ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/338913936

ENHANCED QUERY EXPANSION ALGORITHM: FRAMEWORK FOR EFFECTIVE ONTOLOGY BASED INFORMATION RETRIEVAL SYSTEM By ENESI FEMI AMINU * ISHAQ OYEBISI OYEFOLAHAN ** MUHAMMAD BASHIR ABD….

Article · January 2020

CITATIONS
0

READS
74

4 authors:

Enesi Aminu
Federal University of Technology Minna
17 PUBLICATIONS **20** CITATIONS
SEE PROFILE

Ishaq Oyebisi Oyefolahan
Federal University of Technology Minna
37 PUBLICATIONS **41** CITATIONS
SEE PROFILE

Muhammad Bashir Abdullahi
Federal University of Technology Minna
53 PUBLICATIONS **100** CITATIONS
SEE PROFILE

Muhammadu Salaudeen
Federal University of Technology Minna
76 PUBLICATIONS **72** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Semantic Web View project

Project Screening of Rice Germplasm for African Gall midge Resistance View project

All content following this page was uploaded by Enesi Aminu on 30 January 2020.

The user has requested enhancement of the downloaded file.

ENHANCED QUERY EXPANSION ALGORITHM: FRAMEWORK FOR EFFECTIVE ONTOLOGY BASED INFORMATION RETRIEVAL SYSTEM

By

ENESI FEMI AMINU * ISHAQ OYEBISI OYEFOLAHAN ** MUHAMMAD BASHIR ABDULLAHI ***

MUHAMMADU TAJUDEEN SALAUDEEN ****

** Lecturer, Department of Computer Science, Federal University of Technology, Minna, Nigeria.*
*** Senior Lecturer, Department of Information and Media Technology, Federal University of Technology, Minna, Nigeria.*
**** Head, Department of Computer Science, Federal University of Technology, Minna, Nigeria.*
***** Senior Lecturer and Head, Department of Crop Production, Federal University of Technology, Minna, Nigeria.*

Date Received: 11/01/2019 *Date Revised: 28/01/2019* *Date Accepted: 05/04/2019*

ABSTRACT

The strength of an Information Retrieval System lies on its ability to retrieve relevant information or documents according to user's intent by considering a high level of precision and a low level of irrelevant recall of results. A recent development to actualize this dream is the application of ontology. Therefore, Ontology-Based Information Retrieval is becoming an interesting area in the current research trend of ontology and semantic web. However, the sufficiency of developing domain ontology alone to efficiently and effectively take care of information retrieval becomes a research issue. Thus, to address the research gap, a technique called Query Expansion has been identified as a veritable tool. Query Expansion is a process of expanding initial user's query term(s) with the aid of a technology such as wordNet to return relevant results according to user's intent. But returns of query results using the existing wordNet is challenging in normal or inflected terms, such as synonyms or polysemy (word mismatch). Therefore, this paper proposes improved query expansion algorithm as framework to effectively and efficiently develop ontology based information retrieval system.

Keywords: Query Expansion, WordNet, Ontology, Information Retrieval, Synonym, Polysemy.

INTRODUCTION

Currently, Information Retrieval (IR) is drawing an interesting level of research attentions in the field of Computer Science. It is a science or mechanism of retrieving relevant information based on the user's query search (Bhogal, MacFarlane, & Smith, 2007). IR on a general note has to do with representation, storage, searching, and retrieving of knowledge–based information from database. Typically, the quality of IR techniques can be measured using two metrics. They are Precision and Recall. There are various models and techniques for information retrieval processes. Basically, IR models can be classified into Boolean, vector, probabilistic and inference network model (James & Kannan, 2017). In inference model, document retrieval is modeled as an inference process in an inference network. Most techniques used by IR systems can be implemented under this model; techniques such as Query Expansion, Semantic Annotation, and Semantic Indexing. Various searching algorithms for these techniques include: TF-IDF algorithm, Brute-force algorithm, linear search, and Binary search.

Conversely, Information Retrieval Systems (IRSs) are broadly categorized into two folds: The syntactic search systems, otherwise known as keyword-based systems and semantic search systems also called conceptual-based systems (Haav & Lubi, 2001). Semantic search systems is the application of ontology in semantic information retrieval.

In recent times, ontology has been progressively used for information retrieval systems however; it has not yielded a considerable integration with searching methodologies (Chauhan, Goudar, Sharma, & Chauhan, 2013). The

applications of ontology in semantic information retrieval are diverse in nature and cannot be underestimated. That is, not limited to contextual semantic search only, but also to semantic similarity measures (Akmal, Shih, & Batres, 2014). Semantic search is introduced in the field of Information Retrieval to proffer solution to the pitfall of keyword based search (Thangaraj, & Sujatha, 2014). However, there are some advanced keyword-based query tools, such as Google, Yahoo, and Bing that can retrieve the entire documents from web, but users are still expected to perform appropriate query on the return search results before they can be considered to be appropriate (Alfred et al., 2014).

