Ontology Engineering Techniques for Biological Data

Asma Fardoos, Jamshaid Iqbal Janjua, and Amjad Farooq
Al-Khawarizmi Institute of Computer Science, University of Engineering & Technology, Lahore, Pakistan
{asma.fardoos, jamshaid.janjua}@kics.edu.pk, amjadfarooq@uet.edu.pk

Abstract—Now-a-days the challenge that is increasingly being faced in different domains is how to integrate a large amount of available data and information for a particular domain. The purpose of this integration of data is to get an easy understanding of that particular domain. In this paper our main focus is integration of biological data in single ontology which will be developed using an efficient engineering algorithm. Data integration is being provided through a series of tools and techniques but lack of reliable terminologies and concepts cause hurdles. As a result, on internet a large amount of the available data remains detached & disjoint, even with having access to the internet, we cannot get the advantage of this available data at its maximum level. The dire need to deal with such type of issues has demanded the construction of a large number of ontologies for these domains. Our goal is to provide new technique for gene-ontology to increase the efficiency, completeness and correctness of ontology engineering technique as compared to the previous ontologies developed by different experts.

Index Terms—Ontology engineering techniques; bioinformatics; heterogeneity; RDF Graph; Web Ontology Language; gene ontology;

I. INTRODUCTION

Ontology is a way of representing and organizing knowledge in some logic-based language. It also describes a set of basic terminologies and concepts within a particular domain, and relationship among these terms and concepts [1]. Ontology Engineering is needed when we have to develop a new ontology or combine two ontologies by applying different techniques of ontology matching i.e. mapping, merging and alignment to get the new ontology [2].

We apply different algorithms on these two ontologies by calculating the similarity among them. The algorithm selected to get the new ontology uses the structural characteristics and lexical features of the existing ontologies [3]. These ontology engineering techniques are being used for different types of data in many domains. Currently diverse ontology matching techniques of ontology mapping, ontology merging & ontology alignment are being used to create new ontologies for the biological data [4]. Although there has been done a lot of work in these ontology engineering techniques for different domains, but still some major problems exist in these techniques. These problems are efficiency, completeness and correctness. These ontology engineering techniques are not too much efficient. Our main motivation is to provide an ontology engineering technique for the field of bioinformatics to increase the efficiency, completeness and correctness of ontology engineering technique as compared to the previous ontologies developed by different experts to their maximum level. In this paper our main focus is the development of ontology in the field of bio-informatics specifically for the Gene-ontology. Gene-ontology is the ontology that is constructed and integrated using the data regarding human genes [5]. In this paper, we propose a new technique of Ontology Engineering for Biological data to achieve the following objectives:

1) Improve ontology engineering process to make it efficient.
2) Improve ontology engineering process to make it complete.
3) Improve ontology engineering process to make it correct.
4) Improve ontology engineering process to enhance its quality.

II. LITERATURE REVIEW

The web information resources are growing rapidly in volume and number. With this rapid growth of web, it is becoming tough and time-consuming to retrieve relevant information from these resources [6]. The main cause for the irrelevancy obtained in information retrieval is the way to store web information resources because these are not being stored in the machine-understandable format. It is the estimation that retrieved results are almost 48 percent to 63 percent irrelevant and 37 percent to 52 percent results obtained are relevant that is far away from accuracy. To overcome this problem semantic web has been introduced that is the extension of the current web [7]. This semantic web organizes the content of web...
through ontologies sequentially so that they can be understood by machines. This idea of web semantic can be supported through different techniques of design, development, population and integration of ontology [8].

A. What Is Ontology

Even though it is necessary from ontology to be defined formally, there is no universal definition of the term "ontology" itself. The definitions of ontology can be defined as the conceptualization. It can also be defined as the knowledge body which describes domains usually knowledge domains of common sense [9].

B. Importance Of Web

Now world has become the global village. Today’s fastest growing technology has shrunk the globe into a single huge village. Frequent and rapid movement of information from one node to another in the world at the same time has become possible and the credit goes to the web and internet [10].

