An Enhanced WordNet Query Expansion Approach for Ontology Based Information Retrieval System

Enesi Femi Aminu[1(✉)], Ishaq Oyebisi Oyefolahan[2], Muhammad Bashir Abdullahi[1], and Muhammadu Tajudeen Salaudeen[3]

[1] Department of Computer Science, Federal University of Technology Minna, Minna, Nigeria
{enesifa,el.bashir02}@futminna.edu.ng

[2] Department of Information and Media Technology, Federal University of Technology Minna, Minna, Nigeria
o.ishaq@futminna.edu.ng

[3] Department of Crop Production, Federal University of Technology Minna, Minna, Nigeria
mtsalaudeen@futminna.edu.ng

Abstract. Ontology-based information retrieval is described as a cutting-edge approach capable to enhance the returns of semantic results from documents. This approach works better when similar and relevant terms are added to user's initial query terms using data sources such as wordnet; such technique is known as query expansion. However, the precision of the added term(s) tends to be inaccurate because of the existing WordNet's deficit to handle inflected forms of words. In lieu of this development, this research aims to design Rule based Web Ontology Language (OWL) Information Retrieval System with an enhanced wordnet for query expansion but only limited to the noun subnet database. A combined ontology development methodology was implored; and OWL-2 to develop the ontology for a novel domain of maize crop considering primarily soil, fertilizer and irrigation knowledge. Its rule-based ontology because Competency Questions were modeled using First-Order-Logic (FOL) and encoded with Semantic Web Rule Language (SWRL). Similarly, the wordnet was enhanced on python environment considering the lemmatization's lookup table and the third party modules of Natural Language Tool Kits (NLTK), pattern.en and enchant. Therefore, in this research, the improved wordnet can handle inflected word without stemming it to the root word. It also correctly suggested related words in the case of user's wrong spelt word thereby; reduces minimally time wastage and fatigue. This development invariably aids ontology validation along with the other forms of validations carried out. The research ultimately offers an effective ontology-based information retrieval system based on the proposed algorithmic framework.

Keywords: Inflected words · Maize ontology · Query expansion · Soils fertilizer and irrigations knowledge · WordNet

S. Misra and B. Muhammad-Bello (Eds.): ICTA 2020, CCIS 1350, pp. 675–688, 2021.
https://doi.org/10.1007/978-3-030-69143-1_51

1 Introduction

In this present age, while the exponential growth of data across different repositories is described as heartwarming development; but on the other hand, it also present a challenge for efficient retrieval of relevant information [1]. No doubt that the evolution of semantic based Information Retrieval (IR) is gradually negating the syntactic forms of IR. This development is attributed to the word-based depiction of initial query and corpus documents in repositories of syntactic techniques, which is seen as its pitfall [2]. Based on literature, the challenges of retrieving relevant results as to user's intent is most often characterized by unstructured formats of data that is, data are not machine represented. Thereby, making it difficult for machine to understand and gives due meaning to user's query. To this end, literature have identified a cutting-edge technology known as semantic approach as a proven and possible solution [3]. Semantic search is a type of IR system that operate based on linguistic and knowledge models to proffer solutions to the limitations of keyword based search; that is, syntactic approach [4]. Semantic approach, on one hand entails ontology; and information retrieval technique on the other hand [5]. As a matter of emphasis, ontologies largely utilize semantic web's technologies for modeling [6]. While ontologies provide the platform of representing knowledge in a structure format, information retrieval on the other hand focuses on providing (additional) meaning of data about data in order to achieve relevant hit of information.

Ontology has many definitions but one of the most acceptable definitions according to literature is Gruber [7], which states that ontology is a formal and explicit specification of a shared conceptualization. Similar to standard software development, ontology development equally follows some laid down procedures otherwise known as ontology development methodologies. Examples are Fox and Gruninger, Methontology, Uschold and Kings and Food and Agriculture Organization of United Nations (FAO) Based methodologies. However, the scenario ontology developed in this research work to incorporate the proposed enhanced information retrieval technique is developed based on adoption of combined methodologies as proposed in the review literature of [8]. The adopted methodology consists of six activities or processes, which are collection of domain knowledge; specification of ontology's terminologies; definition of competency questions for ontology's purpose and scope; ontology formalization; ontology evaluation and ontology evolution. Thus, task-based or rule ontology has been developed for this proposed system. It is beyond domain ontology because Semantic Web Rule Language was implemented and the domain under consideration is primarily soil, fertilizer and irrigation knowledge for maize crop as an OWL rule-based ontology. And based on literature covered, the domain is considered novel to be ontologically design.

Information Retrieval is described as science of retrieving data or information relevant to user's need. Therefore, the challenge is no longer availability of information, but retrieving the relevant information according to user's intent [9]. This research issue is commonly attributed to the natural language ambiguities or word mismatch issue, for instance; synonyms and polysemy [10]. Naturally, words are synonymous that is; different words with a common meaning such as maize, corn or *zea may*. Similarly, most natural words are equally polysemous that is; a word that gives more than one descriptions. For example; ear which means the sense organ for hearing or fruiting spike of a cereal plant (maize). Obviously, these words ambiguities make it difficult for existing methods to retrieve accurate and balanced information without much compromise on recall and precision of results. Therefore, to proffer solutions to this issue, the technique of extending initial query constructs otherwise known as query expansion considering the query input's similarity or relatedness has been reliably considered in the literature of information retrieval.