The strength of an Information Retrieval System lies on its ability to retrieve relevant information or documents according to user's intent by considering a high level of precision and a low level of irrelevant recall of results (Uthayan & Mala, 2015). However, this potential of IRS have been bedeviled by the terms; word mismatch owing to the natural languages in form of synonyms and ploysemy for instance. Therefore; it has attracted various degrees of research attentions. One of the recent developments to tackle the problem is the application of ontology. However, Bonacin, Nabuco, and Junior (2016) in their research reported that the disambiguation capability of ontological model has to be given attention.

Word mismatch is a serious challenge of Information Retrieval (Wei, Hu, Tai, Huang, & Yang, 2007). It is described as a situation of IR system, where the concept of query terms of the user is at variance with the concepts of the documents. In other words, the vocabulary issues of synonyms, polysemy, and the likes. Typically, when a user invokes query, IRS is to choose information that is likely to satisfy user's intent. In the light of this, Ranwez et al. (2013) described IRS as a function that maps a query Q (user' request) from HQ (a set of query) to a set of m documents within collection D of all indexed documents.

$$IR : H_Q \rightarrow D^M \qquad (1)$$

1. Ontology-Based Information Retrieval

Ontology-Based Information Retrieval is becoming an interesting area in the current research trend of ontology and semantic web (Alfred et al., 2014). The capability and applicability of ontology are diverse in nature, but more importantly in knowledge management and information sharing. To this end, the work of Bonacin et al. (2016) presented an insight for modeling ontology for information sharing, recovery, and interoperability amid multidisciplinary areas.

Search on the existing web is by keyword-based approach. While commendable efforts in terms of retrieving relevant documents have shown towards IRS by the researchers, there is still room for further improvement. For instance, ability to explore conceptualization lies in user's request and corpus meanings by using additional knowledge. In other words, the capacity of IRS to draw inference or relation between the query terms (Zidi & Abed, 2013). From literatures, Ontology design and its inherent technologies without developing further algorithm or applying an IR technique is considered to address the pitfalls associated with information retrieval based systems. While on the other hand, it is worth mentioning that besides ontology itself as a mechanism for relevant information retrieval; its accuracy is also determined by the efficiency of its retrieval techniques as affirmed in the work of Jain and Singh (2013).

As earlier stated, the sufficiency of ontology alone to efficiently take care of information retrieval becomes a research issue. Therefore, in an attempt to address the research gap of effective information retrieval based on ontology, Query Expansion Technique has been identified as a veritable tool.

2. Query Expansion (Q E) Technique

To retrieve a relevant hit of information based on a search term (query), additional useful terms may be added to the initial search term. Such mechanism is described as Query Expansion or Query Augmentation Technique of IRS. The scheme of adding knowledge to the existing initial query term in order to retrieve relevant information has been catching attentions of researchers in the field of study. From the literature reviewed, the mechanism of additional useful terms can be achieved manually, interactively, or automatically (Bhogal et al., 2007;

Colace, De Santo, Greco, & Napoletano, 2015).

Manual QE is achieved based on the skillful decision of user. User takes decision of which term to add to the new query. The interactive (semi-automatic) QE is often described as user-assisted QE. In semi-automatic, once the system produces additional useful terms, user decides which of the term to add to the initial query. For Automatic QE, weights are computed for each of the additional terms and the term with the highest value will be added to the initial query autonomously by the system. There are different weighting functions that give different results. Thus, it is important to note that the efficiency of retrieval system in this scheme depends on how the weighting values have been computed.

Essentially, with query expansion, the initial query is augmented with an additional meaningful terms and the terms are usually a derivative of knowledge collections or the returned documents of the initial query. There are two classes of knowledge collections, they are: knowledge independent collections and knowledge dependent collections (Alfred et al., 2014). Query expansion technique can be effectively used both in common knowledge bases (wikipedia) and expert-design knowledge bases such as ontologies (de Boer, Schutte, & Kraaij, 2016). Thus, the new search term produced from the mechanisms described is regarded as contextual information for the initial search term hoping to enhance the retrieval results. The contextual information may be obtained from the term co-occurrence, relevance feedback and the latest which can be deduced from knowledge models for instance, ontologies.

The classification of this technique varies in literatures. Different authors presented different categories however; they are still related, but with some level of ambiguity. For instance, while Alfred et al. (2014) categorized the technique into three major categories, (Sodanil, Phonarin, & Porrawatpreyakorn, 2013; Xu, & Croft, 2000; Wu, Ilyas, & Weddell, 2011) categorized it into two approaches differently and respectively. However, in this review paper the categorization process of the technique had been revised Oyefolahan, Aminu, Abdullahi, & Salaudeen, 2018). That is, the authors classified QE as global and local techniques. The former is intended to achieve when wordNet or/and domain ontology (ies) is used to expand an initial query. The review contained the detail and comprehensive classification processes of the technique. Table 1 summarily presented the categorization process of the technique.

3. Related Studies: Ontology-Based Query Expansion Technique of Information Retrieval

Over a period of time, query expansion technique of information retrieval has been continuously reviewed and applied under various expansion mechanisms. This has evidently yielded retrieval results. However, since web of knowledge is currently trending and issue of word mismatch poses a challenge, enhanced knowledge-based query expansion techniques are required (Wu et al., 2011).

