

The Quality of Lexical Semantic Resources: A Survey

Hadi Khalilia
University of Trento / Italy
hadi.khalilia@unitn.it

Abed Alhakim Freihat
University of Trento / Italy
abdel.fraihat@gmail.com

Fausto Giunchiglia
University of Trento / Italy
fausto@disi.unitn.it

Abstract

With the increase of the lexical-semantic resources built over time, lexicon content quality has gained significant attention from Natural Language Processing experts such as lexicographers and linguists. Estimating lexicon quality components like synset lemmas, synset gloss, or synset relations are challenging research problems for Natural Language Processing. Several lexicon content quality approaches have been proposed over years in order to enhance the work of many applications such as machine translation, information retrieval, word sense disambiguation, data integration, and others. In this research, a survey for evaluation the quality of lexical semantic resources is presented.

1 Introduction

Lexical Semantic Resources (LSRs) are lexical databases that organize the relations between their elements (synsets) via lexical and semantic relations. The basic element of LSR is a synset. A synset is a set of lemmas (dictionary form of a word), gloss (natural language text that describes of the synset), and synset examples. A lemma within a synset has a meaning which we call a meaning (sense). Synset examples are used to understand the meaning of a lemma in a synset. For example, the following is a synset:

```
#1 person, individual, someone, somebody, mortal, soul: a human being; "there are too much for one person to do".  
"Person, individual, someone,
```

```
somebody, mortal and soul" are lemmas of the synset, "a human being" is the gloss, and "there are too much for one person to do" is the synset example (Miller et al., 1990). Lexical relations organize the relationships between senses. For example the antonym lexical relation expresses that two senses are opposite in meaning such as love is antonym of hate. Semantic relations organize the relationships between synsets". For example the synset (a) is a hypernym (is-a) of (b) (Miller et al., 1990; Chandrasekaran and Mago, 2021).
```

```
(a) chicken, Gallus gallus: a domestic fowl bred for flesh or eggs; believed to have been developed from the red jungle fowl.  
(b) domestic fowl, fowl, poultry: a domesticated gallinaceous bird thought to be descended from the red jungle fowl.
```

The quality of synset components and also, the quality of its lexical semantic relations are the main factors that influence its quality that participate in increasing or decreasing a LRS quality. Therefore, synset quality measurement is important in order to evaluate the quality of LRSs.

Building LRSs such as WordNet faces many challenges. These challenges are polysemy, missing lemmas, missing senses, and missing relations. For example, one of the main problems that makes Princeton WordNet (Miller and Fellbaum, 2007; Freihat, 2014) difficult to use in natural language processing (NLP) is its highpolysemous nature due

to too many cases of redundancy, too fine grained senses, and sense enumerations. On the other hand, it has several synsets that have missing lemmas and missing relations with other synsets. Also, some lemmas in WordNet have missing senses.

In order to solve these challenges, researchers have proposed three categories of approaches, which are: the category of synset lemmas evaluation approaches, the category of synset gloss evaluation approaches, and the category of synset relations evaluation approaches.

In this survey, these categories are described by tracking the evaluation of synset quality approaches over the past years. Also, the survey focuses on recent researches that have not been covered in the previous surveys.

The paper is organized as follows. In Section 2, we discuss the lexicon quality challenges. In Section 3, we describe the current approaches for synset-quality evaluation. In Section 4, we conclude the paper and discuss future research work.

2 Lexicon Quality Challenges

Lexical-semantic resources face several challenges, they are categorized into two main categories: OVERLOAD work or UNDERLOAD. Inappropriate senses, inappropriate lemmas and inappropriate connections between synsets are needed some extra works and produce OVERLOAD components in LSRs. On the other hand, missing senses, missing lemmas and missing connections between synsets produce UNDERLOAD problem. Therefore, in the following sections we present some of the challenges that produce lexicon with low quality.

2.1 Polysemy

LSR, e.g, WordNet organizes the relation between terms and synsets through senses (term-synset pair). A term may have many meanings (one or more senses) which is called polysemous term. For example, **head** has 33 senses in WordNet which indicates that there are 33 relations between the word head and associated synsets. The ambiguity of a term that can be used (in different contexts) to express two or more different meanings is called polysemy. Due to synonymy and polysemy, the relation between terms and synsets is many-to-many relationship. Really, wrong semantic connection can be occurred in WordNet. A misconstruction that results in wrong assignment of a synset to a term is called *Sense enumeration* (Freihat et al.,

2015).

In WordNet, a compound-noun which contains two-parts (modifier and modified) causes polysemy this is called *compound-noun polysemy*. It corresponds to "the polysemy cases, in which the modified noun or the modifier is synonymous to its corresponding noun compound and belongs to more than one synset". WordNet contains a substantial amount of this type of ploysemy such as: *center* and *medical center* in WordNet (Kim and Baldwin, 2013).

Also in WordNet, a special case is founded when there are related some senses (synsets) with a specific polysemous term and not connected with it. For example, a hierarchical relation between the meanings of a polysemous term (Freihat et al., 2013b). "In case of abstract meanings, we say that a *meaning A* is a more general meaning of a *meaning B*. We say also that the *meaning B* is a more specific meaning of the *meaning A*" which is called *specialization polysemy*. In this case, synset connections require reorganizing the semantic structure (using semantic relations) to cover and reflect the (**implicit**) hierarchical relation between all such senses.

So, the big challenge in WordNet is polysemy, because it may produce OVERLOAD connections (overload of a number of term-synset pairs). For example wrong assignments of a synset to terms in sense enumeration add overload relations in WordNet which decrease the synset quality implicitly.

2.2 Missing Senses, Lemmas and Relations

Despite "the highpolysemous nature of wordNet, there are substantial amount of *missing senses* (term-synset pairs) in WordNet" based on Ciaramita and Johnson's work that cause UNDERLOAD of term synsets problem which is the opposite of the overload of term synsets. For example, new added words in languages cause missing senses (synsets) for some terms in lexical resources (e.g, WordNet). Such as *Crypto Mining* sense is missing from the synsets of *mining* term in WordNet and only two synsets are founded in WordNet for it (Ciaramita and Johnson, 2003).

