CHAPTER 5

SEMANTIC SEARCH IN QAAL BASED ON
GRAPH MATCHING AND SEMANTIC
INDEXING WITH INFERENCE

5.1 INTRODUCTION

In answer extraction module, the correct answer is retrieved from knowledge base. Indexing is a technique used to match the terms in query with the terms in concepts of ontology which is used to improve the answer extraction pattern. Two types of indexing mechanisms are available namely as syntactic indexing and semantic indexing. Syntactic indexing is based on keyword based indexing mechanism. In semantic indexing, two approaches are used which are stemming and decompounding. Stemming is the process whereby a word is reduced to its stem form based on rules. Stemming works well for languages which have derivation and modulation. In stemming, suffix stripping is available. Stemming is used to increase the recall and not loss in precision. Decompounding splits up words which consist of other words that are glued together. Decompounding is used for languages where compounds are productive.

Semantic indexing is one of the types of indexing scheme which is applied in knowledge base. It is also a weighted annotation model in query processing. Annotation weights are computed by using an
adaptation of Term Frequency Inverse Document Frequency (TF-IDF) algorithm. The TF-IDF calculates value for each word in a sentence through an inverse proportion of the frequency of the word in sentence. Words with high TF-IDF imply a strong relationship with the sentence that they appear. Annotation can be done manually or automatically. Annotation is done by manually in this proposed QAAL system. Entities represented in ontology are taken as semantic entities. In this proposed QAAL system, a query vector and document vector are calculated.

After query processing stage, QAAL system returns RDF graph. The user does not understand this RDF graph format. So, there is a need of displaying answer in single sentence form by using the answer extraction module. In this chapter, the benefits of semantic indexing mechanism in answer extraction model are analyzed and the integration of semantic indexing with semantic inference mechanism is proposed to improve the performance of QAAL system. The evaluation is done based on precision, recall, F-Measure and MRR ratio. The input of answer extraction module is a RDF graph which is converted into semantic tree model for matching in knowledge base. After finding the answer from the knowledge base, the user is in need to verify the retrieved answer is correct or not. So, confidence ratio metric is proposed for measuring the correct answer based on semantic search with integration of semantic indexing and inference mechanism.

The organization of the chapter 5 is as follows: Semantic indexing and semantic indexing in QAAL system are explained in Section 5.2 and Section 5.3. In Section 5.4 and Section 5.5, RDF graph matching with synonym factor and RDF graph matching with inferences using semantic indexing are discussed. In Section 5.6, the performance
evaluation and discussion is based on synonym factor, semantic indexing, semantic search with semantic indexing and confidence ratio. Finally concluding remarks as summary is made in Section 5.7.

5.2 SEMANTIC INDEXING

In ontology, semantic IR is possible by setting the concepts and instances in KB as annotations. Semantic indexing is an index type which can make use of information extraction and annotation to index terms of ontology is in semantic manner. In this proposed QAAL system a semi-automatic annotation mechanism is implemented in DS ontology. These annotations are later used for retrieval and ranking process.

A scalable alternative way to answer extraction is semantic indexing. In this approach, semantic data in RDF knowledgebase are indexed in a way and made directly available to use queries for retrieving answer. In this proposed QAAL system, inference mechanism is taken as offline for reduction in time. This improves the scalability of the system and also the efficiency of query. It is proposed to design a QAAL system with semantic indexing with inference which extends the extracted and inferred information by using domain ontology.

IR vector space model is described by Salton et al (2007). In this model, keywords in question were assigned as weight with the fact of
Where \( \text{freq}_{i,Q} \) is the number of occurrences in \( Q \) with the keywords attached to the instance \( i \). \( \text{maxfreq}_{i,Q} \) is the frequency of the most repeated instance in \( Q \). \( n^Q \) is the number of words annotated with \( Q \) and \( D \) is the set of all words in question in the search space. \(|D|\) is the number of set of all words in question.