Technically, literature [11] and [12] defined Query Expansion as process of adding useful terms to initial query terms manually, interactively or automatically. Essentially, the addition of meaningful terms is carried out and derived from data sources or knowledge collections such as wordnet, domain ontologies and the likes [13, 14]. In this research work, an enhancement of wordnet for query expansion is considered for the proposed framework of the ontology-based information retrieval system. The motivation of this improvement is as a result of the notable issue of the existing wordnet which lacks capacity to precisely output the correct results of word inflected forms [15]. Word inflected forms are parts of speech (for instance, noun) that can stem out existence in various forms such as addition of s, es, to form the plural forms of regular noun and different plural forms entirely for irregular noun from the root word. The existing wordnet works by only assume the root singular word. This issue clearly affects the precision and recall of additional terms from the wordnet to perform query expansion.

Therefore, this research work aims to design an OWL rule ontology-based information retrieval framework of an enhanced WordNet for query expansion approach. The remaining sections of this paper is organized as follows: Sect. 2 dealt with related studies of the subject matter and Sect. 3 presented the framework of the proposed system. The improvement of wordnet for query expansion was accounted for in Sect. 4. While Sect. 5 discusses the results, conclusion and suggested future work was presented in Sect. 6.

2 Related Works

Ontology-based information retrieval is described as a reliable approach that is capable of improving returns from sematic documents [16]. Besides the relevant methodology to develop ontology for information retrieval system, knowledge representation languages such as RDF/S, OWL [17]; and Protégé, TopBraidComposer [18, 19] as editors are equally required to formalize ontology. Despite the emergence of other document collections such as Wikipedia for query expansion, the relevance of wordnet in IR remains steady however, not without notable gap such as its capability to handle inflected forms of words. WordNet, a lexical resource is considered to be one of the global largest word collection corpus that offer a hierarchical structure of Synset (set of one or more synonyms) and semantic properties of every words [20, 21]. Wordnet has three databases. These are noun, verb, adjective and adverb but this research is limited to noun database.

In the research work of [22], a framework was proposed to enhanced query expansion for efficient information retrieval on ontology-based system. The aspect of enhancement was particularly on wordnet's issue of effectively handling inflected forms of word. However, it was not implemented and the algorithm for the system does not take into account ontology evolution. In the quest of noticeable gaps to expand query semantically, [23] carried out a survey research on query augmentation with the aid of sematic data sources. Incidentally, issue of increase in the number of expansion terms was identified for mixed mode technique despite being described to optimally perform. Therefore, the proposed research has duly considered it. The work of [2] observed that some concepts such as proper nouns, new words and other technical terms are not contained in wordnet and domain ontologies. The research therefore, proposed a new approach of obtaining a more accurate semantic search by considering Wikipedia along with wordnet and domain ontology. However, the issue of synsets, inflected terms and technical terms of both proper and improper nouns of wordnet remain unsolved. Similarly, reference [10] implored Wordnet and Wikipedia as a more effective data sources and novel technique for query reformulation in order to gain more results that are relevant. However, the issue of synsets to be selected for query expansion when search term reflects in several synsets remain unresolved. A technique of candidate concepts expansion not only from the set of synonyms of the user query, but also considered the synsets of the synsets was proposed. However, not all concepts form set of synonyms in wordnet.

Furthermore, [1] carried out a recent and well extensive survey work on the application of query expansion approach for retrieval of relevant information spanning from the year 1960 to 2017. The techniques reviewed covered both the old approaches such as relevance feedback and the recent and trending approaches (wordnet and Wikipedia). Specifically, data sources, semantic similarity functions and user participation and application were the factors considered to drawn their position for their comparisons and variances. In the end, the survey opened up several field of studies where query expansion could be applied. For example, in the area of Information retrieval systems since there is always a desire to personalize query results owing to user's query and intended results. Consequently, there is also need for IR systems to be manipulated by personalize query argumentation techniques. Hence, the researchers submitted that in the nearest future personalization of web search hits would offer a significant role in the research of Query Expansion (QE) technique. In addition, the importance of QE is also crucial in ontology mapping when used for information retrieval. For example, the precision of relevant results in the robust research works of [24, 25] could also be enhanced when the technique is implored. Finally, the accuracy and efficiency of this technique can be judged by considering precision and recall metrics of IR [26].

3 Framework of Ontology Based Information Retrieval (IR) System

In this research work, the proposed ontology based retrieval system is on maize crop domain but primarily limited to its soils, fertilizers and irrigation knowledge as shown by Fig. 2 following the adopted ontology development methodology. Figure 1 presented schematic representation of the proposed system in a conceptual framework. The ontology is designed to be rule based as clearly shown by Table 1. More importantly, the component shown in blue colour by Fig. 1 signifies the improved corpus collection (wordnet) in the aspect of handling word inflected forms for expansion of initial user's query. It would mitigate the issue of recall and precision metrics of information retrieval when fully implemented with the entire system. At this point, it is good to mention that the work is in progress and nearly at the point of evaluation.