C. Need For Semantic Web

Information on the web is never from a single source rather it is scattered across multiple sources. The reason behind the creation of semantic web is to retrieve only the best possible answer for the input query and to present only the relevant and most useful information to the user. The traditional search engines are not capable of providing satisfactory solutions to the user due to certain shortcomings [11].

D. Ontologies In Bioinformatics

Bioinformatics is an emerging field. Growth of biological data has been flared up in the previous twenty years in multiple database formats handled by heterogeneous structures. Across databases not only the semantics are different in their meaning but the structure also differs a lot. Biological data is complex not only due to large data types but it also needs some additional knowledge to specify & constrain the relationships among the data [12]. Ontologies are basically useful in the enhancing the integration process for the purpose to resolve and communicate the organizational and semantic differences among databases of biological domain. Basically concepts and their mutual existing relationships are described through ontology.

E. Techniques For Ontology Engineering

One basic way to built ontology is to start from scratch without reusing the existing ontological or non ontological resources. Starting from scratch has its own procedure and issues and the other one is to produce ontology from already existing ontologies, when we use existing ontologies to produce new ontology there comes the major issue of heterogeneity. The main purpose of ontologies is the data exchange that is done not only at the general syntactic level but this exchange of data is also done at the public semantic level. Over the web a lot of ontologies have been constructed and are being used by replacing the unfeasible methods. Along with this upturn and start of extensive usage of the ontologies however new issues are being experienced [13] [14]. Among these problems major issue is regarding heterogeneity of data at different levels. To resolve the heterogeneity problem at syntactic & structural level stylist solutions have been developed, while semantic heterogeneity problem is still partially resolved. There have been abundant works for finding structural / semantic similarities between graph entities. Some of them have been developed specifically for the domain of ontology alignment while some others have been developed for various other domains e.g. Wordnet similarity but still all of them are valuable for the solution of ontology alignment problem. Some papers have been summarized here which provide the alignment / matching algorithm to reduce this heterogeneity at various levels. Here in this paper our main concern is regarding the building of ontology from scratch for the biological data by using an efficient, more complete and correct ontology engineering technique. New ontology will be built in such a way that it should define all the concepts, attributes of these concepts and the relationships among these concepts so that we can increase the efficiency, completeness and correctness of ontology engineering technique. Ontology engineering techniques are designed towards advancing their interoperability. In what follows a review/analysis of well-known methods among those known by authors is given here.

1) **ASMOV**: Ontology Engineering Techniques have been used in the area of bioinformatics. A paper [15] describes ASMOV (Automated Semantic Matching of Ontologies with Verification) algorithm is used for matching of ontology particularly in the field of bioinformatics. This algorithm takes two ontologies in the form of input to calculate similarity among these two ontologies by using their lexical and structural characteristics. It is obvious that two ontologies of the same field which have been constructed by different experts can be different from each other in their literals and their hierarchy. After measuring the similarity, alignment is derived and it is ensured to its maximum level that there is no semantic inconsistency present in these two ontologies.

2) **Anchor-Flood**: Another paper [16] presents Anchor-Flood algorithm. The ontology in this algorithm is represented using a directed graph. The algorithm starts from an anchor smoothly exploring new concepts from the collection of existing neighboring concepts. Anchor is a collection of concepts in the pair form which have been gathered from every ontology. Ontologies could be in variable sizes i.e. from small size to large size. For example, bioinformatics has a lot of extensive ontologies. This algorithm is very helpful in aligning variable sized ontologies. This feature improves its performance and scalability factor.

3) **PRIOR+**: One more research paper [17] has proposed ontology mapping approach that is generic as well as adaptive approach named as PRIOR+. While mapping ontology, there comes a serious issue regarding heterogeneity nature of ontologies. This approach is used to reduce heterogeneity problem. First of all, vector space model is used in this approach to calculate the similarity
measure of the structural and linguistic characteristics. Then adaptive method is applied to get the aggregates, and at last neural network is offered to improve the correctness of ontology mapping. Results have proved that this is the competitive approach to measure the reliability of similarities among semantic, syntactic and structural features.