Also, WordNet contains synsets with *missing lemmas* as shown in (Verdezoto and Vieu, 2011). For example, "the term *brocket* denotes two synsets in WordNet, the lemmas of the two synsets are incomplete. This is due to the following: the terms *red brocket* and *Mazama americana* which are syn-

onyms of the lemmas in (b) are missing. The two synsets do not even include the term *brocket deer*. (a) *brocket*: *small South American deer with unbranched antlers*. (b) *brocket*: *male red deer in its second year*”

WordNet relations are “useful to organize the relations between the synsets, while substantial amount of relationships between the synsets remain **implicit** or sometimes **missing** as in the case synset glosses relations. For example, the relation between *correctness* and *conformity* is *implicit*. The relation between *fact or truth* and *social expectations* in the following two meanings of the term *correctness* is *missing*. A human being may understand that *correctness* is a hyponym of *conformity* and *fact or truth* is a hyponym of *social expectations*, but this is extremely difficult or impossible for a machine because *conformity* is neither the hypernym of (a) nor (b). The relation between *fact or truth* and *social expectations* is *missing* because *social expectations* is not defined in WordNet which makes the two synsets are **incorrect** (Freihat et al., 2013a).

Missing senses, missing terms or missing Relations may cause UNDERLOAD problem whether UNDERLOAD in connections or UNDERLOAD in synset itself. Therefore, to enhance synset quality, you have to solve the two main problems: OVERLOAD and UNDERLOAD which are caused by **polysemy** and **missing**, respectively.

3 Quality Evaluation Methods

Lexicon quality estimation methods evaluate the quality of semantic network that a lexical-semantic resource should have. This work depends on the calculation of the synset (acts as a node in the semantic network) correctness and completeness, and also, depends on the connectivity degree of the synset with other synsets in the semantic network. In this section, we introduce and discuss the methods of the lexicon quality evaluation which contains both manual and automatic evaluation methods for the synset quality dimensions (synset correctness, synset completeness and its connectivity) and further classify the evaluation methods into three categories, including synset terms/lemmas evaluation approaches, synset gloss analytical methods, and synset relations with other synsets measures.

3.1 Synset Lemmas Evaluation Methods

Based on the underlying principle of how the synset lemmas are assessed, synset lemmas evaluation methods can be further categorized as Lemmas Validation Methods, and Lemmas Clustering Methods.

3.1.1 Lemmas Validation Methods

The most famous method for lemmas validation is the work of Ramanand in (Nadig et al., 2008). They presented **Validate Synset** algorithm, its principle depends on “**dictionary definitions** to verify that the words present in a synset are indeed synonymous or NOT”. This is due to the availability of synsets in which some members “do not belong”. To accomplish their work they discussed the following research questions: “is a given WordNet complete, how to select one lexico-semantic network over another, and are WordNet synsets INCOMPLETE (may be many words have been **omitted** from the synset) and are WordNet synsets CORRECT (the words in a synset indeed synonyms of each other and the combination of words should indicate the required sense)”. To answer the questions they try to validate the available synsets which are the foundations of a WordNet. “A WordNet synset is constructed by putting together a set of synonyms that together define a particular sense uniquely. This sense is indicated for human readability by a gloss”. To evaluate the **quality of a synset**, they begin by looking for validating the synonyms that the synset has them. They follow these subtasks in the synset validation: are the words in a synset indeed synonyms of each other? Are there any words which have been omitted from the synset? And does the combination of words indicate the required sense? In their work, they focus on the quality of content embedded in the synsets; this is by attempting to verify a given a set of words/lemmas if they were synonyms and thus correctly belong to that synset or not synonyms based on the following **two principles**: “if two words are synonyms, it is necessary that they must share one common meaning out of all the meanings they could possess. And a condition could be showing that the words replace each other in a context without loss of meaning” (Nadig et al., 2008).

A simple block diagram for a synset synonym validation using the system is shown in Figure 1. As we notice from the block diagram, the **input** to the system is: “a WordNet synset which provides the following information: the synonymous words

in the synset, the hypernym(s) of the synset and other linked nodes, gloss, example usages”. The **output** consists of ”a verdict on each word as to whether it fits in the synset, i.e. whether it qualifies to be the synonym of other words in the synset, and hence, whether it expresses the sense represented by the synset”. They used the following **hypothesis**: ”if a word is present in a synset, there is a dictionary definition for it which refers to its hypernym or to its synonyms from the synset” (Nadig et al., 2008).

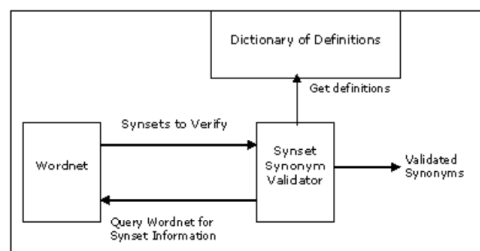


Figure 1: Block Diagram for Synset Synonym Validation.

However, dictionary definitions include useful clues for validating and verifying synonymy. The results show that: the algorithm is simple to implement and depends on the nature (the depth and the quality) of the used dictionary. Many words in WordNet are not validated, around 0.18 of total words in WordNet and 0.09 of total WordNet synsets that couldn’t be validated. Also, the algorithm cannot detect omissions from a synset. To overcome this shortcoming of the algorithm, they proposed that expanding the validation to the synset gloss, and synset relations; using more dictionaries in validation; running the algorithm to other language WordNets, and applying the algorithm on other parts of speech in English. The same team proposed in (Ramanand and Bhattacharyya, 2007) an automatic method for **the synset synonyms and the hypernyms validation** based on new rules: 8 rules in synonym validation and 3 rules for hypernyms validation which is the first attempt of automatic evaluations for synsets in WordNet. They focus on the *synsets* because they are the foundational elements of wordnets and focus on the *hypernymy* hierarchy this is due to its importance in semantic linkages with other synsets. The quality of the synset and its hypernymy ensure the correctness, the completeness and the usability of the resource. They evaluate the quality of a wordnet by ”examining the validity of its constituent synonyms