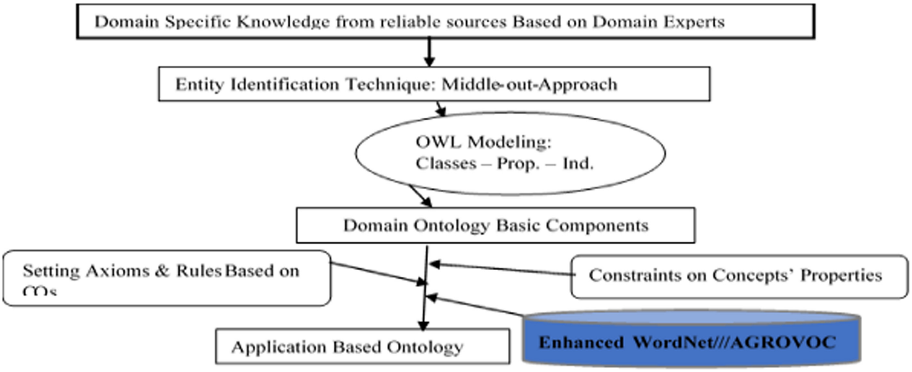

Fig. 1. Conceptual framework of ontology based IR system. (Color figure online)

From Fig. 1, data were collected from reliable research articles and books, published authoritative online data sources, trusted institutes such as CMMYT, IITA, The Institute for Agricultural Research (IAR), Zaria. Subsequently, the collections were validated by domain experts. Middle-out-approach is implored as concept identifying technique; because it first identify the most important concepts; then generalized, and specialized into other concepts. The next step formalized the concepts using OWL, the most highly expressive language that comprises of three major components as Classes, Properties and Individuals but denoted in the figure as ***classes, prop and ind*** respectively. These processes lead to domain ontology. However, the ontology based information retrieval in this research work goes beyond the development of domain ontology or light weight ontology. It is extended by making it tasking and more intelligent by enforcing high-level constraints on concepts' properties (object and data properties of classes) and considered rules and axioms. First-Order-Logic is used to model the competency questions based on contextual information provided by the group of domain experts for axioms and rules; implemented in protégé editor of 5.5.0 via SWRL. Table 1 presents some CQs that were encoded in SWRL and as well query with SQWRL. More importantly, in order to ensure credible information retrieval, an enhanced wordNet that is adequately capable to handle inflected forms of user's query was developed. This is to ensure balanced query augmentation for the proposed system.

The core concepts of the ontology designed includes but not limited to *Maize-Crop/MaizeSeeds* with super class *Crop, Soils, SoilClassificationMethods Fertilizers, FertilizerApplication_Methods, Irrigation, Irrigation_Methods,* as evidently shown in Fig. 3 by OntoGraf of protégé. All the core concepts are subclasses to the main super class *owl:Thing.*

The core concepts of the ontology is well represented by Fig. 2 such as *Soils, Fertilizers, Irrigations* and their related concepts. Currently, the ontology has a total of 4999 OWL axioms, 309 OWL Classes, 423 OWL Object properties, 174 Data properties, 386 Individuals and over 60 SWRL rules based on CQs.

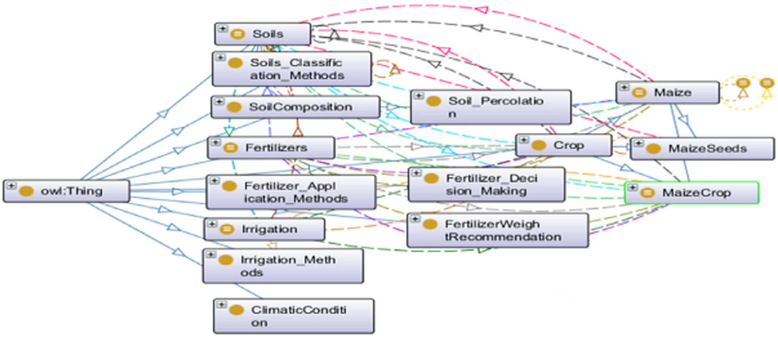

Fig. 2. Ontology's core concepts

The Table presented some samples of the CQs encoded with SWRL and expectedly query with SQWRL. Each of the rule is executed successfully with Drool engine and transferred back the inferred axioms to OWL model that evidently depict the rule richness of the ontology.