4) **CSR**: A paper in the field of ontology alignment [18] presents a method for ontologies alignment that is known as Classification Based Learning of Subsumption Relations (CSR). This technique relates concepts of different ontologies moving from more specific to more general. It takes a couple of ontologies as input, CSR algorithm has to learn the patterns of concepts among these two ontologies and using a subsumption relation to decide that in these ontologies relation exist or not in the concepts pair. This may be done through supervised machine learning methods.

5) **SEMANTIC SIMILARITY ALGORITHM**: In this research paper [19] the technique for similarity measurement of concepts has been proposed. This proposed technique is capable of measuring similarity from single ontology and from multiple ontologies of bioinformatics. An approach for selecting and combining ontologies is also presented. This proposed technique can handle only one form of relation i.e. is-a relation. For experiments, the datasets from biomedical field have been used by experts. The experiments have been done by human employing different ontologies in the framework of “Unified Medical Language System” (UMLS). The experimental results obtained using several datasets of biology and with several ontologies confirmed efficiency of proposed algorithm.

6) **HCOME METHODOLOGY**: In this paper [20] the methodology named as “Human-Centered Ontology Engineering Methodology” (HCOME) has been presented to build and assess the alive ontologies in the communities of knowledge workers. This methodology intends to allow the workers of knowledge the management of their proper conceptualization continuously in daily routine and fill the information gap by actively occupying in the life-cycle of ontology. This HCOME methodology spots the way which is used by knowledge domain workers for the development of their own conceptualization constantly passing through various phases of requirement specifications, ontologies development and maintenance, deployment, evaluation and exploitation.

7) **MEOZO**: In this paper [21] the Methodology for Engineering OWL 2 Ontologies is presented. The basic objective of this paper is to develop ontology by combining the already existing routines of database, ontology and modeling field conceptually. This methodology starts from gathering ontology requirements and ends at maintenance stage by individual validation passing through the stages of ontology creation, instance creation, ensurance of completeness, normalization and ontology validation. This methodology is capable to work with uncertain and in-complete ontology by keeping those ontologies consistent.

8) **ROD**: Another paper [22] presents “Rapid Ontology Development” (ROD) algorithm. The basic objective of this paper is to develop new ontology without having the vast technical domain knowledge. This ROD algorithm has three stages. First stage of ROD is pre-development, second is development and last but not the least one is post-development. Existing ontologies can be employed as whole and partially as well. Therefore, Semantic Ontologies can be built by taking data from sources of semi-structured data.

### III. PROBLEM STATEMENT

Accessing the constantly growing biological data for retrieving and extracting useful information and avoiding the retrieval of irrelevant information is quite tedious task. The main reason behind this irrelevant retrieval is use of improper index terms while storing the information. For the retrieval of relevant information ontology engineering techniques are being developed and used. Although the present ontology engineering approaches are good start, but still the research is required to be able to build up high quality industrial power ontology engineering approaches particularly for biology domain. In this paper, existing approaches will be critically reviewed and a new technique will be proposed.

### IV. PROPOSED ONTOLOGY ENGINEERING TECHNIQUE

#### A. Introduction To Proposed Technique

The main purpose of ontologies is the data exchange that is done not only at the general syntactic level but this exchange of data is also done at the public semantic level. On the web a lot of ontologies have been constructed and are being used by replacing the unfashionable methods. Along with this upturn and starting extensive usage of the ontologies however new issues are being experienced. Sometimes due to the constraint of the provided time ontologies can’t be developed. Normally Ontology life-cycle has three main phases, i.e. Specification phase, Conceptualization phase and the Exploitation phase as discussed previously in HCOME methodology. This technique is not sufficient enough to design and develop the desired ontology within limited time. And if one further divides the conceptualization phase into two phases e.g. knowledge acquisition and ontology development then one can achieve the target within limited time due to the factor of parallelism. Because in this case one is able to develop the ontology while one is acquiring the domain knowledge as the new concepts for a particular domain are gathered / received by one team, at the same time, other team is busy in the implementation of the ontology [23][24]. In this way ontology engineering activity can be accomplished within the provided time. The main objective of this paper is to provide a technique sufficient enough that can fulfill our ontology design and engineering requirements. This proposed technique should be efficient, complete and correct.