### 5.3 SEMANTIC INDEXING IN QAAL SYSTEM

The proposed QAAL system computes semantic similarity between the query and the resource by using adaptation of class vector space IR model. The query vector is generated with the variables in SELECT clause of SPARQL query and is set to 1. Users are allowed to manually set this weight according to the interest. If the vector is constructed, the similarity measurement between the Query \( qr \) and the Resource \( r \) will be calculated. \( \text{sim}(qr,r) \) is termed as the semantic similarity between question and resource which is computed as

\[
\text{sim}(qr,r) = \frac{qr \cdot r}{\sqrt{|qr| \cdot |r|}}
\]
which \( k_{\text{sim}}(qr,r) \) is computed by semantic similarity value between query and resource and the value of \( \lambda \) is calculates as a range from 0 to 1. In this proposed work, the similarity value (\( \lambda \)) is set as a threshold value 0.5 to perform the experiments. The following parameters are used to generate semantic index by using Random Indexing (RI) method. The parameters are seed length which has the value may as set as 1 or -1 in sparse random vector. Another parameter is dimension of semantic vector space which has a predefined number of dimensions to use sparse random vectors. The last parameter is the minimum term frequency which is a term frequency to get in index.

Once the semantic indexing is created, it can be used to find similarities between resource and question with RDF graphs. In this proposed work, a cosine function is used to calculate the similarity between the input term vector and the existing vector which is generated by vector space model. Damljanovic et al (2011) have discussed this type of random indexing scheme based on vector space model for finding nodes in large RDF graphs.

Scalability can be achieved in QAAL system by applying inference which is assured by dividing the whole logical model into individual independent model. Inference on single large model is more complex than inference on independent smaller models. Semantic data in RDF knowledge bases are indexed in structure way which is directly available and used with keyword queries in this proposed work.

Semantic indexing is used to index all the extracted RDF triples with the free text. In this scheme, naive indexing scheme is used. Random indexing scheme is also used to create orthogonal random vectors for each question. This random vector is created by setting a certain number of
randomly selected values which is either +1 or -1. Each term in question are represented by a vector which is a combination of all index vectors.

Current approaches used for index weighting for information retrieval are based on statistical method. Indexing method can also be implemented by considering semantics and concepts of an ontology. This type of indexing method is called as semantic indexing. Semantic indexing can be implemented by concept vector space technique. Graph representation mainly focused for setting semantic indexing mechanism. So, every concept in ontology is related with question frame model in graph by using semantic indexing scheme. A sample graph form a data structure ontology term with semantic indexing which is represented in Figure 5.1.

![Figure 5.1 Graph Representation for Concepts in Ontology](image)

In Figure 5.1, rectangle represents a concept and arrow represents relation between concepts. RDF triple representation is noted
for linear data structure concept which is designed in the proposed QAAL system in data structure ontology. If user asks a question like “How to implement a linear data structure?” then, it is easy to identify the semantic indexing mechanism for the graph with ‘How’ relationship between the ‘Implementation’ concept and ‘Array’, ‘Linked List’ concepts from the given graph representation. The query representation is like “(How(Implement)→Linear_DS)” in question template of ontology model.

Figure 5.2  QAAL Representation for Linear_DS Concepts in DS Ontology
Figure 5.2 represents the sample index structure stored in template model for given linear_DS concept in the proposed QAAL based DS ontology. Template model identifies the answer which can be automatically returned from the knowledge base by using the indexing scheme. This process is a rule based method in which rules are automatically extracted based on indexing scheme. The accuracy of mapping of semantic role in one-to-one mapping is high. It is important to develop an indexing structure for maintaining the annotation information in semantic approach. It is needed to allow different types of information to be indexed and implemented in this semantic model. It allows fast indexing and retrieval mechanism.

5.4 RDF GRAPH MATCHING WITH SYNONYM FACTOR

Before implementing graph matching in this proposed QAAL system, the system is in need to understand many-to-one relationship of query and answer. For example, a same question may be asked in different manner to get a single answer.