Table 1. Querying of rule based ontology

Input Query (Informal CQs)	**Semantic Web Rule Language (SWRL)**	**Semantic Query-Enhanced Web Rule Language (SQWRL)**
What is the best soil PH (range) for maize cropping?	Soils(?s)^SoilPHvalues(?p)^ MaizeCrop(?c)^MaizeSeeds(?e)^ Maize_Varieties(?v) ^ SoilFertility(?f) -> canGrowWellOnTheRangeOf(?c,5-7_PHvalues)^ underWhichMaizeCanGrowdependsOn(?p, ?e) ^ underWhichMaizeCanGrowdependsOn(?p, ?v) ^ underWhichMaizeCanGrowdependsOn(?p, ?s) ^ underWhichMaizeCanGrowdependsOn(?p, ?f) ^ mayBeUsedToCorrect(?p, Liming)	SoilPHvalues(?p). sqwrl:makeSet(?s1, ?p). sqwrl:size(?d,?s1)^SoilFertility(?f)^MaizeSeeds(?e) . sqwrl:makeSet(?s2, ?e). sqwrl:size(?z, ?s2) ^ areInVariousRangesSuitableToGrow(?p, ?e)-> suitableOrBestAreVariesToGrow(?p, ?e) ^ sqwrl:select(?p)
Which type of fertilizer is suitable for maize cultivation/ what is the best fertilizer type for maize?	Fertilizers(?f)^OrganicFertilizers(?o) InorganicFertili ers(?i)^MaizeCrop(?m)^Urea(?u)^NPK_Fertiliz er(?n)^ isAheavyFeederOf(?m,Nitrogen)^strictlyContains(?u,Nitrogen)^containsGoodProportionOf(?n, Nitrogen)^ isAveryRichFormsOf(PoultryDropping, ?o)^containsAlotOf(PoultryDropping, Nitrogen) ->isTheMostSuitable(PoultryDropping, ?o) ^ isTheMostSuitableForCultivating(?n, ?m) ^ isTheMostSuitableForCultivating (?u, ?m)	Fertilizers(?f).sqwrl:makeSet(?s1, ?f) ^ MaizeCrop(?m) ^ sqwrl:makeSet(?s2, ?m) ^ Urea(?u) ^ NPK_Fertilizer(?n) ^ isAheavyFeederOf(?m, Nitrogen) ^ containsMainlyNitrogen(?u, Nitrogen) ^ containsGoodProportionOf(?n, Nitrogen) ^ containsAlotOfNitrogen(?a, Nitrogen) ^ AnimalManures(?a)^containsAlotOfNitrogen(PoultryDropping,Nitrogen)-> isA_SuitableFertilizerToGrow(?a, ?m) ^ isA_SuitableFertilizerToGrow(?n, ?m) ^ isA_SuitableFertilizerToGrow(?u, ?m) ^ sqwrl:select(?n) ^ sqwrl:select(?u) ^ sqwrl:select(?a)
How many times can irrigation be carried out in maize crop?	Soils(?s)^Irrigation(?i)^Soil_Percolation(?p)^Water(?w)^MaizeCrop(?m)^ SoilNutrient(?n)^ClimaticCondition(?c)^ Season(?e)^ dependsOnAvaliabilityOf(?i, ?w)->mayDetermineTheNumbersOf(?p, ?i)^ mayDetermineTheNumbersOf(?c, ?i)^ mayDetermineTheNumbersOf(?e, ?i)	Soils(?s)^Irrigation_Numbers(?i)^ makeSet(?s1, ?i) ^Soil_Percolation(?p)^MaizeCrop(?m)^ClimaticCondition(?c)^ Season(?e) -> mayDetermineTheNumbersOf(?p, ?i)^ mayDetermineTheNumbersOf(?c, ?i)^ mayDetermineTheNumbersOf(?e, ?i) sqwrl:select(?i)

Furthermore, Algorithm1 duly represents the framework and effectiveness of information retrieval from the ontology. An aspect of uniqueness in this proposed algorithm has to do with the combination of the enhanced Wordnet (as shown by Algorithm2) and the proposed popular Agriculture Vocabulary (AGROVOC) database to serve as data sources to perform query expansion. Besides, the algorithm has the capacity to autonomously perform what we called ontology evolution. Valid concepts that form part of user's query but not in the ontology would be automatically added to it. The rationale behind the combination of the two data sources for query expansion is borne out of the fact that AGROVOC would be able to cover up for the Wordnet. That is, some query's concepts that are not single word cannot be handled by wordnet. And more interestingly, considering the local characteristics or peculiarities of the ontology's domain (agriculture), wordnet is deficit to handle such domain solely.

Algorithm1: A Proposed Ontology Based Information Retrieval Algorithm
Input: User's Query
Output: Semantic Results
Parameters: Query-Q (Competency Question); Enhanced WordNet; AGROVOC; DomainOntology; SeedVariables: Maize-M, Soils-S, Fertilizers-F and Irriga tions-I; CandidateTerms-C = y, y is data structure that stores the candidate terms
Procedure:
Step1: *For Each* Input Q
Step2: *Tokenize* Q string
Step3: *While* (i = ! n) { //i and n represent counter and numbers of tokens in a given string respectively
Preprocessed n *further* to extract C // y is created to count and store C
/* During preprocessing, unwanted terms and punctuations are Eliminated. Candidate Terms-C are extracted based on matching the Derived terms (Initial Token minus unwanted terms) to the Domain On tology…. */
ForEach x executes using enhanced WordNet and AGROVOC /* x stands for each candidate terms which is expanded using the data sources*/
Store x on y //y is a candidate terms data structure
Repeat Execution Until i = = y;
}
End Loop
End For
Step4: Query Expansion (QE) formed = {a, b} or $\{i_1, i_2, i_3 \ldots i_n, a_1, a_2, a_3 \ldots a_e\}$ /* a and b represents the candidate terms and new terms added for expansion. Then TF-IDF applies… */
Step5: *Execute* the QE and If successful, Goto 13.
Step6: *If* QE ≠ {a, b}
Step7: *Then* C do not contain >= two (2) seedVariables of owl:Thing
Step8: *ElseIf* C contain one (1) seedVariable of owl:Thing
Step9: *Then* system suggest one or more seedVariable of owl:Thing to form relation. Goto step5
Step10: *ElseIf* C contain >= two (2) seedVariable of owl:Thing and step4 still not formed.
Step11: *Then* activate ontology evolution and Goto step5.
Step12: *Else* input string is out of subject granularity
Step13: *Output* Semantic Result obtained.
EndFor.

