Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu

Cloud based automated framework for semantic rich ontology construction and similarity computation for E-health applications

T. MuthamilSelvan*, B. Balamurugan

School of Information Technology & Engineering, VIT University, Vellore, Tamilnadu 632014, India

ARTICLE INFO

Keywords:
Ontology
Implicit knowledge
Conditional dependence
Graph derivation representation
Dyadic deontic logic
Cosine similarity

ABSTRACT

Ontology structure, a core of semantic web is an excellent tool for knowledge representation and semantic visualization. Moreover, knowledge reuse is made possible through similarity measure estimation between two ontologies, threshold estimation and use of simple if-then rules for checking relevancy and irrelevancy measures. Reduced semantic representations of the ontology provide reduced knowledge visualization which is critical especially for e-health data processing and analysis. This usually occurs due to the presence of implicit knowledge and polymorphic objects and can be made semantically rich through the construction by resolving this implicit knowledge occurring in the form of non-dominant words and conditional dependence actions. This paper presents the working of the automated framework for the construction of semantic rich ontology structures and store in the repository. This construction uses dyadic deontic logic based Graph Derivation Representation in order to construct semantically rich ontologies. Moreover, in order to retrieve a set of relevant documents in response to the cloud user document, the degree of similarity between two ontologies is estimated using the traditional cosine similarity measure and simple if-then rules are used to determine the number of relevant documents and obtain such document's metadata for further processing. These working modules will be extremely beneficial to the authenticated cloud users for document retrieval, information extraction and domain dictionary construction which are especially used for e-health applications. The proposed framework is implemented using diabetes dataset and the effectiveness of the experimental results is high when compared to other Graph Derivation Representation methods. The graphical results shown in the paper is an added visualization for viewing the performance of the proposed framework.

1. Introduction

Cloud Computing services is highly a brand new paradigm for providing various services at different levels of infrastructure, platform and software. This is an enormously developing domain because of the major benefits like flexibility, pay-per-use model thereby reducing the cost significantly. This comprehensive definition and the major benefits is provided by NIST as indicated in [23]. Accordingly Cloud computing is a payment model according to the usage for the provision of the available, convenient, secured and on-demand network access to a shared and distributed pool of configurable computing resources with regard to networks, servers, memory capabilities, applications, software services. Such pay and use services can be securely and rapidly deployed and maintained with minimal technical management effort or cloud service provider interaction [23]. However, the cloud computing services must facilitate the factors like scalability, pay-per-use utility model, distributed architecture, security essentials and virtualization concepts [24]. Cloud Computing services is essentially a new business management paradigm [25] that empowers the on-demand access, elasticity, pay-per-use, long lasting connectivity, availability, highly secure, shared resource pooling and virtualized infrastructure [26].

The term Ontology is closely related to the semantic structure which means "theory of existence". The main advantage of such semantic structure is that they provide a knowledge-sharing framework that supports the representation, sharing and the subsequent reusability of domain knowledge [1]. Ontologies have been widely applied in many fields such as knowledge management, information retrieval, Semantic Web, information integration, semantic search and recommendation systems. The value added feature in cloud computing service in this paper is identified as an Ontology as a Service with the major effect on Infrastructure as a Service. However, the use and the underlying concept of Ontology as a Service were initially proposed in [27]. The authors have defined this terminology, Ontology as a Service (OaaS) is “a service where Cloud service providers deploy the ontology construction application and infrastructure together based on the users' requirements. In this paper, the ontology construction for the text

* Corresponding author.

http://dx.doi.org/10.1016/j.imu.2016.12.002
Received 19 November 2016; Received in revised form 12 December 2016; Accepted 13 December 2016
2352-9148/ © 2016 Published by Elsevier Ltd.

documents posted by the authenticated cloud users and the estimation of the related documents through the use of ontology alignment procedure is done in the cloud server. This process is facilitated by the cloud service provider.

The syntactic and the semantic knowledge of the target input text document can be expressed using several knowledge representation languages used in artificial intelligence like logic, scripts, frames, etc [19]. The first aspect with respect to the ontology construction is closely related to the expressivity of the structure. Ontology can be expressed in different logics such as predicate, fuzzy, temporal, situational, description logic and modal logic [3]. Many of the applications make use of Description Logic (DL) for knowledge representation. However, for certain applications, the use of DL is not feasible for perfect and expressive structure due to the presence of non-dominant words in the target datasets. On the contrary, the knowledge from the dataset will be perfectly dispatched only if the structure is expressive. In such cases, the expressivity of the target data will be reduced, causing several issues like instability and incompleteness. Moreover, the presence of polymorphic objects in the dataset poses to be a very challenging issue wherein the expressivity is a main critical issue [20]. Therefore, it is necessary to enhance the expressivity by uncovering the implicit semantic knowledge, providing expressivity is by using modal logic that covers non-dominant words and conditional probability events. Dyadic Deontic logic is a kind of modal logic and has a great impact of non-dominant words occurring in the documents. It is the formal study dealing with the statements of obligation, forbidden, permissible, conditional obligation and conditional permissible clauses. It can handle sentences containing negated words like SHOULD_NOT, MUST_NOT, SHALL_NOT, COULD_NOT, WILL_NOT, and conditional dependence statements instead of the conventional negation symbols as used in the other logic languages. In addition to this, it includes the other symbols that are available in description logic.

The second aspect in ontology is the ability to reuse the constructed ontology, since newly generated ontology every time is a time consuming process. This concept of using the semantics again is called ontology reuse. In this process of ontology reuse, the semantic knowledge of an existing ontology can be used for a newly constructed ontology even in a heterogeneous environment [21]. Therefore, reusability estimation is an important parameter to identify the degree of intersection. In order to facilitate this, some measures of similarity or intersection computation can be used. Out of several methods used for similarity computations used in the literature, various the distance measures might be used for measuring the degree of similarity between two ontology structures.

