar words can be considered as synonyms (or nearly synonyms) by most native English users (i.e., cheap, off-the-shelf, affordable, cost-effective).

For verbs, we observed a similar behaviour to those already discussed. For example, the top 5 similar words to the verb “murdered” in the past simple tense (tagged as VBD) are all verbs that describe criminal events, only one of them can be considered as a synonym (“killed”). On other hand, the verb “treated” in the past participial tense (tagged as VBN) was similar to verbs that describe events which usually take place at a hospital or medical institution (the medical sense of the verb treat). Interestingly, the verb “treated” in the simple past tense (tagged as VBD) does not show a

similar behaviour. This observation suggests that fine-grained POS tags might help distinguishing different senses of homographs as well as different senses of the verbs. Finally, the last entry in Table 1 shows the adverb “quickly” and its top 5 similar words.

3.3. Does Distributional Similarity Indicate Synonymy?

To further understand the nature of the distributional similarity between words, we conduct an investigation on a set of target words that compiled by the author. To narrow the scope of our investigation, we only considered adjectives for this investigation. We looked in depth into the most similar words to a set of 100 words. We used WordNet as a reference to discover the relation between the similar words if they are linked somehow in this lexical database. If the two words are not linked in WordNet, we perform a manual evaluation to categorize the relation between the two words. The result of this investigation is shown in Table 2. Moreover, we were able to find the following categories of words similarity/relatedness in our investigation.

- *WordNet synonyms*: words which are recognized as synonyms in WordNet.
- *WordNet antonyms*: words which are recognized as antonyms in WordNet.
- *WordNet similar to*: words which are recognized as similar to each other in WordNet.
- *WordNet see also*: words which are connected to each other in WordNet under the see also category.
- *Nearly synonyms*: words which are recognized as synonyms.
- *Specification*: A word is a specific case of another (more general) word.
- *Domain Similarity*: words that occur in the same domain or topic (topically similar words).
- *Contrasting*: words that contrasting but do not qualify as opposites or antonyms.
- *Association*: words that are associated in most context.
- *Shortening*: A word is an orthographic short form of another full word.
- *Similar*: The words are somehow similar, but they are neither synonyms nor antonyms. They do not exhibit any relatedness.

WordNet contains a lot of words that are connected by the synonymy or antonyms relation. However, some adjective pairs might be considered as synonyms by a human evaluator even though WordNet does not recognize them as direct synonyms. For example, the word “appealing” would be considered as synonym to the word “attractive” by most English speakers. But WordNet puts “appealing” under the category similar to the word “attractive”. The same can be said about

the word “beautiful”, which is put under the category see also to the word “attractive”. Therefore, we added these relations to Table 2.

From Table 2, one can observe that only 6 adjective pairs out of the 100 pairs are actually WordNet synonyms. Interestingly, 21 of the pairs are WordNet antonyms. However, many of the pairs that are identified as synonyms by manual investigation are actually within the similar to category in WordNet. We also identified other pairs that are not directly linked in WordNet to be synonyms (we refer to them as nearly synonyms). We found 23 pairs that can be qualified as nearly synonyms. These observations could justify the low precision of the work of Leeuwenberg *et al.* [7] when evaluation was performed against the synonyms in WordNet compared to the manual evaluation which gave much more optimistic results. Perhaps the most interesting example in Table 2 is (natural-liquefied), since the similarity between the two words seems to be unknown at first. We manually checked the corpus for instances of the two adjectives, we found that the two words co-occur in many contexts in the corpus due to the wide use of the expression liquefied natural gas in newswire corpora such as the one we have used in these experiments. We chose to refer to this particular instance of similarity as association. It seems that the Skip-Gram model of Word2Vec capture this kind of association with high similarity score.

Table 2. Counts per category for 100 target adjective with their most similar words.

Category	Most similar (out of 100)	Example
WordNet synonyms	6	Possible-potential
WordNet antonyms	21	Unsuccessful- successful
WordNet similar to	21	Responsible-accountable
WordNet see also	3	careful-cautious
Nearly-synonyms	23	limited-minimal
Specification	5	healthy-heart-healthy
Domain similarity	4	scientific-peer-reviewed
Contrasting	3	personal-work-related
Association	2	natural-liquefied
Shortening	1	professional-pro
Similar	11	proper- adequate

Finally, we found 11 adjective pairs where the semantic similarity somehow exists but does not qualified to be exactly synonyms and cannot be categorized under any of the aforementioned categories. These pairs also are not directly linked in WordNet and we simply refer to this category as similar.