The set of questions are represented in following manner,

Question 1 : Define a stack
Question 2 : What is stack?
Question 3 : Explain basic functionality of stack
Question 4 : Discuss the functionality of stack

The proposed system is in need to automatically recognize the synonyms of relationship between these above questions and will successfully retrieve the correct answer for a given query. So, phrase
terms are necessary for the above problem. Phrase terms are defined as the words in query that remain after removal of stemmed index terms and stop words. For the set of above questions, phrase terms are “Define”, “What”, “Explain” and “Discuss”. This proposed system must infer this phrase terms. If a QA system is able to recognize the pattern between these phrase terms, then the retrieval process will be easy. For this, Synonym Factor (SF) is considered for every phrase terms $p_i$ and $p_j$ in query. In this thesis, $SF_{i,j}$ is the synonym factor with $i^{th}$ question term with $j^{th}$ query term which is defined by the Equation (5.4).

$$SF_{i,j} = \left( \frac{m_{i,j}}{n_i + n_j - m_{i,j}} \right)$$

(5.4)

Where $n_i$ is the number of question that contain $p_i$, $n_j$ is the number of queries that contain $p_j$ and $m_{i,j}$ is the number of $i^{th}$ question term with $j^{th}$ query term.

**Table 5.1 Question with Index Terms and Phrase Terms**

<table>
<thead>
<tr>
<th>No</th>
<th>Question</th>
<th>Index terms</th>
<th>Phrase terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>What is stack?</td>
<td>stack</td>
<td>What</td>
</tr>
<tr>
<td>2.</td>
<td>Compare stack and queue.</td>
<td>stack, queue</td>
<td>Compare</td>
</tr>
<tr>
<td>3.</td>
<td>Explain the difference between FIFO and LIFO</td>
<td>difference between FIFO, LIFO</td>
<td>Explain</td>
</tr>
</tbody>
</table>

The Table 5.1 is used to discuss the index terms and phrase terms in given question. Index terms are the keywords used in users question and phrase terms are the value which corresponds to the given
user question. These phrase terms are referred with a set of question templates which are referred in this proposed QAAL system automatically.

5.5 RDF GRAPH MATCHING WITH INFERENCES USING SEMANTIC INDEXING

Most of RDF based graph representation is in negation form that can’t be answered in SPARQL queries. Inference mechanism is needed to convert the RDF based queries into answer model which is retrieved from knowledge base. Logic is the discipline that studies the principles of reasoning. It is used in formal languages for expressing knowledge. Inference is a method for knowledge acquisition or expansion. It expands knowledge base with additional information using metadata and rules. It helps to uncover unexpected relationships and inconsistencies. It can also be used by intelligent agents for making decisions and selecting courses of action.

Several methods are used to implement graph based similarity detection mechanism. To increase both recall and precision of a query, query modification with conjunctive and disjunctive technique shall be used. Query modification techniques are classified as manual query modification, query rewriting and graph-based query modification. The way to modify query is that the user enters a query and the system returns the result with the part of ontology also. The user navigates and reformulates the query like add or remove query terms.

Query rewriting is done by the system with the ways of augmentation, trimming and term substitution. Augmentation is the technique to enhance the query with terms that are derived from ontology
of the query terms. Different semantics can be used by depending upon the ontological structure.Trimming is a technique of removing query-terms and has the opposite effect of augmentation. Trimming of query is realized by the result comparison of trimmed query with the original query.

Augmentation and trimming exploit a query consisting of Conjunctive of terms (AND) and Disjunction of terms (OR). With the user information, conjunctive queries give high precision and disjunctive queries lead high recall. Disjunctive query augmentation with conjunctive query trimming will increase the recall and conjunctive query augmentation with disjunctive query trimming will increase precision of a query. So, substitution with synonyms, hypernyms or hyponyms from ontology is used to increase recall or precision.

In Figure 5.3, semantic indexing with graph matching technique is shown in QAAL system. After matching RDF graph formation of given question in query form with the sub graph of ontology for retrieving answer in knowledge base, similarity checking is processed by calculation of semantic similarity measurement among concepts, similarity among relations and similarity among graphs. The final representation is checked with the ontology model which is stored as knowledge base by using semantic indexing scheme with template representation. QAAL graph matching is processed with inference engine and the final answer is retrieved after processing this mechanism and correct answer is retrieved from knowledge base.
Figure 5.3 QAAL System with Semantic Searching with QGT Matching and Semantic Indexing