and its hypernym-hyponym pairs”. The authors defined the synonymy validation as ”the inspection of the words in the synset indeed synonyms of each other or NOT”, and they use the following observation: ”If a word w is present in a synset along with other words w_1, w_2, \dots, w_k , then there is a dictionary definition of w which refers to one or more of w_1, w_2, \dots, w_k and/or to the words in the hypernymy of the synset” which was the hypothesis in the (Nadig et al., 2008) work. In the **synonymy validation algorithm**, the authors apply 8 rules in order which are the basic steps of the algorithm. Also, omissions from synsets aren’t considered. Examples of these are synsets such as: *Taylor, Zachary Taylor, President Taylor*: no definition for the last multiword. Thus the multiword synonyms do share partial words. To validate such multi-words without dictionary entries, they check for the presence of partial words in their synonyms”. They run the algorithm on the *noun synsets* (39840 from the available 81426) of PWN, the inputs of the algorithm are synsets with more than one lemma, by running the validator which uses the online dictionary service *Dictionary.com* in validation, the results show that the percentage of the synsets where all words were validated is (0.701), Pushpak algorithm is simple and acts as a backbone for the synset validation models, also, the applied rules such as: Rule1, Rule2 and Rule7 are the most impact among synonym validation rules, on the other hand Rule4, Rule5 and Rule6 are the lowest. They conclude that many of the words present in PWN aren’t validated and those with rare meanings and usages. ”The wordnet contains synsets that have outlier words and/or missing words”. The limiting factors are ”the availability of dictionaries and tools like stemmers for those languages”. They plan to summarize the quality of the synsets into a single number. The results could then be correlated with *human evaluation*, finally converging to a score that captures the human view of the wordnet. ”The presented algorithm is available only for Princeton WordNet. However, the approach could broadly apply to other language wordnets and other knowledge bases as well. And the algorithm has been executed on noun synsets; they can also be run on synsets from other parts of speech”. Also, in the same area and due to the wide-spread usage of lexical semantic resources, the lexicon quality evaluation became more and more important to tell us how well the applica-

tions and operations based on these resources perform, for example, the authors in (Giunchiglia et al., 2017) describe a general approach to improve the quality of the lexical semantic resources by proposing an algorithm to classify the ambiguity words (based on their senses) in the lexical semantic resources to three classifications for a: polyseme, homonym or unclassified. Also, they present "a set of formal quantitative measures of resource incompleteness". And apply their work and analysis on "a large scale resource, called the Universal Knowledge Core (UKC)". The authors define "two types of incompleteness, i.e., *language incompleteness* and *concept incompleteness*". Language Incompleteness (in a lexical resource): a set of synsets/words/concepts is not lexicalized in a lexical resource (e.g UKC) by a specific language. A model (language incompleteness measurement) that can be used to measure the count (how much) of omitted synsets/words/concepts in the language is described in (Giunchiglia et al., 2017). The notion of "concept incompleteness can be thought of as the dual of language incompleteness. If the language incompleteness measures how much of the UKC a language does not cover, the concept incompleteness measures how much a single concept is covered across a selected set of languages. Concept incompleteness: is the complement to 1 of its coverage". A concept incompleteness model that can be used to measure the concept incompleteness is described in (Giunchiglia et al., 2017). Also in the same research, **lexical ambiguity** is described (it is happened when one word in a language denotes to more than one concept) and they computed the number of ambiguity instances in UKC, e.g., polysemy or homonymy. As an application example they applied the proposed algorithm to "checks whether any two concepts denoted by a single word are polysemes of homonyms or NOT on the UKC concepts". They run the algorithm which consists of 4 steps, and the results showed that, "the UKC contains 2,802,811 ambiguity instances across its pool of 335 languages, these instances were automatically evaluated by the algorithm which, generated 0.32 polysemes among all the ambiguity instances and 0.22 homonyms across all languages". They concluded that when the language coverage increases then the average ambiguity coverage decreases, and vice versa. Also, "increasing the minimal required number of ambiguity instances consistently increases the percentage of

polysemes (up to the 0.74), decreases the percentage of homonyms (down to the 0.11) as well as the percentage of unclassified instances (down to around the 0.15)". Giunchiglia's group presented the language incorrectness evaluation method in UKC in (Giunchiglia et al., 2018), the authors proposed that "the languages in the UKC are far from being complete, i.e., from containing all the words and synsets used in the everyday spoken or written interactions. And far from being correct, i.e., from containing only correct senses, namely, only correct associations from words and concepts to synsets". These limiting factors impact the lexical resource quality. **Language Incorrectness** is the number of psycholinguistic mistakes in a language in a lexical resource per the number of total of concepts in that language in the same resource. They proposed a model to measure the language Incorrectness in (Giunchiglia et al., 2018). Furthermore, this work solves the problem of synset incomplete through presenting a model that transforms the semantic relations nodes from synsets to concepts. This is based the fact that is some words have multiple meanings, and each word is codified as a synset, consisting of a (possibly incomplete) set of synonymous words. The proposed approach describes the UKC design as three-layers: words, synsets and concepts. "Word layer, stores what we call the universal lexicon, the synset layer, stores the world languages, and the concept layer, stores the world (mental) model(s), as represented by the CC". This work makes an improvement in the UKC that influences on its quality; this due the work that becomes a language independent and handles the problem of each synset is associated with one and only one language.

3.1.2 Lemmas Clustering Methods

Lemmas clustering methods retrieve and make clusters for collected synonym lemmas from lexical semantic resources and dictionaries based on the lexical semantic network, for example, the authors in (Lam et al., 2014) proposed an approach that built a new Wordnet from several lexical resources and Wordnets using machine translation (MT- is the core operation for their approach). The presented algorithms use three approaches to generate (translate synsets) synset candidates for each synset in a target language T, the approaches as follow. "1) *the direct translation (DR) approach*: this approach directly translates synsets in PWN to T. 2) *Approach using intermediate Wordnets (IW)*: for each synset,

they extract its corresponding synsets from intermediate Wordnets. Then, the extracted synsets, which are in different languages, are translated to T using MT to generate synset candidates. Synset candidates are evaluated using IW method as shown in Figure 2, also the ranking of synset candidates depends on the ranking equation in (Lam et al., 2014).