4. Neural Network for Synonymy Identification

From the analysis in the previous section, it is clear that similarity (as measured by cosine similarity) does not

necessarily indicate synonymy. It is quite rare that the most similar word is actually a synonym to the target word. **This raises the question: given a vector representation of two words, can we build a system that classifies whether the two words are synonyms or not? If yes, how to obtain labelled data to train the system?**

We addressed the synonymy identification problem as a classification task. To make the problem simpler and doable within the time-frame of the research, only adjectives were considered for the classification. We used a feed-forward neural network with backward propagation as a learning algorithm. The classifier architecture is shown in Figure 2. **To obtain labelled training data, we extracted synonyms pairs from SimLex-999 similarity lexicon with similarity score > 6:5. SimLex-999 is a gold-standard resource for evaluating distributional semantic models** [4]. According to authors’ own words: “Simlex-999 explicitly quantifies similarity rather than association or relatedness”. Simlex-999 was produced using crowd-sourcing of 500 paid native speakers who were asked to rate the similarity, as opposed to association, of different concepts provided a visual interface. Table 3 shows some examples from the data set.

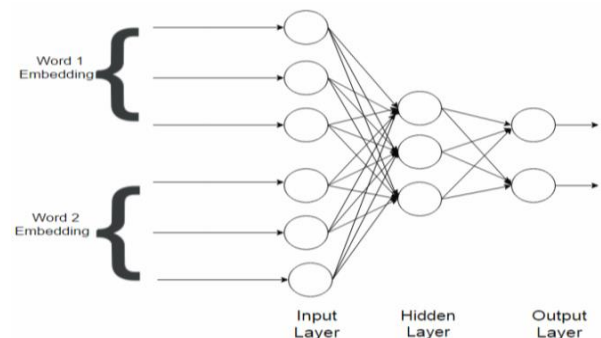


Figure 2. Feed-Forward Neural Network Architecture for synonymy classification.

Table 3. Examples from the SimLex-999 lexicon. Similarity scale ranges from 10 (most similar) to 0 (least similar).

Word1	Word2	POS	Sim-score
old	new	A	1.58
smart	intelligent	A	9.2
plane	jet	N	8.1
woman	man	N	3.33
word	dictionary	N	3.68
create	build	V	8.48
get	put	V	1.98
keep	protect	V	5.4

Considering the examples given in Table 3, one can conclude that SimLex-999 was designed to reflect semantic similarity due to synonymy relation (as shown in the similarity score of smart and intelligent) more than any other relation. For example, even though the words “man” and “woman” denote similar concepts (i.e., co-hyponyms of the hypernym person or human), they received a relatively low similarity score. In addition, and despite the strong functional

similarity between “old” and “new”, they received a low similarity scores. Therefore, the SimLex-999 can be used to perform a quantitative evaluation of the effectiveness of semantic models to reflect synonymy with high confidence.

Then, we expanded the dataset by linking the pairs to the WordNet lexicon. For example, for a given pair (good, great), we searched in the synsets of “good” where “great” is one of the synonyms and included all other synonyms within the same synset. To obtain negative examples, antonyms in the same synset were extracted and antonym pairs were formed. Using this method, we obtained 128 synonymy pairs (SYN) and 91 antonymy pairs (ANT). We added 90 pairs of words that are neither synonyms nor antonyms (ELS) which have been annotated manually.

We performed several experiments on both 2-way classification (SYN/ANT) and 3-way classification (SYN/ANT/ELS). We used the vector representations (the output of word2vec) for each pair as an input to the neural network. In each of the classification runs, the data were split into 75:25 training and testing, respectively. The results are reported in Table 4.

Table 4. Results of the classification experiments.

Experiment	# Hidden units	Accuracy
2-Class	20	0.76
	134	0.74
3-Class	25	0.60
	136	0.63

Even though the training dataset was small, the results for the classification experiments are encouraging. In both cases (2-class and 3-class), the accuracy is way above random guessing, which indicate that supervised learning is feasible for the synonym identification task using word embeddings as features.