The function of semantic similarity based semantic indexing with semantic inference algorithm is described follows: The input of this algorithm is the RDF graph which is converted into semantic tree form by using QGT matching and the output is a correct answer which is retrieved from knowledge base by using integration of semantic indexing with semantic inference approach. Indexing mechanism is applied in each level of semantic tree with terms $T_g$ used in RDF graph. Before applying indexing and inference mechanism, semantic tree $ST$ is initially checked in ontology. If $ST$ is in ontology ON then, semantic indexing $SI$ is applied for extraction. The terms represented in semantic tree is checked and stored in $T_g$. 
Inference technique is used when the answer is not retrieved correctly from ontology. It is used to improve the time of search to identify the terms in ontology. After applying indexing and inference in semantic tree ST then, the answer is retrieved from ontology and validity of answer is verified. The process is stopped. If the answer A has terms \( T_g \), answer will be displayed. Otherwise, the answer is taken as ‘not available’ type.

<table>
<thead>
<tr>
<th>Input</th>
<th>Output</th>
<th>Steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>IF InFerence</td>
<td>A Answer</td>
<td>1. Get a semantic tree ST from QGT of Query Processing Model.</td>
</tr>
<tr>
<td>ON ONtology</td>
<td></td>
<td>2. If ST is in ON, then</td>
</tr>
<tr>
<td>SI Semantic Index</td>
<td></td>
<td>3. Goto Step 5</td>
</tr>
<tr>
<td>ST Semantic Tree</td>
<td></td>
<td>4. Else Goto Step 12</td>
</tr>
<tr>
<td>( T_g ) Terms in RDF graph in triple set form</td>
<td></td>
<td>5. Apply semantic indexing SI in ST and store it in ( T_g )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6. Apply knowledge and expand with inference IF when answer is not retrieved correctly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7. Check ( T_g ) with ontology ON and get answer A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8. If A is in ON, then</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9. Display A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10. Goto step 13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11. Else</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12. Display ‘Not Available’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13. Stop the process</td>
</tr>
</tbody>
</table>

Algorithm 5.1 Semantic Similarity based on Semantic Indexing with Semantic Inference Algorithm
The algorithm describes the semantic similarity based semantic indexing with inference mechanism for given RDF graph that is represented as semantic tree form. The flow of above steps is represented in algorithm 5.1.

The Table 5.2 represents the question model with the inference representation of data structure ontology in QAAL system. Inference mechanism is implemented for users question with seven categories as definition, comparison, concept specification, example type, quantification, expectation type and verification type methods. In this proposed QAAL system, the integration of semantic indexing with inference mechanism is implemented and after successful formation of inference mechanism with semantic indexing, the performance will be improved in enormous way.

**Table 5.2 Inference Model in Data Structure Ontology**

<table>
<thead>
<tr>
<th>Question Type</th>
<th>Inference model</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td>What is X?</td>
<td>Define a stack?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>What is a tree?</td>
</tr>
<tr>
<td>Comparison</td>
<td>How X differs from Y?</td>
<td>Compare stack and queue. How binary tree differs from binary search tree?</td>
</tr>
<tr>
<td>Concept specification</td>
<td>What? Who? When? Where?</td>
<td>What are the applications of graph?</td>
</tr>
<tr>
<td>Example</td>
<td>What is the example of X?</td>
<td>What are the implementations of stack?</td>
</tr>
<tr>
<td>Quantification</td>
<td>How many?</td>
<td>In how many ways linear data structure can be implemented?</td>
</tr>
<tr>
<td>Expectation type</td>
<td>Why not X?</td>
<td>Why can’t we use one variable for implementing operations in queue?</td>
</tr>
<tr>
<td>Verification</td>
<td>Is X true?</td>
<td>Is graph is a type of tree?</td>
</tr>
</tbody>
</table>
5.6 PERFORMANCE EVALUATION AND DISCUSSION

The evaluation is based on comparison of phrase term factors used in user’s question in query form with the question template model of the ontology in proposed QAAL system. The implementation based on semantic search matching terms with integration of semantic indexing model with inference mechanism is checked.

Carroll et al (2004) have introduced a tool Jena which is an open source framework for building java applications with semantic indexing scheme. Sirin et al (2007) have introduced a reasoner called pellet which is a reasoner used for setting inference mechanism with ontology based environment. To evaluate the performance of QAAL system and AQUALOG QA system, Jena tool with pellet reasoner is used in this thesis for construction of ontology with semantic indexing mechanism.