Query expansion algorithm in sports domain for semantic information retrieval was proposed in (Devi & Gandhi, 2014). The proposed SIRSD algorithm works by using WordNet and domain ontology to enhance the returned results of query. The research depicted the significance of query expansion using ontology (wordNet) over query expansion using relevance feedback. However, WordNet, a lexical ontology to an extent provided solutions to the

Query Expansion Techniques	Types
Alfred et al. (2014)	Manual QE Semi-automatic QE Automatic QE
Sodanil, Phonarin, and Porrawatpreyakorn (2013, December)	Local Analysis (Relevance feedback, local feedback) Global Analysis (Ontology, Thesaurus)
Xu and Croft (2000)	Automatic Query Expansion (local and global) Manual Query Expansion (Relevance, Manual Thesaurus)
Wu, Ilyas, and Weddell (2011)	Syntactical (Using Statistic analysis) Semantic (Ontology, Thesaurus)
Oyefolahan, Aminu, Abdullahi, and Salaudeen (2018)	Local Approach (Relevance feedback, Pseudo-relevance feedback) Global Approach (Ontologies- WordNet or Domain specific). Any of the approach can be manually, interactively or automatically achieved.

Table 1. Categorization Process of Query Expansion Technique

problem of word mismatch, but not without limitations such as polysemy issues, word-inflected form of WordNet, and ambiguous synset of WordNet for various domains. Vijayarajan, Dinakaran, Tejaswin, and Lohani (2016) proposed an ontology based Object- Attribute-Value (O-A-V) generic framework for information retrieval. It provides meaningful insight into the content of the documents, but related terms of those search words are not taken into consideration, thereby results in low precision and recalls of results. Similar weakness was found in the work of Tulasi, Rao, Ankita, & Hgoudar, 2017), who proposed a semantic based system with automatic semantic annotation for efficient retrieval of documents in the domain of sports.

An ontology based search system as a knowledge search and retrieval system developed for a knowledge area has the capacity to address the limitations of the existing keyword based methods. Ma and Tian (2015) developed ontology for mechanical domain in RDF scripts with aid of Protégé 4.2 and adopted query expansion technique for its knowledge retrieval. It was reported that the proposed method outweighs the performance of keyword based search methods in terms of precision and recall metrics. Similar approach was used by (Zhang, Hou, Chen, & Zhuang, 2013; Xinhua & Xutang, 2012) for engineering domain and technique preparation domain ontologies, respectively. An ontology based information retrieval was presented in the domain of soccer Kara et al. (2012). In their research work, three semantic search issues were taken into consideration. Among which is the retrieval's performance issue.

The work of Alfred et al. (2014) maintains that the retrieval mechanisms of an ontology knowledge based system are drawing research attentions presently. Ontology based Query Expansion technique for the domain of agriculture was the focus of their research work. The authors stated that the format of agricultural information on the web poses a lot of challenges in terms of knowledge base redundancy and also complexity to manage the different natures of relationship that exist among the concepts of the knowledge base. Therefore, it is concluded that modeling agricultural information ontologically is capable to resolve the research problem.

The research work of Sodanil et al. (2013) focused on query expansion using global analysis to improve the performance of information retrieval of agricultural domain ontology using simple query terms and association rule mining based method for inference. The ontology is based on agricultural expertise retrieval framework (ARGIX) of Phonarin et al. (2012). It is constructed using Protégé and generated the agricultural OWL files. The research therefore recommended that more agricultural vocabularies have to be considered in order to expand the ontology.

The use of Relevance feedback and Pseudo relevance feedback techniques of query expansion based on the information from news domain ontology in a probabilistic retrieval system were examined in the work of Bhogal and Macfarlane (2013). The main goal of the research is to evaluate the effect of ontology in terms of recall and precision metrics. The ontology written in XML was developed by Kallipolitis, Karpis, and Karali (2007). The authors concluded that the results of pseudo-relevance feedback outweigh the performance of the relevance feedback. TREC document collection was selected for evaluation owing to the chosen domain. The authors therefore recommended that in order to improve intelligence, the system should include concept's synonyms into consideration which was not included in their system.

An easy access to research information of any domain or field of study remains fundamental for qualitative research development. However, the intent of user's search cannot be semantically interpreted by the conventional search engines. To this end, Thunkijjanukij, Kawtrakul, Panichsakpatana, and Veesommai (2009) proposes query expansion-based information retieval methodology for plant production ontology specifically, rice domain. OWL and AGROVOC CS WB were used as ontology representation language and ontology editor, respectively. Thus, the IR technique along with the reasoning components of the ontology were reported to greatly enhance the performance of the ontology based IR. The ontology was evaluated by domain experts and users following how satisfactory the ontology to the

competency questions. The query efficiency was measured by precision and recall.