5. Future Work

The work that has been presented in this paper can be extended in a few potential directions. In our work, we used the word embeddings as features to feed the neural network. Thus, the features of the words reflect the linear environments in which they tend to occur. For future work, we intend to enrich the features by adding additional features that are inferred from translational context to help distinguish synonymy from other sense relations. Translational context can be obtained either by either using parallel corpora with word alignment or word-to-word translation dictionaries. Nevertheless, using parallel corpora means that one has to deal with the morphological variations of lexical items in morphology-rich languages.

Another direction is to extend the work presented in this paper is to expand the problem into sense relations identification. That is, given two words that exhibit distributional similarity, the task would be to identify the sense relation of the two words (e.g, synonyms,

antonyms, (co) hyponyms, etc.). This task can be tackled as a multi-class classification in a similar fashion to that we used to address the synonymy identification task.

6. Conclusions

In this paper, we addressed the problem of extracting synonyms from text corpus using word embeddings and supervised neural network. We used word2vec to construct word embeddings and performed a qualitative evaluation of the most similar words to a few target words. Our investigation showed that distributional similarity does not always indicate synonymy but the similarity might be due to other functional similarity (e.g., antonym) or domain similarity (e.g., association). Then, we showed that embeddings constructed using word2vec can be used as features to feed a neural network for synonymy classification task. For future work, we suggested extending our approach to sense relation identification instead of synonymy discovery which can be tackled as multi-class classification task.

References

- [1] Bahashwan M., Abu-Bakar S., and Sheikh U., “Efficient Segmentation of Arabic Handwritten Characters Using Structural Features,” *The International Arab Journal of Information Technology*, vol. 14, no. 6, pp. 870-879, 2017.
- [2] Crystal D. *The Cambridge Encyclopedia of Language*, Cambridge: Cambridge University Press, 1987.
- [3] Harris Z., “Distributional Structure,” *Word*, vol. 10, no. 2-3, pp. 146-162, 1954.
- [4] Hill F., Reichart R., and Korhonen A., “Simlex-999: Evaluating Semantic Models with (genuine) Similarity Estimation,” *Computational Linguistics*, vol. 41, no. 4, pp. 665-695, 2016.
- [5] Jurafsky D. and Martin J., *Speech and Language Processing*, Pearson, 2014.
- [6] Khan K., Baharudin B., and Khan A., “Identifying Product Features from Customer Reviews Using Hybrid Patterns,” *The International Arab Journal of Information Technology*, vol. 11, no. 3, pp. 281-286, 2014.
- [7] Leeuwenberg A., Vela M., Dehdari J., and van Genabith J., “A Minimally Supervised Approach For Synonym Extraction with Word Embeddings,” *The Prague Bulletin of Mathematical Linguistics*, vol. 105, no. 1, pp. 111-142, 2016.
- [8] Levy O. and Goldberg Y., “Dependency-Based Word Embeddings,” in *Proceedings of the 52nd Annual Meeting of the Association for*

- Computational Linguistics*, Baltimore, pp. 302-308, 2014.
- [9] Mikolov T., Chen T., Corrado G., and Dean J., "Efficient Estimation of Word Representations in Vector Space," in *Proceedings of the 1st International Conference on Learning Representations*, Scottsdale, pp. 1-12, 2013.
 - [10] Miller G., "WordNet: A Lexical Database for English," *Communications of the ACM*, vol. 38, no. 11, pp. 39-41, 1995.
 - [11] Pennington J., Socher R., and Manning C., "Glove: Global Vectors for Word Representation," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Doha, pp. 1532-1543, 2014.
 - [12] Toutanova K., Klein D., Manning C., and Singer Y., "Feature-Rich Part-of-Speech Tagging with A Cyclic Dependency Network," in *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, Edmonton, pp. 173-180, 2003.
 - [13] Turney P., "Domain and function: A Dual-Space Model of Semantic Relations and Compositions," *Journal of Artificial Intelligence Research*, vol. 44, pp. 533-585, 2012.
 - [14] Van der Plas L. and Tiedemann J., "Finding Synonyms Using Automatic Word Alignment and Measures of Distributional Similarity," in *Proceedings of The COLING/ACL on Main Conference Poster Sessions*, Sydney, pp. 866-873, 2006.



Nora Mohammed is currently a researcher at the college of engineering, Al-Qadisiyah University, Iraq. Her main research interest is in the field of natural language processing and computational linguistics. She has worked on mining natural language text for information retrieval, relation extraction, and synonymy discovery. She received her Master degree in Computer Science and Engineering from Osmania University, India.