3) *Approach using intermediate Wordnets and a dictionary (IWND)*: in this approach, one bilingual dictionary is added to the intermediate languages to T. They translate synsets extracted from intermediate Wordnets to English, then translate them to the target language using English-Target language dictionary. For each synset, they have many translation candidates. A translation candidate with a higher rank is more likely to become word belonging to the corresponding synset of the new Wordnet in the target language". To improve the quality of the Wordnet synsets; feedback and comments from communities (mother-tongue) can be used.

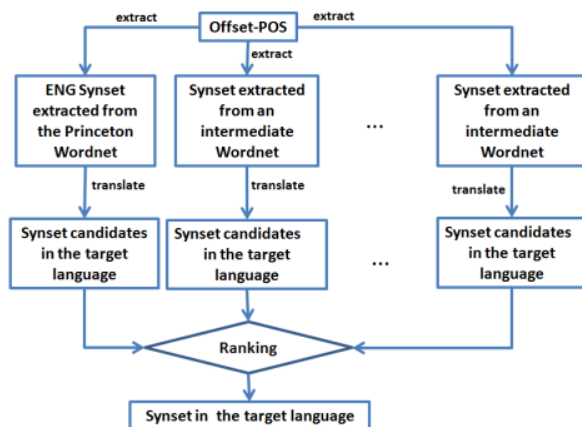


Figure 2: The Intermediate Wordnets Method for Synset Ranking.

Various synset construction (lemmas clustering) methods were proposed combining the **commutative** methods. (Fierdaus et al., 2020) proposed a novel approach for Automatic Indonesian WordNet Development (automatic synset creation). In previous researches, manual methods were used in the establishment of synsets such as the clustering approach. The approach in (Fierdaus et al., 2020) creates synsets from the Indonesian Thesaurus. The input to the system is a set of words which is taken from Thesaurus Bahasa Indonesia (*the used thesaurus is in pdf format which was published in 2008.*), and then the system works on the set of words (input) to find out the semantic similarities between words through successive

steps. The initial input is a word, after that the commutative method is used for the **synset extraction**. Then through the **pre-processing stage**, the system removes excessive characters in the synset that produced from previous stage. After that, the synset that produced from pre-processing will be clustered and combined using **Agglomerative Hierarchical Clustering** algorithm. All these steps are applied for the **Automatic WordNet Development**. Also, the resulted synsets are evaluated using the F-measure method which involves the calculation of precision (P) recall (R) and the evaluation using the gold standard as in (Fierdaus et al., 2020). Commutative method focuses on a commutative relation between the synonyms/lemmas. If a commutative relation is available between the lemmas then the synset will be Valid. Synonym relations should be commutative which means that "if a word k_1 has a synonym k_2 , then k_2 also must be a synonym of k_1 ". Finding a synset that has a valid value is done using a matrix table. Synset extraction is carried out in several steps of the algorithm as follows (Ananda et al., 2018): searching for a sense of the entry word, searching for synonyms on every sense from the previous step, searching for "the chosen word" in the sense that being sought, identify the prospective synset to be sought: by looking for candidates for the synset that can be generated from each item from the words in the dataset, determine whether every word in the prospective synset has a commutative relationship, elimination of candidate synset which is a subset of the other synset and take the remainder of the elimination synset candidate. **Clustering process** is important for making the extracted synset better. They applied Agglomerative Hierarchical Clustering. It is a bottom-up approach that grouped data based on distance value and "the clustering process will be stopped after it reached a condition decided by threshold value". And in (Fierdaus et al., 2020), the authors selected 80 words (as the **test** data) that taken randomly from the thesaurus. And then the system processed the selected word by the system and will produce one or more synset using Agglomerative Hierarchical Clustering method. And also, the authors used the **gold standard** "it finds out how much the correlation between the score issued by the system and the relevance of the words being tested" which is the result of validating synonym sets performed by lexical experts (lexicographers). However, in the validation process F-Measure value for the research

approach was 0.84. It is expected to apply and measure the performance of Agglomerate Hierarchical Clustering and another clustering method on a bigger data scale in the development of synset for Indonesian WordNet. Finally, another synset construction method is **fuzzy synsets** extraction, where fuzzy synsets are a special type of synsets which is discovered from textual definitions. For example, the authors in (Oliveira and Gomes, 2011) present a fuzzy synsets extraction method based on "the term senses are not discrete" fact. Fuzzy synsets are extracted from three (Portuguese) dictionaries **automatically**, which are: Dicionario Aberto and the Portuguese Wiktionary as public domain dictionaries and PAPEL 2.0 as a public domain lexical network. They proposed the following steps in the approach: they specify general textual patterns for extracting synonymy-pairs. Compute the similarity value between terms in synonymy-pairs. Using the similarity value, the clusters are created and they are called fuzzy clusters (fuzzy synsets). Then they build a graph using these fuzzy clusters. Lastly, the built graph participates in creating a fuzzy thesaurus in Portuguese language. The method of creating of fuzzy synsets is based on two stages: First, use a dictionary to extract synonymy graph (where they revealed that the number of collected synsets from Portuguese dictionaries in order to create Portuguese WordNet was larger than the number of synsets in Portuguese thesauri which are discovered manually), and then use the created graph to cluster the words/terms in synsets. Synpairs (two nouns are connected as synonyms) can be extracted from the definitions in dictionaries. Also, they use the clustering-algorithm which is shown in (Oliveira and Gomes, 2011) to make fuzzy synsets clusters which are discovered from the synonymy graph [$G = (N, E)$, where N are the number nodes in G and E are the number of edges in E]. The main steps of the algorithm are: 1) Empty sparse matrix creation. 2) Fill the cells of the Empty sparse matrix with the similarity ratio between of the words in the adjacency vectors. 3) Normalize the cell values in the sparse matrix. 4) extract fuzzy clusters. 5) If two clusters have the same elements then they will be merged in a bigger cluster. The *input* of the algorithm is synonymy graph G and the outputs are the resulting synsets in Portuguese. The authors used the manual evaluation for the created synsets (Padawik thesaurus). The average of corrected synonyms pairs in Padawik is 0.75 and

for the synsets is higher than 0.73. In order to get more improvement, they will focus on each fuzzy synset by specifying individual cut-points, and also to work on new relations between words not only the similarity (synonymous words).