5.6.1 Evaluation Based on Synonym Factor

To improve the performance of QA system, the semantic indexing takes a major role. Based on the Equation (5.4) synonym factor for different phrase terms with question term and query terms are noted. The Table 5.3 shows the synonym factor for phrase terms used in user’s question. Phrase Term A represents terms in users question represented in query form and Phrase Term B represents terms in ontology which is either represented as concept or properties or individuals.

For every question, synonym factor is calculated with the phrase terms used in question model and query model. The models are compared and analyzed. The synonym factor range is set from 0 to 1. If the question and query terms are having the same words and meaning then, the factor
value will be reached to 1. Otherwise, the value range is varied between the given values 0 and 1.

Table 5.3 Synonym Factor with Phrase Terms in Question and Query Template

<table>
<thead>
<tr>
<th>Phrase Term A</th>
<th>Phrase Term B</th>
<th>Synonym Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>What</td>
<td>Define</td>
<td>0.45</td>
</tr>
<tr>
<td>Explain</td>
<td>Discuss</td>
<td>0.33</td>
</tr>
<tr>
<td>Compare</td>
<td>Difference</td>
<td>0.25</td>
</tr>
</tbody>
</table>

5.6.2 Evaluation Based on Semantic Indexing

QAAL system performance is evaluated based on syntactic indexing and semantic indexing mechanism. The Figure 5.4 shows the comparison of syntactic and semantic indexing mechanism with various metrics of test set five question type. In this figure, precision and recall values are compared for both syntactic indexing and semantic indexing mechanism. F-Measure is calculated based on precision and recall values. MRR ratio is also plotted in Figure 5.4 with ranking order.

All metrics are measured in percentage. Precision value of syntactic indexing at test set five is 57.43% where semantic indexing mechanism produces 83.76 % precision value. Recall ratio for syntactic indexing is 58.23% and 67.45% is achieved by using semantic indexing mechanism in QA system. Based on the precision and recall ratios, F-Measure value is calculated and syntactic indexing mechanism achieves 57.83% in and 74.73% is produced by semantic indexing mechanism. MRR ratio is based on ranking order of the answer. So, semantic indexing
mechanism achieves 75.61% MRR ratio where syntactic indexing produces 52%. These are all noted in Figure 5.4.

![Figure 5.4 Syntactic and Semantic Indexing Evaluation in Test Set Five](image)

It is concluded that the performance of QA system is improved based on semantic indexing compared with syntactic indexing type.

5.6.3 Evaluation Based on Semantic Search with Semantic Indexing

To choose semantic indexing in semantic search mechanism, different evaluation metrics such as precision, recall, F-Measure and MRR values with semantic search, semantic search with syntactic indexing and semantic search with semantic indexing approaches are used for evaluation in AQUALOG QA system and proposed QAAL system. Precision values are noted in Figure 5.5 for the test set five. Three types of
mechanisms are implemented and tested in test set five for AQUALOG QA system and proposed QAAL system.

The test set five contains maximum number of complex questions in it. Three types of mechanisms namely semantic search, semantic search with syntactic indexing and semantic search with semantic indexing are implemented and tested in AQUALOG QA system and QAAL system.

![Figure 5.5 Precision Ratio with Indexing Types in QAAL Vs AQUALOG in Test Set Five](image)

Precision value is reached to 0.55 in QAAL system by using semantic search mechanism at test set five and 0.51 in AQUALOG QA system. 0.64 precision value is reached in AQUALOG QA system based
on semantic search with syntactic indexing which is comparatively low than proposed QAAL system. Because, QAAL system produces 0.75 precision value. Semantic search with semantic indexing mechanism has reached high value compared to semantic search and semantic search with syntactic indexing mechanism. AQUALOG QA system achieves 0.74 precision values by using semantic search with semantic indexing mechanism and the proposed QAAL system achieves a high value 0.84 by using this mechanism.

Figure 5.6 Recall Ratio with Indexing Types in QAAL Vs AQUALOG in Test Set Five

Figure 5.6 shows recall ratios performance with QAAL system and AQUALOG QA systems based on semantic search, semantic search
with syntactic indexing and semantic search with semantic indexing mechanism.