An architectural framework of Semantic Based Information Retrieval System (SBIRS) for a semantic search was designed and implemented by Thangaraj and Sujatha (2014). Semantic annotation and indexing was designed for a component of the SBIRS called semantic indexer; where weights of terms were computed by an adapting TF-IDF algorithm. However, it was noted that all web pages could not use the same ontology even for the same domain. The system was implemented using C#.net as web-based system in Visual Studio 2010. In order to test the effectiveness of the system, the proposed system, SBIRS was compared with a keyword based approach, KBIRS in terms of precision and recall.

A prototype tool that enables semi-automatic annotation of document for publishing on the web with the aid of RDFa was presented by Fontes, Cavalcanti, and Moura (2013). The authors explored the ontology inference capability for semantic annotation and on meta-annotation concepts. The system was developed in Java 1.6 using the APIs such as OWLAPI, Pellet-Reasoner. TREC based approach was implored to carried for evaluation. The evaluation was done with and without the reasoner by taken precision and recall metrics into consideration. However, concepts of the domain (wine) ontology as terms that are present in the web document are exact match therefore, recall may likely be affected. Synonyms of those terms have to be taken into consideration.

A semantic annotation information retrieval system based on corn plant ontology was proposed in the research of Qi, Zhang, and Gao (2010). The semantic annotation technique was RDF triple-based of Formal Concept Analysis (FCA) approach. The approach equally implored TF-IDF algorithm to compute the term weight. Terms are chosen as core domain concepts if only their weight values are more than the threshold value.

The research of Li, Raskin, and Ramani (2007) described an ontology development methodology that would be incorporated with protégé and promised that the proposed method would have the capacity to achieve efficient ontology based information retrieval. Similarly, cash crop ontology developed by Ekuobase and Ebietomere (2016) is reported to provide precise cash crop farmers market's information in Nigeria.

While Ruban, Tendolkar, Rodrigues, and Shetty (2014) argued that the keyword based information retrieval does not satisfy user's query intent in terms of recall and precision. Therefore, in order to enhance the relevance of results return to user, an information retrieval model based on domesticated plants ontology is developed. The ontology was used to produce an information retrieval system that aid context based search consequently, changing the keyword based search system. Also, Pokharel, Sherif, and Lehmann (2014) described how agricultural information in Nepal became accessible by converting different available agricultural data which are not readily and easily accessible before through the use of Resource Description Framework (RDF). RDF is a triple based ontology representation language. Nonetheless, the authors hoped in future to integrate some domain related concepts such as trade and transportation into the ontology. More so, non experts cannot use the system.

Similarly, some researches consider the approach of ontology mapping to address the problem of information retrieval. For examples; Jimeno-Yepes, Berlanga-Llavori, and Rebholz-Schuhmann (2010) proposed ontology refinement algorithm to refine ontologies for information retrieval, by mapping some popular biomedical ontologies (MeSH and Gene). Uthayan and Mala (2015) proposed hybrid fuzzy ontology whereby concepts which are similar from two ontologies of a domain are retrieved and aligned. The authors stated that fusion of ontology into accessing information and retrieval is an effective approach to implement searching effects of appropriate information that users want.

Therefore, ontology-based query expansion technique of information retrieval can be viewed, analyzed, or designed in three forms depending on the goals of a research, namely: query expansion technique using wordNet ontology, query expansion technique using domain-specific ontologies, and combination of WordNet and domain-specific ontologies. The first two

forms of QE design have their own potentials and limitations as discussed by the subsequent sections.

3.1 Query Expansion Technique using WordNet

WordNet is described as knowledge independent collection or a corpus independent knowledge model. That is, a collection of documents that is not restricted to particular domain of knowledge. Generally, WordNet is a lexical data repository that is ontologically designed for linguistic purpose, specifically for parts of speech. It has three databases. These are noun, verb, adjective and adverb. The set of synonyms denoted as synset makes available various semantic relations, such as synonyms, antonymy, hyponymy, hypernymy, holonymy, and meronymy (Uthayan & Mala, 2015). Despite its limitations, WordNet has played significant role in information retrieval techniques from the literatures reviewed (Lu, Sun, Wang, Lo, & Duan, (2015).

Besides web search, it has been tested on patent search tasks (Magdy & Jones, 2011); all these for query expansion. Beyond query expansion, it has the capacity to disambiguate the sense of query words (Pal, Mitra, & Datta, 2014) all to achieve efficient information retrieval. In contrast, the limitations of this technique includes most of the semantic associations between two terms that are not found in the WordNet and since WordNet has wide coverage, ambiguous terms within the ontology can pose a challenge (Bhogal et al., 2007).

3.2 Query Expansion Technique using Domain-Specific Ontologies

In recent times, ontology has been progressively used for developing applications for a domain and information retrieval systems. As earlier stated, ontology may be developed or reused in any field of knowledge or domain. So, when ontology is used as a knowledge model to cushion the effect of retrieving relevant information in a particular domain to form conceptual information based on the initial query, such phenomenon is described as query expansion using domain-specific ontology. In the literatures of (Alfred et al., 2014; Sodanil et al., 2013; Liao, Li, & Liu, 2015), query expansion was implored for agriculture domain ontologies. Similarly, (Chauhan et al., 2013; Devi & Gandhi, 2014; Kara et al., 2012; Tulasi et al., 2017) for sports based ontology.