3.2 Synset Gloss Evaluation Methods

Measuring lexical semantic relatedness for a synset or a concept generally requires certain background information about the synset. Such information is often described in the synset **gloss**, which includes a different number of examples. The authors in (Zhang et al., 2011) introduced a new model to measure the semantic relatedness. The model exploits the WordNet gloss and semantic relations as features in building concept vectors. Also, they use other features in the designed model: "**wn-synant** merges WordNet synonyms and antonyms. **wn-hypoer** merges WordNet hypernyms and hyponyms, and **wn-assc** merges WordNet meronyms, holonyms and related, which are features corresponding to associative relations". This work participates in the improvement of the quality of WordNet and Wikipedia operations.

Hayashi and his team used a gloss as an indicator in semantic relatedness is their work in (Hayashi, 2016) paper which measures the strength of the evocation relation between the lexicalized concepts. The authors in (Hayashi, 2016) defined the evocation as "a directed yet weighted semantic relationship between lexicalized concepts". Evocation relations are "potentially useful in several semantic NLP tasks, such as the measurement of textual similarity/relatedness and the lexical chaining in discourse, the prediction of the evocation relation between a pair of concepts remains more difficult than measuring conventional similarities (synonymy, as well as hyponymy/hypernymy) or relatednesses (including antonymy, meronymy/holonymy)" as in (Cramer, 2008). The work in (Hayashi, 2016) made good improvements on evocation relations by applying a novel approach in to prediction of the strength and direction of the evocation relations. For example, PWN dataset includes (39,309) synset pairs. If we compare the work of Y. Hayashi with the results of (Ma, 2013), Y. Hayashi considered "evocation as a semantic relationship between lexicalized concepts, rather than a relation between words", which were considered in (Ma, 2013). Also, the authors in (Maziarz and Rudnicka, 2020) worked on the possibility of the WordNet construc-

tion based on "a distance measure which performs better than other knowledge-based features in evocation relations" (Hayashi, 2016). They used the Dijkstra's algorithm to "measure the distance between nodes (words/synsets) in WordNet structure using a new method for evocation strength recognition based with four types of relations: "wn: pure WordNet relations (directed WordNet edges), g: gloss relations (directed gloss relation instances), polyWN: the set of all pairs of polysemous lemma senses taken from WordNet (bidirectional relations between different senses of the same polysemous lemma) and polySC: the set of all pairs of polysemous lemma senses co-occurring in SemCor corpus" as described in (Chklovski and Mihalcea, 2002). "Dijkstra's distance measuring algorithm was applied on the four structures (one structure for each relation type) to get the minimum points between lexical concept pairs. Then 3-similarity measures are used in each time in order to obtain the best predictions of evocation strength in all cases" (Maziarz and Rudnicka, 2020). Marek Maziarz and his team presented a novel approach for *evocation relation measurement which based on the combination of three types of relations: "gloss relations, pairs of polysemous lemma senses and instances derived from the SemCor corpus, and using the proposed inverse Dijkstra's distance* for improving lexical WordNet structure for the needs of evocation recognition". Like the categorization of methods in the preceding subsection, the next group of methods that we present attempt to explain the importance of the synset gloss in the synset quality evaluation by incorporating an additional examples in the gloss.

3.2.1 Synset Gloss Properties

In this section, we discuss the features and properties of the synset gloss, and how the instructions in (Jarrar, 2006) coverage the gloss properties. In addition, we will define some of the problems that can be solved with help of the synset gloss. Synset gloss writing has several rules and instructions; each synset developer has to apply them during a synset creation for example 6 instructions are explained in the paper of (Jarrar, 2006). In this study, the notion of gloss for ontology engineering purposes and the significance of glosses have been introduced. Gloss is "a useful mechanism for understanding concepts individually without needing to browse and reason on the position of concepts". For example, the work in (Jarrar, 2006) introduced

the notion of gloss for concepts/terms in lexical resources by suggesting a list of instructions for writing a gloss. These **instructions** are the following:

1. "It should start with the principal/super type of the concept being defined. For example, Search engine: A computer program that ..., University: An institution of ...".
2. "It should be written in the form of propositions, offering the reader inferential knowledge that helps him to construct the image of the concept". For example, instead of defining Search engine as "A computer program for searching the internet" one can say "A computer program that enables users to search and retrieve documents or data from a database or from a computer network..."
3. "It should focus on distinguishing characteristics and intrinsic properties that differentiate the concept from other concepts (it is the most important)".
4. "The use of supportive examples is strongly encouraged".
5. "A gloss should not contradict the formal axioms" and vice versa.
6. "It should be sufficient, clear, and easy to understand".

The supportive examples in glosses are important due to "clarify true cases (commonly known as false), or false cases (commonly known as true); and to illustrate and strengthen distinguishing properties". They will implement the proposed algorithm in lexical resources like WordNet. And they plan to investigate "how much the process of validating glosses can be (semi-) automated". The synset gloss is the explanation of the synset that can cause correct or wrong assignment in synset-term pairs. Sense enumeration in WordNet is "one of the main reasons that results in wrong assigning of a synset to a term". The authors in (Freihat et al., 2015) proposed a novel approach to "discover and solve the problem of sense enumerations in compound noun polysemy in WordNet". Compound noun polysemy in WordNet is classified into three types such as: "metonymy polysemy cases where the modified noun belongs to two synsets, one of these synsets is base meaning and the other is derived meaning. The specialization polysemy cases where the modified noun belongs to two synsets, one of these synsets is a more general meaning of the other or both synsets are more specific meanings of a third synset. And Sense enumeration

means a misconstruction that results in wrong assignment of a synset to a term, i.e., assignment the noun modifier or the modified noun as a synonym of the compound noun itself". They reduced "the number of sense enumerations in WordNet without affecting its efficiency as a lexical resource". This research improves the lexicon quality by **removing irrelevant semantic relations** between synsets (Freihat et al., 2015).