In order to have an efficient ontology-based information retrieval, Algorithm1 proposes the framework. From step1 to 5, user input query gets tokenize. For example, a query as *what is the best soil to grow maize?* Breaks down into tokens as what, is, the, best, soil, to, grow, maize,?, so as to pave way for preprocessing where irrelevant concepts or terms and symbols such as *what, is, the, to,?,* in the first instance are discarded. This is as a result of matching the input tokens against the ontology. More so, in order to mitigate the effects of word mismatch (synonyms), the candidate terms are expanded using the two data sources: enhanced Wordnet and AGROVOC. Thus, the expanded relevant terms known as candidate terms are formed and stored in the data structure. At this point, TF-IDF algorithm is implored to assigns weight and ranked the new terms added (as denoted as b in step 4) to finally store out the best ranked. And thereby, execution of {a, b} follows and output the results accordingly.

Step 6 to 12 suggest that if query expansion do not form, the algorithm assumes that the candidate terms considering core knowledge of the ontology which is coded as seedVariable of owl: Thing are not equal to or greater than two so as to form relation. In that case, it is either the candidate terms contain only one term which cannot form relation (step 8) or the relation is formed (step 10) but terms not in the ontology. In the case of step 8, step 9 suggest that the system would add term(s) to form relation and execute. While for step 10 the system would automatically add the terms for the ontology to evolves. Finally, in a situation where the query terms are confirmed to be totally out of ontology purpose and scope, the algorithm would output message that suggest it.

4 Enhanced Wordnet Query Expansion

The use of WordNet, one of the largest corpus collections as data source for query expansion in IR remains highly significant despite its limitations. It has a long-standing history in the area of text processing or natural language processing. However, aside from its limitation to process compound words like Wikipedia, it is also defected in recognizing word in its inflected forms. It only assumes every word based on their root or stem form which inadvertently affects the expected precision of additional terms in query expansion. Therefore, the identified gap motivated the drive to enhance the wordnet considering its enormous advantages in the area of information retrieval.

In this research, Query Expansion (QE) is mathematically defined as follows:

$$\mathrm{QE} = \{\mathrm{i}_1,\ \mathrm{i}_2,\ \mathrm{i}_3 \ldots \mathrm{i}_\mathrm{n},\ \mathrm{a}_1,\ \mathrm{a}_2,\ \mathrm{a}_3 \ldots \mathrm{a}_\mathrm{e}\} \tag{1}$$

This equation one is derived from the second equation below.

$$\mathrm{QE} = \{((\mathrm{i}_1,\ \mathrm{i}_2,\ \mathrm{i}_3 \ldots \mathrm{i}_\mathrm{n} \cup \mathrm{u}_1 \ldots \mathrm{u}_\mathrm{m}) - (\mathrm{u}_1 \ldots \mathrm{u}_\mathrm{m})) + (\mathrm{a}_1,\ \mathrm{a}_2,\ \mathrm{a}_3 \ldots \mathrm{a}_\mathrm{e})\} \tag{2}$$

Where set $\{i_1, i_2, i_3 \ldots i_n \cup u_1 \ldots u_m\}$ represents the initial user's query; set $\{u_1, u_2, u_3 \ldots u_m\}$ represents unwanted terms *m* which is/are contain(s) in *n*. However, the scenarios of *m* in *n* may occur as follows: *m* may be equal, greater or less than *n* (that is, $m > n$, $m > n$, $m > n$). The implication of these scenarios tells the computational overhead (cost) associated with query expansion approach. And lastly, set $\{a_1, a_2, a_3 \ldots a_e\}$ represent the augmented terms. That is, the required relevant and similar terms from the data sources.

In the first instance, the set difference (denoted as -) of the set of initial query terms and the unwanted terms is processed to find the required terms in accordance to the ontology. The unwanted terms consist of stop words, punctuations and the likes. Naturally, initial query terms others known as competency question is a composition or union of relevant and irrelevant query terms. As soon as this first step is achieved, the result (that is, relevant terms otherwise known as candidate term) of the set difference operation is autonomously tagged with similar words from the enhanced wordnet and AGROVOC for query expansion.

Lemmatization technique is implored in improving the inflected forms of word or term taken advantage of its generic lexical lookup table for word's singular and plural. Python environment is used for its implementation considering the NLTK, pattern.en and enchant modules. They make natural language processing easier to modify and improve upon. Figure 3 gives a conceptual representation of how the enhanced wordnet for query expansion is developed.