The recall ratio by using semantic search mechanism achieves 0.56 value in AQUALOG QA system and 0.61 in proposed QAAL system. By using semantic search with syntactic indexing method, AQUALOG QA system achieves 0.72 recall ratio and QAAL system achieves 0.81 recall ratio which is comparatively high than AQUALOG QA system. Semantic search with semantic indexing mechanism produces high recall ratios than semantic search and semantic search with syntactic indexing. AQUALOG QA system achieves 0.72 recall ratio and QAAL system produces 0.85 for the same technique. Based on above analysis of various mechanism, it is concluded that the semantic search with semantic indexing mechanism achieves high precision and recall ratios in QAAL system compared with AQUALOG QA system.

Figure 5.7 shows the graphical representation for the comparison of QAAL system and AQUALOG QA system based on F-Measure value. F-Measure ratio is reached 0.76 in QAAL system based on semantic search and 0.60 in AQUALOG QA system with the same type of search. In semantic search with syntactic indexing, this ratio is reached as 0.78 in QAAL system and 0.68 in AQUALOG QA system. But, compared to semantic search with semantic indexing mechanism the above two methods F-Measure ratio is low. F-Measure ratio has reached 0.84 in semantic search with semantic indexing mechanism in QAAL system which is comparatively high than the F-Measure value 0.73 in AQUALOG QA system with the same mechanism.
So, it is concluded that based on semantic search with semantic indexing mechanism F-Measure ratio value in QAAL has been improved 12% than AQUALOG QA system.

Figure 5.8 shows the MRR ratio in percentage for three types of mechanism in QAAL system and AQUALOG QA system. In semantic search, AQUALOG QA system reached 52% MRR ratio and proposed QAAL system reached 65%. Based on semantic search with syntactic indexing mechanism, AQUALOG QA system reached 64.5% where QAAL system has 72% MRR value. After implementing semantic search with semantic indexing mechanism, MRR ratio is increased high value in both AQUALOG QA system and QAAL system. QAAL system has MRR ratio which is high as 84.12% based on semantic search with semantic
indexing mechanism. But in AQUALOG QA system, MRR ratio has reached 64% which is comparatively lower than QAAL system.

![Figure 5.8 MRR Ratio in Percentage with Indexing Types in QAAL Vs AQUALOG in Test Set Five](image)

From the analysis of results, it is concluded that the semantic search with semantic indexing mechanism gives the MRR metric range as high compared to other two approaches.

### 5.6.4 Confidence Ratio

A set of answers are generated by QAAL system which is needed to verify the correct answer. Confidence metric is the main solution to the user to identify the given retrieved answer is in correct form or not. By implementing the confidence, it is in need to count the number of times of the same answer is retrieved from the system. This is
the level of measuring the confidence. The correlation formula is used to find the confidence value which is described by the Equation (5.5). \( \text{corr}(x,y) \) is the correlation between \( x \) and \( y \) where \( x \) is the question target and \( y \) is the answer target.

\[
\text{corr}(x,y) = 0.5 \times \left( \frac{\text{count}(x \land y)}{\text{count}(x)} + \frac{\text{count}(x \land y)}{\text{count}(y)} \right)
\] (5.5)

Where \( \text{count}(x) \) is the number of relevant questions, \( \text{count}(y) \) denotes a number of relevant answers from knowledge base. The computed correlation score is then equally weighted with a simple linear rank score and the answer is calculated by using confidence of a given question \( x \) from the metric \( \text{Confidence}(x) \) in which \( \text{Confidence}(x) \) is the confidence of given question.

\[
\text{Confidence}(x) = 0.5 \times \left( \text{corr}(\text{target},x) + \frac{\text{NumAns} - \text{Rank}(x)}{\text{NumAns}} \right)
\] (5.6)

Where

- \( \text{NumAns} \) is the number of distinct answers
- \( \text{Rank}(x) \) is the original rank for a question \( x \) and
- \( \text{Target} \) is the actual answer target.

If more than one answer is produced then the answers will be reranked in descending order based on the confidence ratio. It is found that answers with a confidence score of 0.5 or less was usually wrong and thus to be removed from correct answer list. The remaining answers are the
final reranked list of answers. The high confidence ratio value of the answer is selected and it is displayed to the user for getting satisfaction from them.