From literature, query expansion algorithm or technique is mostly developed and achieved using WordNet. On the other hand, the WordNet is limited in terms of ambiguous set of synonyms (polysemy issues) and inability to distinguish the results of word-inflected form of search term from normal-word form. Therefore, this paper aimed to propose an algorithmic framework to address the limitations so as to enhance WordNet for query expansion approach. And at the end, effective ontology based information retrieval system is realized.

4. Proposed Architecture of Ontology based Information Retrieval System

For efficient information retrieval that devoid of word mismatch problem as earlier stated, this section proposed a framework that enhanced query expansion algorithms for design and implementation of ontology based information retrieval system. Therefore, the proposed system as represented by Figure 1 comprises of four layers. They are as follows.

4.1 Query Search Layer

The layer consists of user interface, query processor, and candidate terms. A user interface will be created where user is expected to input query search. Since the query search is inputted in natural language, it has to be preprocessed to remove stop words using Stanford parser. Thereafter, candidate terms or keywords will be extracted based on the ontology design using SPARQL to carry out the query. If the candidate terms are successful in the ontology, it will proceed to the next layer. Otherwise, it will return a message such as not a domain term.

4.2 Semantic Query Expansion Layer

To ensure better relevant results in terms of precision and recall, this layer proposed to use an enhanced WordNet to capture the synonyms and equally design an algorithm to handle the polysemous nature of the candidate terms selected. The terms selected are results of the query on the ontology with the aid of SPARQL. An improved WordNet becomes necessary in order to solve word-inflected form of WordNet. A pseudocode Algorithm has represented the

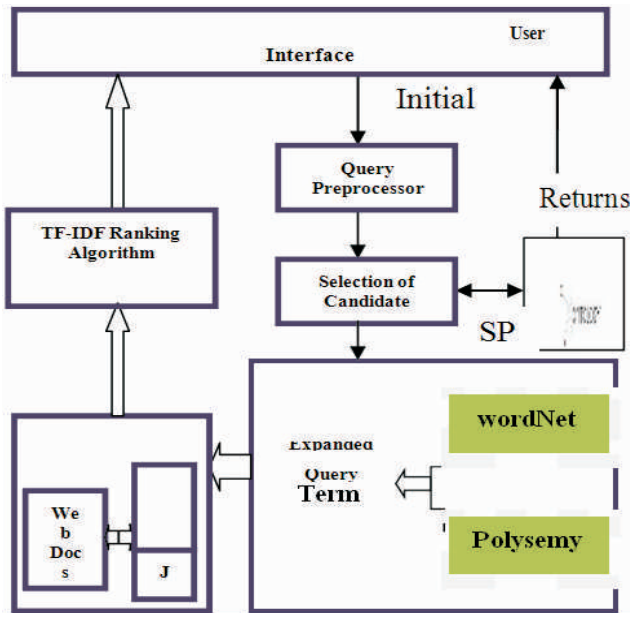

Figure 1. Conceptual Framework of the Proposed System

proposed solution.

This layer is described as the core layer of the proposed system. This is the layer that handles the expansion of the initial query search in order to retrieve relevant hits.

4.3 Semantic Search Layer

This layer is referred to as Knowledge Base because query search terms at this point are expanded. That is, the candidate terms of the initial query terms have been argumented with their synonyms using an enhanced WordNet and the polysemy algorithm to ascertain that they are contextual based terms. Based on this development, the return results such as web documents will be retrieved based on the semantic of the expanded query terms that matched with the ontology's concepts of the proposed domain with the aid of JENA as semantic reasoner.

Algorithm1: Word-inflected form of WordNet Input: Normal noun-terms and inflected noun-terms Output: Returns the appropriate synsets for term

Step 1: Connect to the WordNet noun database

Step 2: Read input string from the search text field

Step 3: If root (normal) term inputted Then GOTO Step 4 Else GOTO Step6

Step 4: Retrieve all synset or senses of the root term

Step 5: Set count equals to zero and pass it to Step 5

Step 6: Declare a map of string words to string meanings (String to String)

Step 7: Check what kind of inflection occurred and retrieve corresponding inflexion action

Step 8: Retrieve an appropriate synset or senses of the inflected term

Step 9: GOTO Step 5

Similarly, Algorithm 2 represented the proposed Polysemy algorithm for query search.

Algorithm 2: Polysemy Word Lookup

Input: Candidate term

Output: Context Based term

Step 1: For Each candidate term from user search query

Step 2: Input terms

Step 3: Related terms/documents returns

Step 4: If key terms returned are found as ontology's concepts

Step 5: Then Do Step 7

Step 6: Else GOTO Step2

Step 7: Pass it to Step 5

Finally, Algorithm 3 presented a pseudocode of how the query expansion will work.

Algorithm 3:

Input: User Search Query

Output: Search Query

Expanded Method: Query Expansion

Step 1: Develop ontology for a domain

Step 2: End user input search query

Step 3: Split the user search query into individual term

Step 4: Candidate term selection

Step 5: For-each candidate term, simultaneously find the synonyms and polysemy using an enhanced WordNet and polysemy algorithm, respectively

Step 6: The initial search query and the related words gotten from Step 5 form an expanded search query

Step 7: Return the new expanded search query for SPARQL.