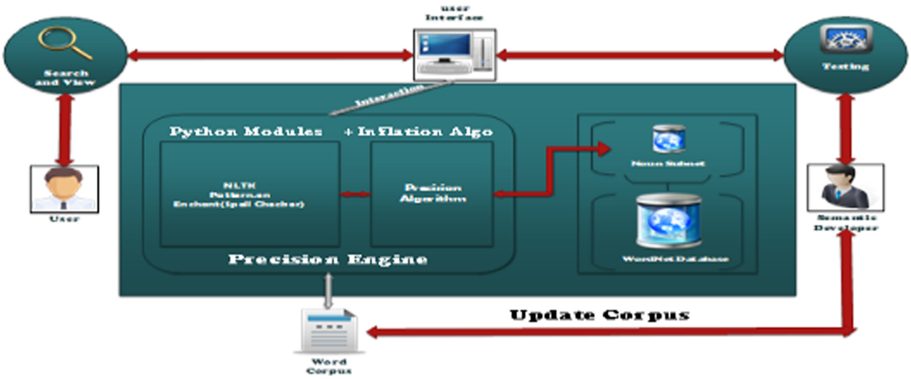

Fig. 3. The enhanced wordnet architecture

The developer accessed the existing wordnet (word corpus) in the python environment via the pattern.en module and NLTK. More so, enchant module was also considered to aid user for cross check spelling in the enhanced version. To deal with the main issue of inflected words, an algorithm named precision was designed as shown by Algorithm2 to finally developed the wordnet. As earlier stated in the previous section, it is only noun subnet or database that is being considered in this research. In addition, an interface was designed to enable end user input search word.

Algorithm2: Word Inflected Forms of (Noun Subnet) Wordnet Algorithm
Procedure:

Step 1: import list of English words
Step 2: initialize list of all English words
Step 3: initialize singular and plural list [data structure]
Step 4: input word and pass through Step1-3
Step 5: if word is not == _exit
Step 6: check singular list
If Found in singular list
Step 7: print -> word is singular
Step 8: if word is spelt correctly
Step 9: return true and print dictionary data
Else
Step 10: get possible suggestion and print suggested words
Step 11: identify Inflection
If Found in plural list
Step 12: repeat step 8 – 11
If Not Found in both singular and plural list
If user input not spelt correctly
Step 13: initialize possible suggestion word
Step 14: print suggested words and execute
Step 15: stop

From step 1–4, through the modules of python programming language and lemmatization technique implored, the algorithm ab initio works by importing and initialize all list of English words in data structure. Then at step 4 user inputs word and the algorithm automatically repeat the previous steps against the inputted word. Step 5–14 checks the existence of the noun word in the lookup table of lemmatization approach and verify if it is in singular or plural words list. At this point also, correctness of word is verified a to ascertain inflected forms of word. If word is correctly spelt out, the system would print out the dictionary meaning and if otherwise, the system suggested possible words. The same processes would be undergone by any forms of inflected word.

5 Results and Discussion

In this research, the OWL rule ontology aspect of the proposed information retrieval system has been duly developed based on the domain. As stated earlier over 60 CQs have been formalized using FOL and as well encoded as rules with the aid of SWRL of protégé 5.5.0 version. All the rules were successfully executed with Drool engine and queried using SQWRL. For example, Fig. 4 depicted the result of querying rule2 of Table 1.

The queried rule in Fig. 4 refers to the second sample rule of Table 1 of the best fertilizer for maize. The question did not specified which of the fertilizer classes (organic or inorganic) consequently, the rule encoded in the ontology thus returns the best fertilizer for inorganic (that is, *NitrogenPhosphorusPotassium_Fertilizer* and *Urea_Fertilizer*) and organic (that is, *PoultryDropping*) from the ontology files as shown in the figure. Letters *n, u* and *a* at the output (that is, at the bottom of Fig. 4) are the variables used in the rule.

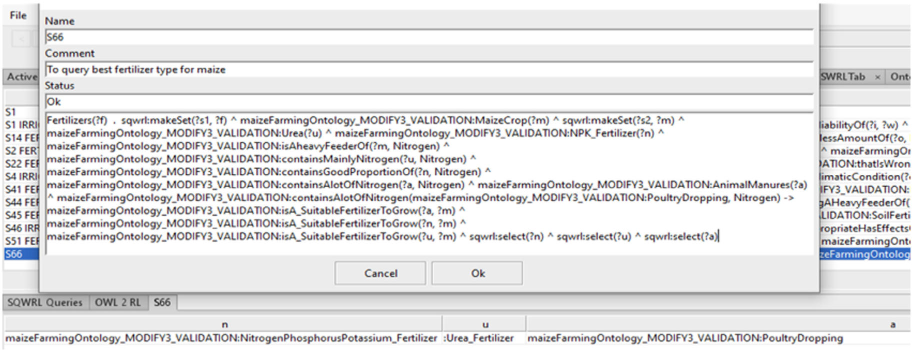

Fig. 4. The SWRL rule output

Furthermore, the proposed algorithm1 serves as a framework in which the ontology-based information retrieval system would be developed. As earlier stated, the development is already in good measures of progress before evaluation.