![Figure 5.9 Semantic Search Indexing in QAAL Vs AQUALOG](image)

**Figure 5.9 Semantic Search Indexing in QAAL Vs AQUALOG**

Figure 5.9 shows the Confidence ratio of QAAL and AQUALOG QA system based on semantic search indexing mechanism in difference test sets. Six different test sets are taken for confidence ratio calculation. A set of questions available in each test set is implemented for confidence ratio calculation and average confidence ratio is taken for each test set. In QAAL system and AQUALOG QA system, confidence ratio is achieved more than 0.5 for all six test sets. By implementing semantic search indexing mechanism, AQUALOG QA system reach 0.52 confidence ratio in test set five and QAAL system achieves 0.67 for the same test set.
In test set ten, confidence ratio has been reached 0.93 in QAAL system which is comparatively high than 0.89 in AQUALOG QA system. From the analysis with the figure 5.9, it is concluded that QAAL system performance is improved 6.7% (approximately 7%) based on confidence ratio with semantic search indexing mechanism than AQUALOG QA system.

![Graph showing confidence ratio for QAAL and AQUALOG](image)

**Figure 5.10 Confidence Ratio for Semantic Search with Semantic Indexing and Inference Mechanism in QAAL Vs AQUALOG**

Figure 5.10 shows the confidence ratio by using semantic search with integration of semantic indexing and inference mechanism in QAAL system and AQUALOG QA system. Totally six set of questions are taken for comparison of different types of searching mechanism implemented in proposed QAAL system. Confidence ratio is calculated for all questions in
each test set and average confidence ratio is taken for graph representation.

Semantic search with semantic indexing and inference mechanism has produced 0.79 confidence ratio in AQUALOG QA system in test set five and proposed QAAL system achieves 0.84 ratio for the same test set. In test set seven, AQUALOG QA system reaches 0.82 confidence ratio where QAAL system achieved 0.86 that is high confidence value. In test set ten, 0.97 confidence ratio is reached in QAAL system compared to 0.89 confidence ratio in AQUALOG QA system.

By comparing all six sets of confidence ratio, semantic search with semantic indexing and inference mechanism is improved upto 6% in the proposed QAAL system compared to AQUALOG QA system. It is concluded that the performance of the QA system based on answer extraction model in QAAL system has been improved by using semantic search with integration of semantic indexing and inference mechanism.

5.7 SUMMARY

Indexing mechanism is mainly used to improve the precision and recall values. Compared to syntactic indexing, semantic indexing has gradually increased the performance of proposed QAAL system with 20% in precision, recall, F-measure and MRR values.

Three types of mechanism are implemented and tested in the proposed QAAL system. These three types of searches are semantic search, semantic search with syntactic indexing and semantic search with semantic indexing. In this proposed QAAL system, these three types of mechanisms are implemented and the performance is evaluated based on precision, recall, F-measure and MRR values. With the comparison of all
types of search, semantic search with semantic indexing mechanism gives a successful answer which is used to improve the proposed QAAL system’s performance. MRR ratio of QAAL system is 84.12% which is high compared to 64% MRR ratio in AQUALOG QA system by using semantic search with semantic indexing mechanism.

Confidence ratio is another metric used for verifying the retrieved answer is correct or not based on semantic indexing and inference mechanisms. Confidence ratio for semantic search indexing mechanism is taken in both QAAL system and AQUALOG QA system. The proposed QAAL system performance is improved 6.7% (approximately 7%) than the existing AQUALOG QA system by applying semantic search indexing mechanism. After implementing semantic inference in semantic search based on semantic indexing, the performance is improved high in proposed QAAL system. Compared to AQUALOG QA system, QAAL system performance is improved 6% by applying semantic search with integration of semantic indexing and inference mechanism.

The main goal of this proposed QAAL system is the performance improvement, neglecting ambiguity for retrieving the answer from knowledge base. So, it is concluded that the implementation of semantic search with integration of semantic indexing with semantic inference methodology is the best suitable answer extraction method in the proposed QAAL system for improving the performance of QA system in excellent manner compared with the existing AQUALOG QA system.