4.4 Results Retrieval Layer

In order to ensure efficient results in terms of low recall high precision, the return results based on the contextual information search of the previous layer need to be further optimized by adopting a dynamic ranking algorithm known as Term Frequency – Inverse Document Frequency (TF-IDF) to compute weights for the results and finally retrieve relevant results as to the intent of user.

Discussion and Conclusion

Algorithm 1 is designed to improve the operability of the existing WordNet considering the importance of the lexical database (WordNet) in the field of Information Retrieval System. The improvement that is, Alogrithm 1 focuses on making sure that appropriate synonyms distinctly returns for root noun terms (normal) and extended terms (inflected noun terms). It is important to mention that three types of databases exist in WordNet that is; noun, verb, and adjective/adverb databases, but this research paper focuses on Noun database.

Similarly, Algorithm 2 is also designed to cater for polysemous terms. In order to have a high recall of relevant results, other related or similar terms in meaning of search terms have to be considered. For example, a user searching for the term corn can still acquire relevant information from related term maize. Since the information retrieval model is ontology based, domain ontology will be used to serve as regulator for the related terms.

When all the necessary synonyms or polysemy of search terms have been performed; the results form the new search query which we refer to as the enhanced query expansion in this paper as represented by Algorithm 3. The expanded new query would be handled by a query language called SPARQL.

As a proof of concepts, when the proposed algorithm framework is implemented and applied, it will assist the existing WordNet to return the synsets (set of synonyms) for normal noun terms and inflected noun terms appropriately. Normal noun terms are the root lexical noun terms (they normally exist in a singular form) such as Chair, Box, Knife, Child and Fishery. On the other hand, Inflected Noun terms are lexical noun terms that normally exist in plural forms or padded with some letters as prefix or suffix. Examples are Chairs, Boxes, Knives, Children, and Fisheries. The algorithm becomes necessary because the existing WordNet returns exactly same synsets for both noun terms.

In order to enhance accuracy of query expansion technique, WordNet is used. In Algorithm 1, search term(s) is inputted and if normal terms; appropriate synset are retrieved and candidate terms are selected in Algorithm 3. Also, this algorithm will trigger Algorithm 1 (an enhanced WordNet) to find the synonyms and Algorithm 2 to find the appropriate polysemy using a given domain ontology as lookup word of the candidate terms simultaneously. Else that is, if the inputted terms are inflected, step 7 of Algorithm 1 will be invoked to check the type of inflection and perform the corresponding actions.

In conclusion, the proposed framework for an effective ontology based information retrieval using query expansion approach can be adopted under any domain ontology to develop an IR system. However, it is still a research in progress. The authors aimed to adopt it to develop ontology based information retrieval system that will be implemented on maize farming domain ontology. Possibilities of some minor changes to the proposed framework may likely come up with no effects to its operability. For instance, on the course of implementation; OWL QL which is also a very good query language may be used instead of SPARQL, especially when semantic expressivity of ontology is the watch word. SPARQL is a query language for RDF ontology representation language. However, it can still be used for OWL based ontology. On the part of evaluation metrics, precision and recall will be used during implementation by considering user/domain expert and comparison with existing system or TREC evaluation technique(s).

References

[1]. Akmal, S., Shih, L. H., & Batres, R. (2014). Ontology-based similarity for product information retrieval. *Computers in Industry*, 65(1), 91-107.

[2]. Alfred, R., Chin, K. O., Anthony, P., San, P. W., Im, T. L.,

Leong, L. C., & Soon, G. K. (2014). Ontology-based query expansion for supporting information retrieval in agriculture. In *The 8th International Conference on Knowledge Management in Organizations* (pp. 299-311). Springer.

[3]. Bhogal, J., MacFarlane, A., & Smith, P. (2007). A review of ontology based query expansion. *Information Processing & Management*, 43(4), 866-886.

[4]. Bhogal, J., & Macfarlane, A. (2013, October). Ontology based query expansion with a probabilistic retrieval model. In *Information Retrieval Facility Conference* (pp. 5-16). Springer.

[5]. Bonacin, R., Nabuco, O. F., & Junior, I. P. (2016). Ontology models of the impacts of agriculture and climate changes on water resources: Scenarios on interoperability and information recovery. *Future Generation Computer Systems*, 54, 423-434.

[6]. Chauhan, R., Goudar, R., Sharma, R., & Chauhan, A. (2013, March). Domain ontology based semantic search for efficient information retrieval through automatic query expansion. In *Intelligent Systems and Signal Processing (ISSP), 2013 International Conference on* (pp. 397-402). IEEE.

[7]. Colace, F., De Santo, M., Greco, L., & Napoletano, P. (2015). Weighted word pairs for query expansion. *Information Processing & Management*, 51(1), 179-193.

[8]. de Boer, M., Schutte, K., & Kraaij, W. (2016). Knowledge based query expansion in complex multimedia event detection. *Multimedia Tools and Application*, 75(15),9025-9043.