The results of Algorithm2's implementation in comparison with the existing wordnet are shown by Fig. 4 and Fig. 5.

Figure 4 shows the outputs of the existing and enhanced wordnets by left hand and right-hand sides respectively. The intent of the user's query is crops in inflected (plural) forms. However, while the existing wordnet returns results by working with the root word; the enhanced system works with the original inflected word and also returns appropriate results. Therefore, stemming input string to its root word as does by the existing wordnet can inadvertently affects its capability to provide a precise means of query expansion. Besides, as stated earlier; the enhanced wordnet has the strength to aid user with respect to wrong spelling of input word as shown by Fig. 5.

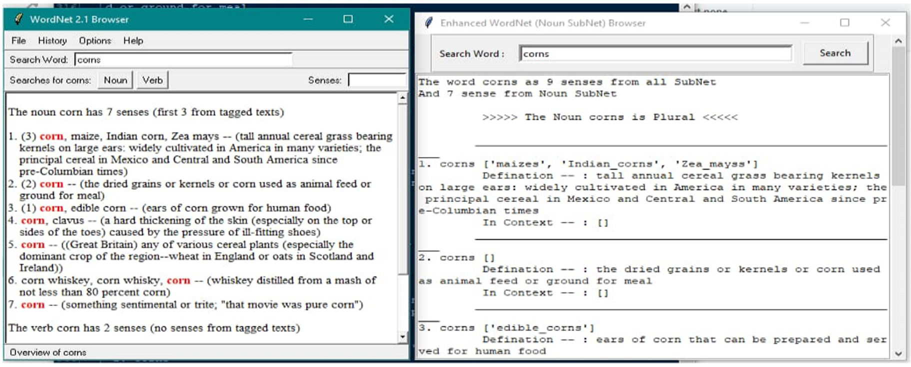

Fig. 5. Wordnets' output

Figure 5 shows a wrong spelling of fertilizer as ***fertelizer***. The existing wordnet at left hand side shows no response. But the enhanced wordnet at the right-hand side suggested some correctly spelt related words to assist user. This constantly saves time wastage and fatigue on part of user. The most interesting advantage of this innovation to wordnet especially in this research largely lies on ontology validation. Since the ontology is hand coded from scratch there is always tendency of human errors such as wrong spell of ontology's concepts. Thus, on the course of query expansion, where candidate terms are expected to tagged with similar words in the data source and variance occur as a result of wrong coded concept in ontology; the enhanced wordnet would be able to trap it and generates suggested related terms.(Fig 6)

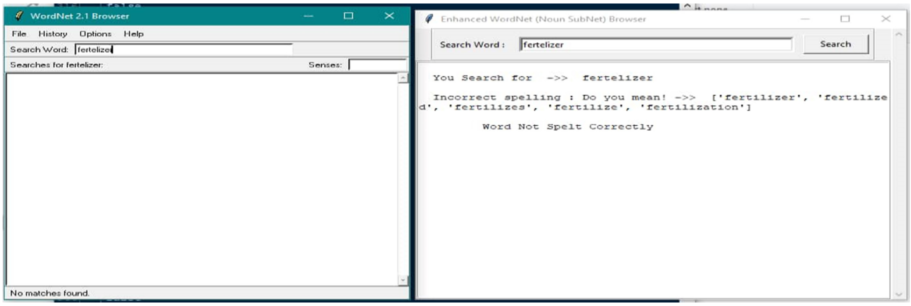

Fig. 6. Wordnets' input string spelling capability

6 Conclusion and Future Work

Query Expansion technique of Ontology-Based Information Retrieval may be designed in three forms. These are query expansion technique using independent knowledge model (WordNet); query expansion technique using domain-specific ontologies and combination of techniques. In this research, the combination of domain ontology and WordNet which is lexical database. More so, the developed OWL rule ontology considering primarily the domain of soil, fertilizer and irrigation knowledge for maize crop implored combined methodologies. The proposed ontology based IR system is premised on a framework shown by algorithm1 for effective information retrieval. The domain ontology is developed based on OWL2 using protégé 5.5.0 version. The ontology goes beyond mere classification in that, some set of verified competency questions (CQs) from domain experts were modeled using FOL, and encoded with SWRL of the editor. Hence, the named OWL Rule ontology. The ontology was validated by experts and also evaluated based on ontology's vocabulary and competencies.

More importantly in this research, the wordnet for query expansion has been improved based on the existing issue of inflected forms of words which is capable of affecting precision of input query to be expanded. The enhancement is only limited to noun subnet database, and the technique of lemmatization was implored based on algorithm2. It was implemented using python programming language taken the advantages of the third party modules of pattern.en, enchant and NLTK. Therefore, the enhanced wordnet has the capability to aid user in suggesting correctly related terms when wrong spelt word occurs. Thereby, reduces minimally user's time wastage and fatigue. This development equally aids in ontology validation. However, the AGROVOC proposed in the framework is yet to be implemented. More so, interested researchers may work on the other databases of WordNet. Besides, a technique can be devised to reduce the issue of computational costs associated with query expansion approach as rightly indicated in Sect. 4 of this research. It is also important to mention that the implementation of the framework (algorithm1) for the ontology-based information retrieval system is in good progress. Therefore, complete evaluation of the system will be carried out upon completion in due course.