In this methodology there is no special change in Specification and Exploitation phase. It only introduces
two new phases those are Knowledge Acquisition and Ontology Development.

**B. Proposed Technique**

Proposed methodology is basically enhancement of HCOME methodology. The phases, goals and tasks for proposed methodology / technique have been described in Table I.

<table>
<thead>
<tr>
<th>Phases of Ontology Life Cycle</th>
<th>Goals</th>
<th>Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Specification Phase</strong></td>
<td>Defining the scope / requirements / aim / teams</td>
<td>1) Discussion (S) 2) Producing documents (S) 3) Identification of coworkers (S) 4) Specifying scope and aim of ontology (S)</td>
</tr>
<tr>
<td><strong>Knowledge Acquisition Phase</strong></td>
<td>Acquiring Knowledge</td>
<td>1) Importing concepts from libraries of the ontologies (P) 2) Consulting from top level generic ontologies (P) 3) Consulting with domain experts through discussions (S)</td>
</tr>
<tr>
<td><strong>Ontology Development Phase</strong></td>
<td>Developing &amp; Maintaining Ontology</td>
<td>1) Improving the ontologies (P) 2) Managing the conceptualizations (P) 3) Merging of developed versions (P) 4) Comparison of own ontology versions (P) 5) Simplify the versions (P) 6) Addition of the documentation (P)</td>
</tr>
<tr>
<td><strong>Exploitation Phase</strong></td>
<td>Using Ontology</td>
<td>1) Browsing ontology (P) 2) Using ontology in different applications</td>
</tr>
<tr>
<td><strong>Evaluation Phase</strong></td>
<td>Evaluating Ontology</td>
<td>1) Start arguments and analysis (S) 2) Compare with others produced versions (S) 3) Evaluate the agreed ontologies (S) 4) Management of discussions recorded on this ontology (S) 5) Suggest versions of new ontology by applying recommended changes (S)</td>
</tr>
</tbody>
</table>

**C. Phases Of Proposed Technique**

The tasks in phases are executed iteratively until a harmony is reached among the workers related to this knowledge. Workers perform these tasks either individually or may be in group form. In the first individual case, we assume that these tasks will be executed in the workers personal space which is marked as “P”. In latter group form, these tasks will be executed in the workers shared space which is marked as “S”. The major tasks for our proposed technique have been discussed in the following sections.

1) **Specification Phase**: This initial Specification phase for the life cycle of ontology is executed in the shared space environment and it includes following missions:
   a) In this phase knowledge worker is doing planning of aim and scope of ontology.
   b) To obtain the common decided requirement specifications, a discussion among the team members is made.
   c) The final decided specifications are recorded in proper documents and the forms.

2) **Knowledge Acquisition Phase**: This Knowledge Acquisition phase for the life cycle of ontology includes following missions:
   a) For the purpose of re-using conceptualizations, existing ontologies are imported.
   b) For the better understanding and clarification of the conceptualizations of a domain, discussion ontologies is done.

3) **Ontology Development Phase**: This Ontology Development phase of ontology life-cycle includes following tasks:
   a) For the purpose of re-using conceptualizations, existing ontologies are imported.
   b) The reuse and progress of ontologies is supported through the various operation of management, mapping and merging of their different versions.
   c) For the purpose of identification of ontologies the comparison for the different versions of ontology is done.