3.2.2 Synset Gloss Validation

Synset gloss validation methods are computationally simple but they need much effort when someone works on the gloss validation manually, and the synset resources as lexical databases and dictionaries act as a strong backbone for the gloss extraction models, for example, Purnama and his group presented a supervised learning based approach which is an automatic gloss extractor for Indonesian synsets (Purnama et al., 2015). The main sources and datasets used are web documents containing the gloss of the synsets. The proposed approach includes three main phases which are: **preprocessing, features extraction, and classification phase.**

Preprocessing phase includes several sub tasks such as they fetch a collection of web documents using a search engine, raw text extraction and clean-up, and they extract sentences from gloss candidates. Also, in *the features extraction* phase, seven features are extracted for each gloss: "the number of characters in a sentence, the number of words in a sentence, the position of a sentence in a paragraph, the frequency of a sentence in the document collection, the number of important words in a sentence, the number of nouns in the sentence and the number of gloss sentences from the same word". In the final phase- *Classification*, the supervised learning approach depends on these features to accept or reject the candidate which is a gloss in a test. In the classification operation, they used two models which are: Backpropagation feedforward neural networks (BPFNN) and decision tree DT models. BPFNN is a multilayer architecture with seven input nodes; these nodes represent the extracted features (attributes) that extracted in the second phase. But the output node is used for deciding which one of two classes (ACCEPT or REJECT) through the gloss prediction operation. The nodes are shown in the BPFNN architecture in Figure 3. On the induction using DT, they consider all of the features as continuous value with positive inte-

ger and the internal branch nodes in DT are binary splits, –each node has a label (value) $\leq n$ and $> n$.

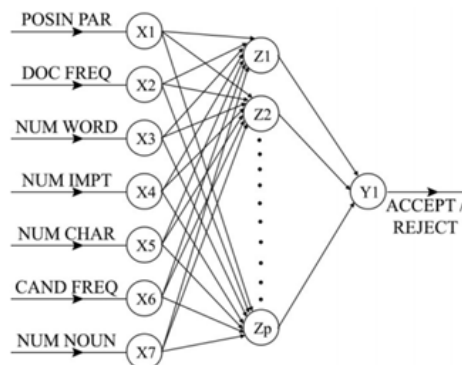


Figure 3: BPFNN architecture for Gloss Candidate Classification.

In this research, the system was successful in collecting 6,520 Indonesian synset glosses, and the accuracy of using the decision tree and BPFNN is then calculated, and the accuracy average is 0.74 and 0.75, respectively. This work represents an improvement in a gloss sentence candidate validation. The authors recommend applying the method of the acquisition of gloss natural languages in the world other than Indonesian.

3.3 Synset Relations Evaluation Methods

In this section, we discuss two types of synset relations: Implicit relations and Special relations. In addition, we will present the sub-types of relations, several examples and how they can coverage the synset-connectivity with other synsets.

3.3.1 Implicit Relations

The methods of synset relatedness based on implicit relations were emerging to measure the lexicon quality such as the work of Bhattacharyya and his team in the research (Nadig et al., 2008). The authors proposed an approach for **hypernymy validation**, this approach receives two synsets as input and states whether they have a hypernym-hyponym relationship between them or NOT". Also, this work is the first attempt for hypernymy validation. The approach consists of three steps (Nadig et al., 2008): *First step*: "prefix forms as an indicator of hypernymy: this is using the following rule: If one term of a synset X is a proper suffix of a term in a synset Y, X is a hypernym of Y" *Second step*: "using web search to validate hypernymy: this is using the following hypothesis: If two words in the form of a Hearst pattern show a sufficient number

of search results on querying, the words can be validated as coming from a hypernym-hyponym synset pair". *Third step*: "using coordinate terms to validate hypernymy: this is using the following hypothesis: If two terms are established to be coordinate terms, a hypernym of one can be stated to be the hypernym of the other". The hypernymy validation was tested on "the set of all direct hypernyms for noun synsets in the PWN". There are a total of 79297 hypernym-hyponym pairs constituting this set. A synset is validated if it gives non-zero search (using Microsoft Live search) results for any 2 of the 9-patterns-Hearst patterns (Hearst, 1992) tested in the algorithm. And "the utilization of coordinate terms is achieved by using Wikipedia as corpus". In all, the authors were able to validate 0.71 of noun hypernymy relation pairs in the Princeton WordNet using their algorithm. The authors concluded in (Nadig et al., 2008) that "many of the synsets present in PWN contain semantic relations may be **inappropriately** set up or may be **missing** altogether".

Also, another example, in (Freihat et al., 2013a) worked on the extraction of the explicit relations from implicit ones in order to enhance WordNet. They added "new explicit hierarchical and associative relations between the synsets which reorganized the semantic structure of the polysemous terms in wordNet". The authors transform "the **implicit** relations between the polysemous terms at **lexical** level to **explicit** relations at the **semantic** level between synsets". However, their approach deals with all polysemy types at all ontological levels of WordNet such as Metonymy, Specialization polysemy, Metaphors and Homographs polysemy. They identified the relations: is-homograph, has-aspect and is-metaphor as extracted semantic relations between synsets. In addition, specific relations for a specialization polysemy are extracted. The explicit relations at the semantic level are: "*Homographs*: there is no relation between the senses of a homograph term. They use the relation is-homograph to denote that two synsets of a polysemous term are homographs. For example, this relation holds between the synsets saki alcoholic drink and saki as a monkey. *Metonymy*: in metonymy cases, there is always a base meaning of the term and other derived meanings that express different aspects of the base meaning. For example, the term chicken has the base meaning a domestic fowl bred for flesh or eggs and a de-

ived meaning the flesh of a chicken used for food. To denote the relation between the senses of a metonymy term, they use the relation has-aspect, where this relation holds between the base meaning of a term and the derived meanings of that term. *Metaphors*: in metaphoric cases, they use the relation is-metaphor to denote the metaphoric relation between the metaphoric meaning and literal meaning of a metaphoric term. For example, this relation is used to denote that cool as great coolness and composure under strain is metaphoric meaning of the literal meaning cool as the quality of being at a refreshingly low temperature". Also, in some cases (e.g. in *Specialization polysemy*), the authors suggested to add a new (missing) parent; they established a new (missing) is-a relation and affix a number of synsets to one synset. This work improved the WordNet quality by "transforming the implicit relations between the polysemous senses at lexical level into explicit semantic relations", and they used the manual evaluation to measure the quality of the approach. The approach was applied on all polysemous nouns. So, they recommended applying the algorithm to handle "verbs, adjectives and adverbs".