[9]. Devi, M. U., & Gandhi, G. M. (2014). WordNet and ontology based query expansion for semantic information retrieval in Sports domain. *Journal on Computer Science.*

[10]. Ekuobase, G. O., & Ebietomere, E. P. (2016, May). Ontology for alleviating poverty among farmers in Nigeria. In *Proceedings of the 10th International Conference on Informatics and Systems* (pp. 28-34). ACM.

[11]. Fontes, C. A., Cavalcanti, M. C., & Moura, A. M. D. C. (2013, September). An ontology-based reasoning approach for document annotation. In *Semantic Computing (ICSC), 2013 IEEE Seventh International Conference on* (pp. 160-167). IEEE.

[12]. Haav, H. M., & Lubi, T. L. (2001, September). A survey of concept-based information retrieval tools on the web. In *Proceedings of the 5th East-European Conference ADBIS* (Vol. 2, pp. 29-41).

[13]. Jain, V., & Singh, M. (2013). Ontology based information retrieval in semantic web: A survey. *International Journal of Information Technology and Computer Science (IJITCS)*, 5(10), 62.

[14]. James, N. T., & Kannan, R. (2017). A survey on information retrieval models, techniques and applications. *International Journals of Advanced Research in Computer Science and Software Engineering*, 7(7), 16-19.

[15]. Jimeno-Yepes, A., Berlanga-Llavori, R., & Rebholz-Schuhmann, D. (2010). Ontology refinement for improved information retrieval. *Information Processing & Management*, 46(4), 426-435.

[16]. Kallipolitis, L., Karpis, V., & Karali, I. (2007, February). World news finder: How we cope without the semantic web. In *Proceedings of the 25th Conference on IASTED International Multi-Conference: Artificial Intelligence and Applications* (pp. 616-621). ACTA Press.

[17]. Kara, S., Alan, Ö., Sabuncu, O., Akpınar, S., Cicekli, N. K., & Alpaslan, F. N. (2012). An ontology-based retrieval system using semantic indexing. *Information Systems*, 37(4), 294-305.

[18]. Liao, J., Li, L., & Liu, X. (2015). An integrated, ontology-based agricultural information system. *Information Development*, 31(2), 150-163.

[19]. Li, Z., Raskin, V., & Ramani, K. (2007). A methodology of engineering ontology development for information retrieval. In *Proceedings of the 16th International Conference on Engineering Design (ICED'07)*.

[20]. Lu, M., Sun, X., Wang, S., Lo, D., & Duan, Y. (2015, March). Query expansion via WordNet for effective code search. In *Software Analysis, Evolution and Reengineering (SANER), 2015 IEEE 22nd International Conference on* (pp. 545-549). IEEE.

[21]. Ma, S., & Tian, L. (2015). Ontology-based semantic retrieval for mechanical design knowledge. *International Journal of Computer Integrated Manufacturing*, 28(2), 226-238.

[22]. Magdy, W., & Jones, G. J. (2011, October). A study on query expansion methods for patent retrieval. In *Proceedings of the 4th Workshop on Patent Information Retrieval* (pp. 19-24). ACM.

[23]. Oyefolahan, I. O., Aminu, E. F., Abdullahi, M. B., & Salaudeen, M. T.. A. (2018). Review of ontology-based information retrieval techniques on generic domains. *International Journal of Applied Information Systems (IJAIS)*, 12(13), 8-21.

[24]. Pal, D., Mitra, M., & Datta, K. (2014). Improving query expansion using WordNet. *Journal of the Association for Information Science and Technology*, 65(12), 2469-2478.

[25]. Phonarin, P., Nitsuwat, S., & Haruechaiyasak, C. (2012). AGRIX: An ontology based agricultural expertise retrieval framework. In *Advanced Materials Research* (Vol. 403, pp. 3714-3718). Trans Tech Publications.

[26]. Pokharel, S., Sherif, M. A., & Lehmann, J. (2014, August). Ontology based data access and integration for improving the effectiveness of farming in Nepal. In *Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)* (Vol. 2, pp. 319-326). IEEE Computer Society.

[27]. Qi, H., Zhang, L., & Gao, Y. (2010, August). Semantic retrieval system based on corn ontology. In *Frontier of Computer Science and Technology (FCST), 2010 Fifth International Conference on* (pp. 116-121). IEEE.

[28]. Ranwez, S., Duthil, B., Sy, M. F., Montmain, J., Augereau, P., & Ranwez, V. (2013). How ontology based information retrieval systems may benefit from lexical text analysis. In *New Trends of Research in Ontologies and Lexical Resources* (pp. 209-231). Springer, Berlin, Heidelberg.

[29]. Ruban, S, Tendolkar. K, Rodrigues, A. P, & Shetty, N, (2014). An ontology-based information retrieval model for domesticated plants. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(5).

[30]. Sodanil, M., Phonarin, P., & Porrawatpreyakorn, N. (2013, December). An ontology-based query expansion for an agricultural expert retrieval system. In *Proceedings of International Conference on Information Integration and Web-based Applications & Services* (p. 358). ACM.