4) **Exploitation Phase**: This last Exploitation phase of ontology life-cycle includes following tasks:
   a) For the purpose to evaluate, review and criticize the specific conceptualizations, by the individuals the assessment of agreed/shared ontologies is done in this last phase.
   b) For identification of differences among these versions comparison of the ontology’s shared versions is done.
   c) To support the decisions of workers posting of different arguments on the ontologies versions is also done.

**V. VALIDATION OF PROPOSED TECHNIQUE**

This section presents the validation of proposed technique through a case study of Gene Ontology that has been generated by us. The validation continues with all the phases of our proposed technique from Specification phase to Exploitation phase.

A. **Gene Ontology**

In this section we will describe Gene Ontology generated by ourselves. Genes are composed of Deoxyribonucleic acid (DNA) & Ribonucleic acid (RNA) [5]. As we already described there are some technologies that are used to represent the data. Here we will use Resource Description Framework (RDF) technique to represent the data related to human genes. RDF is a common framework for the description of resources. Primarily this is intended to represent the metadata that
can be parsed and processed by machines rather than just displayed to people. Anything with identity can be described in RDF and RDF is a good candidate for recording and sharing knowledge on Web. The RDF Schema is being used for describing the properties and classes of the RDF resources [25]. Along with this RDF graph OWL language will also be used which is used for defining the meanings of the terms which are used in the domain relationship and vocabularies of terms according to the domain specifications. The relationship of these terms can also be explained by using the Ontology. OWL is the more powerful language for defining the meanings of these terms in the particular domain and also the semantic meanings of these terms as compared to the RDF, RDF-S and XML [26]. This Gene Ontology will be generated by following all the phases of our proposed technique and those phases are Specification phase, Knowledge Acquisition phase, Ontology Development phase and Exploitation phase.

B. Specification Phase

In this phase scope of the case study will be described. The researcher has taken human genes as a case study to prove the correct working of the proposed technique. Genes are composed of Deoxyribonucleic acid (DNA) & Ribonucleic acid (RNA). As already described there are some technologies that are being used to represent the data. Here proposed technique will be used to represent the data related to human genes and generate the gene ontology. The Ontology Web Language is being used for development of the application that understands and processes meaning of the information instead of only presenting the information to human.

C. Knowledge Acquisition Phase

In knowledge acquisition phase RDF Triples are made by collecting / accumulating all the concepts related to genes from the existing ontologies and also by discussing with the gene experts.

1) RDF Triples: In RDF statements are made. At its basic level, simply the RDF gives us a means for the description of things in terms of properties and property values. A statement in RDF is called RDF triple. Here the researcher tabulated some collection of RDF triples for Gene provided in Table II.

<table>
<thead>
<tr>
<th>Gene</th>
<th>contains</th>
<th>DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>contains</td>
<td>RNA</td>
</tr>
<tr>
<td>DNA Nucleotide</td>
<td>has</td>
<td>Nucleotide</td>
</tr>
<tr>
<td>DNA Nucleotide</td>
<td>has</td>
<td>3 Parts</td>
</tr>
<tr>
<td>DNA Nucleotide</td>
<td>has</td>
<td>Sugar</td>
</tr>
<tr>
<td>DNA Nucleotide</td>
<td>has</td>
<td>Phosphate</td>
</tr>
<tr>
<td>DNA Base</td>
<td>has</td>
<td>Base</td>
</tr>
<tr>
<td>DNA Base</td>
<td>has</td>
<td>4 Parts</td>
</tr>
<tr>
<td>DNA Base</td>
<td>consists of</td>
<td>Adenine</td>
</tr>
<tr>
<td>DNA Base</td>
<td>consists of</td>
<td>Guanine</td>
</tr>
</tbody>
</table>

D. Ontology Development Phase

To check the phase of ontology development, OWL code and RDF graph have been generated. OWL code has been generated from classes and RDF graph from RDF triples.