Finally, a new research in implicit relations is the paper of T.Dimitrova and her group (Dimitrova and Stefanova, 2019). They have added semantic relations between *nouns* in WordNet that are *indirectly* linked via verbs and adjectives. This assigns new semantic properties to nouns in WordNet. The work reveals **hidden** (indirect) semantic relations between nouns (a noun-noun pair) by using information that is already available from the inter-POS derivative and (morpho) semantic relations between noun – verb, and noun – adjective synsets. "Most relations between synsets connect words of the same part-of-speech (POS), such as : noun synsets are linked via hypernymy / hyponymy (superordinate) relation, and meronymy (part-whole) relation, verb synsets are arranged into hierarchies via hypernymy / hyponymy relation, adjectives are organized in terms of antonymy and similarity, and relational adjectives (pertainyms) are linked to the nouns they are derived from, and adverbs are linked to each other via similarity and antonymy relations". But the authors work on the following *two main categories* for hidden semantic network extraction (Dimitrova and Stefanova, 2019):

1. *Noun – noun relations through verbs*: noun

synsets that are derivationally related to a verb synset and linked through semantic relations that are inherited from the (morpho)semantic relations between noun and verb synsets. The authors worked on 10 categories of the relations, as follow: Instrument Relation, Actor Relation, Causator Relation, Agent Relation, Theme Relation, Result Relation, Location Relation, Uses Relation, Property Relation, and Time Relation

2. *Noun – noun relations through adjectives*: both sides of the relations in this category are nouns and connected through adjectives. 4 types are selected for the category relations, as follow: Property Relation, Part-of Relation, Related Relation and Result Relation.

Dimitrova's work participated in the increasing of the SYNSET CORRECTNESS ratio. They identified the semantic relations between nouns in WordNet that are indirectly linked via derivative relations through verbs and adjectives. Also, the formulated relations will not only increase the inter-relatedness and density of WordNet relations but would allow us to assign new semantic properties to nouns, these properties will explicitly assist the synset to interconnect with the appropriate synsets (senses) that also, improve the SYNSET CORRECTNESS (Dimitrova and Stefanova, 2019).

3.3.2 Special Relations

Special types of synset relations are discussed in this section, these relations added more semantic properties on the synset lattice in lexicons, such as the work of Hayashi that was discussed in Section 3.2, they proposed "a supervised learning approach to predict the strength (by regression) and to determine the directionality (by classification) of the evocation relation that might hold between a pair of lexicalized concepts PWN evocation dataset" (Hayashi, 2016). The authors used neural network (NN) model for classifying evocation relations into **FOUR** categories which are: "outbound", "inbound", "bidirectional" and "no-evocation". And the forest regression model for the prediction of evocation strength is presented. The features of evocation relations are: *Similarity/relatedness features*: 4 similarity/relatedness features are utilized; two of them are synset-based such as "wupSim computes Wu-Palmer similarity which gives the depth of node s from the root" whereas others are word-based such as "ldaSim feature provides the cosine similarity between the word vectors". *Lexical resource features*: these features have been

captured some asymmetric aspects of evocation relationships such as lexNW that finds "the difference in graph-theoretic influence of the source/target concepts in the underlying PWN lexical-semantic network". And *Semantic relational vectors*: in this feature category, they depended on the rule of (Mikolov et al., 2013). "all pairs of words sharing a particular relation are related by the same constant (vector)" to implement the features of the evocation relation. This paper proposed "a supervised learning approach to predict the strength and to determine the directionality of the evocation relation between lexicalized concepts"; which directly impacts the synset connectivity through improving the strength and directionality measurements. The best case in their experiments was the combination of the proposed features "Similarity/relatedness features, Lexical resource features and Semantic relational vectors" which outperformed the individual baselines (Hayashi, 2016). In addition, the authors of the paper (Maziarz and Rudnicka, 2020) focused on a special type of evocation relation which is polysemy, in order to recognize evocation strength. Strong polysemy links participate in constructing a high-quality lexical resource. The framework consisted of three **steps**. *First*: they studied the topologies (3-topologies) of the network of polysemy (graphs of senses). All relations of these topologies are polysemy. Spearman's correlation is used for evaluating the similarity measure in the 3-topologies in order to choose the best topology for lexical resource construction. *Second*: the evocation strength is measured based on the selected topology in step 1 and using the Neural Network and Random Forest models. Also, the authors presented a novel approach which is based on Dijkstra's algorithm to calculate distances between lexical concepts in WordNet structure, and using three types of relations: "A complete polysemy graph (WN-g-co): for a given lemma- linked all senses together". SemCor-based polysemy graph (WN-g-sc): an incomplete graph built by extracting polysemy links from SemCor. It makes groups for such sense pairs that co-occur in the corpus, giving poor completeness but probably good precision". And "the chaining graph (WN-g-ch) tries to connect senses based on contemporary semantic relations between senses of all polysemous words/lemmas that are the closest as in the WordNet graph using the nearest-neighbor chaining algorithm". The **chaining topology** is the best one among the three

listed topologies for lexical resource construction. Therefore; the polysemy network of the lexical resource structure is constructed using "chaining procedure executed on individual word senses of polysemous lemmas". In the measuring of the evocation strength, the work (Maziarz and Rudnicka, 2020) used the inverse of Dijkstra's Distance. For each synset in the evocation set, they "calculated the $dist_{Dijkstra}$ measure and its inverse to find the evocation strength" using semantic relational vectors and the lexical resource features (Hayashi, 2016). Third, they applied Neural Network-NN and Random Forest-RF models to measure evocation strength on the chaining topology with the selected features. Good accuracy in the measurements of the evocation strength is resulted from applying both NN and RF models. Therefore, the authors recommended to utilizing the results in the applications of the polysemy such as Word Sense Disambiguation and Information Retrieval.

Conclusion

Three categories of approaches that influence synset quality in lexical semantic resources which are used in NLP applications were discussed. These are: synset lemmas evaluation category, synset gloss evaluation category, and synset relations evaluation category.