References

1. Azad, H.K., Deepak, A.: Query expansion techniques for information retrieval: a survey. Inf. Process. Manage. **56**, 1698–1735 (2019)
2. Jiang, Y.: Semantically-enhanced information retrieval using multiple knowledge sources Cluster Comput. **23**, 1–20 (2020)
3. Sánchez, D., Batet, M., Isern, D., Valls, A.: Ontology-based semantic similarity: a new feature-based approach. Expert Syst. Appl. **39**, 7718–7728 (2012)
4. Thangaraj, M., Sujatha, G.: An architectural design for effective information retrieval in semantic web. Expert Syst. Appl. **41**(18), 8225–8233 (2014)
5. Sánchez, D., Isern, D., Millan, M.: Content annotation for the semantic web: an automatic web-based approach. Knowl. Inf. Syst. **27**, 393–418 (2011)
6. Shadbolt, N., Berners-Lee, T., Hall, W.: The semantic web revisited. IEEE Intell. Syst. **21**(3), 96–101 (2006)
7. Gruber, T.R.: A Translation Approach to Portable Ontology Specifications. Knowl. Acquisition **5**(2), 199–220 (1993)
8. Aminu, E. F., Oyefolahan, I. O., Abdullahi, M. B., Salaudeen, M. T.: A review on ontology development methodologies for developing ontological knowledge representation for various domains. Int. J. Inf. Eng. Electron. Bus. **2**, 28–39 (2020)
9. Yi, M.: Information organization and retrieval using a topic maps-based ontology: result of a task-based evaluation. J. Am. Soc. Inf. Sci. Technol. **15**(12), 1898 – 1911 (2008)
10. Azad, H.K., Deepak, A.: A new approach for query expansion using Wikipedia and WordNet. Inf. Sci. **492**, 147–163 (2019)
11. Bhogal, J., Macfarlane, A., Smith, P.: A review of ontology based query expansion. Inf. Process. Manage. **43**, 866–886 (2007)
12. Francesco, C., Massimo D. S., Luca, G., Paolo, N.: Weighted word pairs for query expansion. Inf. Process. Manage. **51**, 179–193 (2014)
13. Rayner, A., et al.: Ontology based query expansion for supporting information retrieval in agriculture. In: The 8th International Conference on Knowledge Management in organizations, Springer proceedings in Complexity (2014)
14. de Boer, M., Schutte, K., Kraaij, W.: Knowledge based query expansion in complex multimedia event detection. Multimedia Tools Appl. **75**(15), 9025–9043 (2015). https://doi.org/10.1007/s11042-015-2757-4
15. Fawei, B., Pan, J.Z., Kollingbaum, M., Wyner, A.Z.: A semi-automated ontology construction for legal question answering. New Gener. Comput. **37**(4), 453–478 (2019)
16. Baziz, M., Boughanem, M., Aussenac-Gilles, N.: Conceptual indexing based on document content representation. In: International conference on conceptions of library and information sciences, pp. 171–186. Springer (2005)
17. Allemang, D., Hendler, J.: Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL. Elsevier, Waltham (2011)

18. Chujai, P., Kerdprasop, N., Kerdprasop, K.: On transforming the ER model to ontology using protégé OWL tool. Int. J. Comput. Theor. Eng. **6**(6), 484 (2014)
19. Alatrish, E.S.: Comparison some of ontology. J. Manage. Inf. Syst. **8**(2), 018–024 (2013)
20. Raja, B., John, P., Chakravarthi, B.R., Arcan, M., Mccrae, J.P.: Improving wordnets for under-resourced languages using machine translation (2018)
21. Uthayan, K.R., Mala, G.S.A.: Hybrid ontology for semantic information retrieval model using keyword matching indexing system. Sci. World J. **2015** (2015)
22. Aminu, E. F., Oyefolahan, I. O., Abdullahi, M. B., Salaudeen, M. T.: Enhanced query expansion algorithm: framework for effective ontology based information retrieval system. I-Manager's J. Comput. Sci.,**6**(4), 1–11 (2019)
23. Raza M,A, Mokhtar, R, Ahmad N, Pasha M, Pasha, U.: Taxonomy and survey of semantic approaches for query expansion. IEEE Access **7,** 17823–17833 (2019)
24. Arogundade, O.T., Abayomi-Alli, A., Misra, S.: An ontology-based security risk management model for information systems. Arab. J. Sci. Eng. **45**(8), 6183–6198 (2020). https://doi.org/10.1007/s13369-020-04524-4
25. Sowunmi, O.Y., Misra, S., Omoregbe, N., Damasevicius, R., Maskeliūnas, R.: A semantic web-based framework for information retrieval in E-learning systems. In: International Conference on Recent Developments in Science, Engineering and Technology, pp. 96–106 (2017)
26. Abioye, T. E., Arogundade, O.T., Misra, S., Akinwale, A. T., Adeniran, O. J.: Toward ontology-based risk management framework for software projects: an empirical study. J. Softw. Evol. Process, **32**(12), p. e2269. (2020)