1) OWL: The given ontology has been implemented in the logical language i.e. OWL. All the concepts of Gene Ontology are implemented as independent classes in this language.

2) RDF Graph: RDF triples are used to create RDF graphs. RDF graph is the conceptual representation for ontology. Now RDF graph, can be created for Gene Ontology by using the collection/set of RDF triples, given in Fig 1.

E. Exploitation Phase

To check the validation we have evaluated the performance of our proposed technique with a Case Study of Gene Ontology. We have also consulted domain experts regarding this developed gene ontology. They declared our developed ontology as satisfactory. Our proposed ontology engineering technique has made the
process of ontology engineering more efficient, complete and correct. Its quality has also been enhanced.

F. Results: Analysis And Discussion

The validation of work can be done through following 5 ways:

1) Experimentally
2) Implementation
3) Case Study
4) Simulation
5) Mathematically

We have done it using case study and now to prove our results we are going to do this mathematically. The basic criteria for the comparison include following four main parameters.

1) Efficiency: The efficiency of an ontology engineering technique is similar to the efficiency [27], used in information retrieval. The efficiency can be written as:

\[
\text{Efficiency} = \frac{\text{InputConceptsCount}}{\text{CorrectConceptsFound}}
\]

2) Completeness: The completeness of an ontology engineering technique is similar to the precision measures [27], which is used in information retrieval. The completeness can be written as:

\[
\text{Completeness} = \frac{\text{CorrectConceptsFound}}{\text{TotalConceptsFound}}
\]

3) Correctness: The correctness of an ontology engineering technique is similar to the recall measures [27], which is used in information retrieval. The correctness can be written as:

\[
\text{Correctness} = \frac{\text{CorrectConceptsFound}}{\text{CorrectConceptsExpected}}
\]

4) Overall Quality of Results: The result of the last parameter overall quality (OQ) is basically based upon the results of correctness and completeness. The overall quality (OQ) can be written as:

\[
\text{OQ} = 2 \times \frac{\text{Completeness} \times \text{Correctness}}{\text{Completeness} + \text{Correctness}}
\]

5) Performance Evaluation with Test Case: For the performance evaluation of our proposed technique, we have conducted an experiment. For this purpose we have taken 25 sample input concepts. We expect that this process will provide at least 15 correct concepts out of total concepts found as output. Following are the steps performed in this procedure:

1) Determine the scope and domain of the ontology
2) Consider reusing the existing ontologies
3) Enumerate the important concepts in the ontology
4) Define classes, their relationships and class hierarchy
5) Define properties of the classes
6) Create instances / concepts

We have performed all these steps for the given techniques in this paper. The domain we have taken is Bio-informatics and the scope has been limited to Human Gene. First of all we have extracted the concepts from existing techniques. After extraction, we listed all the important concepts from these ontologies. Now we have defined classes, inter-relationships and class hierarchy for these listed concepts. After defining the classes, we have described their respective properties. At last step we have created new concepts using the respective techniques. The results for all of these techniques / algorithms have been provided below in the Table III.