The challenges of synset quality were also discussed, these challenges might cause an OVERLOAD or UNDERLOAD the number of LSR components. They also negatively affect lexicon quality.

These approaches were related explicitly or implicitly with synset quality. Despite of each approach gave a good solution, it couldn't solve all problems/challenges of synset quality. They presented partial solutions that handled with one or two challenges at most. Each approach was a complement to each other as shown in Table 1. It shows a tabulation of these approaches according to synset quality dimensions that are influenced by the challenges. A comprehensive definition for synset quality and an approach that evaluated synset quality with all categories weren't studied in previous researches. An approach that combines all these partial solutions to reach a comprehensive evaluation of LSR quality is recommended.

Section	Completeness	Correctness	Connectivity
3.1.1	✓	✓	
3.1.2	✓	✓	
3.2.1		✓	✓
3.2.2		✓	
3.3.1			✓
3.3.2			✓

Table 1: The coverage of discussed methods

References

- I Putu Prima Ananda, Moch Arif Bijaksana, and Ibnu Asror. 2018. Pembangunan synsets untuk wordnet bahasa indonesia dengan metode komutatif. *eProceedings of Engineering*, 5(3).
- Dhivya Chandrasekaran and Vijay Mago. 2021. Evolution of semantic similarity—a survey. *ACM Computing Surveys (CSUR)*, 54(2):1–37.
- Timothy Chklovski and Rada Mihalcea. 2002. Building a sense tagged corpus with open mind word expert. In *Proceedings of the ACL-02 workshop on Word sense disambiguation: recent successes and future directions*, pages 116–122.
- Massimiliano Ciaramita and Mark Johnson. 2003. Supersense tagging of unknown nouns in wordnet. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 168–175.
- Irene Cramer. 2008. How well do semantic relatedness measures perform? a meta-study. In *Semantics in Text Processing. STEP 2008 Conference Proceedings*, pages 59–70.
- Tsvetana Dimitrova and Valentina Stefanova. 2019. On hidden semantic relations between nouns in wordnet. In *Proceedings of the 10th Global Wordnet Conference*, pages 54–63.
- Valentino Rossi Fierdaus, Moch Arif Bijaksana, and Widi Astuti. 2020. Building synonym set for indonesian wordnet using commutative method and hierarchical clustering. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 4(3):778–784.
- Abed Alhakim Freihat. 2014. *An organizational approach to the polysemy problem in wordnet*. Ph.D. thesis, University of Trento.
- Abed Alhakim Freihat, Fausto Giunchiglia, and Biswanath Dutta. 2013a. Regular polysemy in wordnet and pattern based approach. *International Journal On Advances in Intelligent Systems*, 6.
- ABED ALHAKIM Freihat, FAUSTO Giunchiglia, and BISWANATH Dutta. 2013b. Solving specialization polysemy in wordnet. *International Journal of Computational Linguistics and Applications*, 4(1):29.

- Abed Alhkaim Freihat, Biswanath Dutta, and Fausto Giunchiglia. 2015. Compound noun polysemy and sense enumeration in wordnet. In *Proceedings of the 7th International Conference on Information, Process, and Knowledge Management (eKNOW)*, pages 166–171.
- Fausto Giunchiglia, Khuyagbaatar Batsuren, and Gabor Bella. 2017. Understanding and exploiting language diversity. In *IJCAI*, pages 4009–4017.
- Fausto Giunchiglia, Khuyagbaatar Batsuren, and Abed Alhakim Freihat. 2018. One world—seven thousand languages. In *Proceedings 19th International Conference on Computational Linguistics and Intelligent Text Processing, CiCling2018, 18-24 March 2018*.
- Yoshihiko Hayashi. 2016. Predicting the evocation relation between lexicalized concepts. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1657–1668.
- Marti A Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Coling 1992 volume 2: The 15th international conference on computational linguistics*.
- Mustafa Jarrar. 2006. Position paper: towards the notion of gloss, and the adoption of linguistic resources in formal ontology engineering. In *Proceedings of the 15th international conference on World Wide Web*, pages 497–503.
- Su Nam Kim and Timothy Baldwin. 2013. Word sense and semantic relations in noun compounds. *ACM Transactions on Speech and Language Processing (TSLP)*, 10(3):1–17.
- Khang Nhut Lam, Feras Al Tarouti, and Jugal Kalita. 2014. Automatically constructing wordnet synsets. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 106–111.
- Xiaojuan Ma. 2013. Evocation: analyzing and propagating a semantic link based on free word association. *Language resources and evaluation*, 47(3):819–837.
- Marek Maziarz and Ewa Rudnicka. 2020. Expanding wordnet with gloss and polysemy links for evocation strength recognition. *Cognitive Studies— Études cognitives*, (20).
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.
- George A Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J Miller. 1990. Introduction to wordnet: An on-line lexical database. *International journal of lexicography*, 3(4):235–244.
- George A Miller and Christiane Fellbaum. 2007. Wordnet then and now. *Language Resources and Evaluation*, 41(2):209–214.
- Raghuvar Nadig, J Ramanand, and Pushpak Bhattacharyya. 2008. Automatic evaluation of wordnet synonyms and hypernyms. In *Proceedings of ICON-2008: 6th International Conference on Natural Language Processing*, volume 831. Citeseer.
- Hugo Gonçalo Oliveira and Paulo Gomes. 2011. Automatic discovery of fuzzy synsets from dictionary definitions. In *Twenty-Second International Joint Conference on Artificial Intelligence*.
- I Purnama, Mochamad Hariadi, et al. 2015. Supervised learning indonesian gloss acquisition. *IAENG International Journal of Computer Science*, 42(4).
- J Ramanand and Pushpak Bhattacharyya. 2007. Towards automatic evaluation of wordnet synsets. *GWC 2008*, page 360.
- Nervo Verdezoto and Laure Vieu. 2011. Towards semi-automatic methods for improving wordnet. In *Proceedings of the Ninth International Conference on Computational Semantics (IWCS 2011)*.
- Ziqi Zhang, Anna Lisa Gentile, and Fabio Ciravegna. 2011. Harnessing different knowledge sources to measure semantic relatedness under a uniform model. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 991–1002.