[31]. Thangaraj, M., & Sujatha, G. (2014). An architectural design for effective information retrieval in semantic web. *Expert Systems with Applications*, 41(18), 8225-8233.

[32]. Thunkijjanukij, A., Kawtrakul, A., Panichsakpatana, S., & Veesommai, U. (2009). Ontology development: A case study for Thai rice. *Kasetsart J. (Nat. Sci.)*, 43(3), 594-604.

[33]. Tulasi, R. L., Rao, M. S., Ankita, K., & Hgoudar, R. (2017). Ontology-based automatic annotation: an approach for efficient retrieval of semantic results of web documents. In *Proceedings of the First International Conference on Computational Intelligence and Informatics* (pp. 331-339). Springer.

[34]. Uthayan, K. R., & Mala, G. S. A. (2015). Hybrid ontology for semantic information retrieval model using keyword matching indexing system. *The Scientific World Journal*, 2015.

[35]. Vijayarajan, V., Dinakaran, M., Tejaswin, P., & Lohani, M. (2016). A generic framework for ontology-based information retrieval and image retrieval in web data. *Human-centric Computing and Information Sciences*, 6(1), 18.

[36]. Wei, C. P., Hu, P. J. H., Tai, C. H., Huang, C. N., & Yang, C. S. (2007). Managing word mismatch problems in information retrieval: A topic-based query expansion approach. *Journal of Management Information Systems*, 24(3), 269-295.

[37]. Wu, J., Ilyas, I., & Weddell, G. (2011). *A study of ontology-based query expansion*. Technical Report CS-2011-04.

[38]. Xinhua, L., & Xutang, Z. (2012). A domain ontology-based Information retrieval approach for technique preparation. *Physics Procedia*, 25, 1582-1588.

[39]. Xu, J., & Croft, W. B. (2000). Improving the

effectiveness of information retrieval with local context analysis. *ACM Transactions on Information Systems (TOIS)*, 18(1), 79-112.

[40]. Zhang, X., Hou, X., Chen, X., & Zhuang, T. (2013). Ontology-based semantic retrieval for engineering domain knowledge. *Neurocomputing*, 116, 382-391.

[41]. Zidi, A., & Abed, M. (2013, May). A generalized framework for ontology-based information retrieval: Application to a public-transportation system. In *Advanced Logistics and Transport (ICALT), 2013 International Conference on* (pp. 165-169). IEEE.

ABOUT THE AUTHORS

Enesi Femi Aminu presently Lectures in the Department of Computer Science at Federal University of Technology, Minna, Nigeria and also a Ph.D student in the same Department. He teaches courses including Operating Systems, Database Design and Management, Object Oriented Programming, and Web Design. He obtained both his B.Sc and M.Sc Degrees in Computer Science from University of Jos, Jos and Ahmadu Bello University, Zaria, respectively. His current research interest is on Ontology Design and Semantic Search. He is also a member of professional bodies: Nigeria Computer Society (NCS) and International Association of Computer Science and Information Technology (IACSIT).

Dr. Ishaq Oyebisi Oyefolahan is currently a Senior Lecturer in the Department of Information and Media Technology at Federal University of Technology, Minna, Nigeria. Prior to joining his current department, he was an Assistant Professor in the Department of Information Systems, International Islamic University Malaysia. His fields of interest are Business Intelligence, Web and Mobile Applications Development, Utilization and Evaluation, Business-IT Alignment, and Knowledge Management. He has published several technical papers in International, National Journals, and Conferences.

Dr. Muhammad Bashir Abdullahi is currently the Head of the Department of Computer Science at Federal University of Technology, Minna-Nigeria. He received B.Tech (Honors) in Mathematics/Computer Science from Federal University of Technology, Minna-Nigeria, and PhD Degree in Computer Science and Technology from Central South University, China. His research interests are mainly in the areas of Network and Information Security, Internet of Things, Machine Learning, Big Data Technology, and Cloud Computing.

Dr. Muhammadu Tajudeen Salaudeen is currently a Senior Lecturer and Head in the Department of Crop Production at Federal University of Technology, Minna, Nigeria. He holds a PhD in Plant Pathology from the University of Ibadan, M.Sc Crop Protection from Ahmadu Bello University, Zaria, and B. (Agric) Tech. General Agriculture (First Class Honours) from the Federal University of Technology, Minna. He was a Research Fellow at the International Institute of Tropical Agriculture (IITA), Ibadan between 2009 and 2012 and a Visiting Scientist from 2013 – 2014 at the same Institute. He has received specialist training on Molecular Breeding and Analyses of Agricultural Research Data in the Netherlands (2012), Spain (2013 and 2014) and IITA, Ibadan. He won an International research grant in 2013 to investigate the status of Maize Lethal Necrosis Disease (MLND) in Nigeria. He conducts Research on Survey of Plant Viruses, Plant Virus Diversity, Virus Host Range, and Screening for Resistance. He teaches Biostatistics, Crop Protection, and Plant Pathology courses for undergraduate and postgraduate levels.

View publication stats