Sample input concepts: 25
Correct concepts (expected): 15

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Correct Concepts</th>
<th>Total Concepts</th>
<th>Efficiency</th>
<th>Completeness</th>
<th>Correctness</th>
<th>Overall Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASMOV Algorithm[15]</td>
<td>5</td>
<td>7</td>
<td>0.2</td>
<td>0.71</td>
<td>0.33</td>
<td>0.45</td>
</tr>
<tr>
<td>Anchor-Flood Algorithm[16]</td>
<td>7</td>
<td>9</td>
<td>0.28</td>
<td>0.77</td>
<td>0.47</td>
<td>0.58</td>
</tr>
<tr>
<td>PRIOR+ Algorithm[17]</td>
<td>8</td>
<td>9</td>
<td>0.32</td>
<td>0.88</td>
<td>0.53</td>
<td>0.66</td>
</tr>
<tr>
<td>CSR Algorithm[18]</td>
<td>6</td>
<td>10</td>
<td>0.24</td>
<td>0.6</td>
<td>0.4</td>
<td>0.48</td>
</tr>
<tr>
<td>Semantic Similarity Algorithm[19]</td>
<td>8</td>
<td>9</td>
<td>0.32</td>
<td>0.88</td>
<td>0.53</td>
<td>0.66</td>
</tr>
<tr>
<td>HCOME Methodology[20]</td>
<td>7</td>
<td>8</td>
<td>0.28</td>
<td>0.88</td>
<td>0.47</td>
<td>0.61</td>
</tr>
<tr>
<td>MEO2O Methodology[21]</td>
<td>7</td>
<td>8</td>
<td>0.28</td>
<td>0.88</td>
<td>0.47</td>
<td>0.61</td>
</tr>
<tr>
<td>ROD Algorithm[22]</td>
<td>6</td>
<td>10</td>
<td>0.24</td>
<td>0.6</td>
<td>0.4</td>
<td>0.48</td>
</tr>
<tr>
<td>Proposed Methodology</td>
<td>15</td>
<td>16</td>
<td>0.6</td>
<td>0.93</td>
<td>1</td>
<td>0.96</td>
</tr>
</tbody>
</table>

The graphical representation of comparison is given below in Fig 2-5.
The existing techniques [15-22] used in survey are already over-viewed in Section II. The achievement of the objectives has been confirmed during the analysis of results in Fig 2-5.

VI. CONCLUSION & FUTURE WORK

In this paper an efficient, complete and correct technique for the development of new ontologies has been presented.

A. Conclusion

This paper has provided a new ontology engineering technique to built ontology after better and useful review of already proposed different ontology engineering algorithms. All these algorithms have some issues regarding the correctness and completeness. This newly proposed ontology engineering technique is very important because it increases these two issues up to some extent. It’s been tried to develop an ontology engineering technique to increase the efficiency, correctness and completeness along with its quality. The proposed technique has also been validated by using the case study of human genes ontology by following the phases. This technique is very useful for everybody especially the biologist and computer scientist who can use this algorithm for the data representation and integration.

B. Future Recommendations

Although this algorithm deals only with gene ontology of bioinformatics domain but this research can be extended to other areas of bioinformatics domain to engage the concepts, vocabularies and databases from DNA and structured protein ontologies as well as to the bio-medical domain. We can extend this ontology engineering approach to remove the semantic heterogeneity in future & to enhance the re-usability of this approach so that the advantage of available data over the internet can be achieved at maximum level.

REFERENCES


Asma Fardoos received her Bachelor degree in Computer Science from University of Engineering and Technology, Lahore in 2009. She completed her Masters in Computer Science from University of Engineering & Technology, Lahore in 2013. She is working at Al-Khawarizmi Institute of Computer Science, University of Engineering and Technology, Lahore as Research Officer. Her research interests are Semantic Web, Bioinformatics, Ontology Engineering, Software Engineering, Information Retrieval and Web Methodologies.

Jamshaid Iqbal Janjua received his Masters degree in Computer Science from University of Central Punjab, Lahore in 2009. Currently, he is pursuing his Ph.D in Computer Science from University of Central Punjab, Lahore. The author became a Member of IEEE in 2011. He is an active member of IEEE Computer Society, Lahore Section. He is Senior Manager Technical Research at with Al-Khawarizmi Institute of Computer Science (KICS), University of Engineering & Technology, Lahore, Pakistan. His research interests are Business Intelligence, Software Engineering, Energy Management Systems & Surveillance Science.

Amjad Farooq accomplished his Masters degree in Computer Science from Bahauddin Zakariya University, Multan in 1996. He completed his Ph.D degree in Computer Science from University of Engineering and Technology, Lahore in 2009. Currently he is teaching at Computer Science & Engineering Department, University of Engineering and Technology, Lahore as Associate Professor. His research interests are Ontology Matching, Semantic Web, Software Engineering and Web Engineering.