1.1. Need of the hour - semantics

The process of automatically exchanging, sharing and reusing the data or information in the World Wide Web is critical and often challenging. In the midst of the advancement of information technology, the usage of the above issues in the web are very limited due to the heterogeneity problem prevailing in the information resources and the non-semantic nature of HTML, XML and their underlying URL [2]. There are many techniques available in the literature to solve syntactic and structural heterogeneity problems [21]. However, the semantic heterogeneity problem is always a great challenge to be resolved. Semantic heterogeneity is a problem that two contexts do not share the similar understanding of information. Some of the semantic heterogeneity problems are synonym sets, concept lattices, features and constraints [6]. These problems are solved to an extent in the past. However, effective techniques are necessary to resolve this problem completely.

1.2. Reusability – degree of similarity

Semantic heterogeneity problems are solved by ontology structure. The process of Ontology Alignment in semantic web aims to find semantic correspondences between similar elements of different Ontologies using ontology reuse measures. Ontology and the subsequent ontology alignment process is widely used in many applications areas, such as knowledge management [5], electronic commerce, E-Learning, and information retrieval systems [8], semantic search and recommendation systems [22].

Manual ontology alignment is very critical and time-consuming when it is performed manually as the size and complexity of the ontology structure increases. Hence, automatic ontology alignment became a well-known technique in many practical applications including information transformation and data integration, query processing, E-commerce, E-Learning, Information Retrieval and Recommendation systems [4]. The Ontology Alignment techniques existing in the literature are methods based on Strings, Languages, Constraints and Semantics [7,9]. However, most of the existing Ontology Alignment techniques used in the literature suffer from two main limitations:

1. Reduced semantic expressivity of the constructed ontology,
2. Concepts, Relationships between the concepts, axioms and the path links in the existing frameworks are retrieved based on the occurrence of only dominant words in the input text documents. Therefore, it is necessary to provide intelligent techniques for effective Ontology Alignment for the purpose of ontology reuse.

1.3. Objectives

In this paper, an automated framework is proposed which provides separate working modules for ontology construction, measuring ontology expressivity and estimation of the degree of similarity between two different ontologies. This degree of similarity estimation facilitates the cloud service provider to provide related documents to the authenticated cloud users. Such retrieval is provided by the threshold value in the similarity degree and the use of ordinary if-then rules for related documents retrieval. In case of ontology construction module dyadic deontic logic based GDR (Graph Derivation Representation) technique is used for constructing sematic rich expressive ontology. There are four different phases in the proposed framework. In the initial phase, the cloud users are properly authenticated using the traditional username-password mechanism. Subsequently, the authenticated cloud users post their unprocessed but rather meaningful documents to the cloud service provider for further processing. These unprocessed documents are converted into dyadic rules representation to construct highly expressive ontology structure. In the second phase, a GDR for each concept, the different relations and their corresponding instances in a given ontology is generated. This is facilitated by the recursive process of graphical derivations. Later, an integration technique is applied to merge multiple graph node structures in order to produce an initial integrated GDR for the given ontology. As a result, a complete GDR representation of the given ontology is generated by deleting the unstable relationships for semantic measurements are done. In the third phase, the sematic expressivity factor of ontology is estimated and the degree of similarity between two different ontology structures is identified using cosine similarity metric. In the final fourth phase, the related documents are retrieved and provided to the authenticated cloud users. This is facilitated by the threshold estimation module and the ordinary if-then rules construction. The major objectives of the proposed framework are given below:

- To facilitate the deployment of raw text document to the cloud service provider by the authenticated cloud users.
- To provide a semantically stable ontology structure of the underlying knowledge using GDR
- To visualize the semantically expressive ontology structure using the implicit knowledge, non-dominant words and conditional probability event occurrences.

- To estimate the expressivity factor of the semantically rich ontology structure.
- To compute the degree of similarity between two different ontology structures using cosine similarity metric.
- To retrieve and provide the related documents metadata to the authenticated cloud users based on the rules metric.

1.4. Quick analysis on the objectives

There are six notable objectives in the proposed framework. The following discussion is a quick step by step analysis of the objectives using an example.

Step 1: Since the dataset used in this paper is diabetes dataset, the users are preferably the people working in hospital environments. The cloud users in this paper can be doctors, nurses, lab technicians, dean of the hospital, etc. have the facility in posting some text documents related to medical records. These documents can be in any format and posted by any authenticated cloud user. For instance, a cardiologist can post a document pertaining to a recent technology in analysis of the disease or related to surgery.

Step 2: The size of this posted document can be enormous and also the technical aspects of the document should be help to any person reading the document even if he is not related to the hospital domain. For instance, the document can contain many keywords related to the heart disease and surgery. To obtain the underlying knowledge from the document a sematic knowledge representation must be developed. In this paper, an ontology structure is developed.

Step 3: With reference to step 2, for the purpose of constructing ontology structure there are several existing techniques. However, some of the logic representations will not be accurate in case of the occurrence of non-dominant words like can, will, cannot, may not, etc. and other conditional probability events. In the proposed framework, a highly expressive ontology structure is developed.

Step 4: For the purpose of comparing the expressivity with respect to the other logic representations, some factors are identified like the total number of classes, relationships and instances. These numbers are very high in the proposed framework compared to the other existing techniques.

Step 5: This is the next step for identifying the similarities (differences) between two different documents. In case many cardiologists are posting different documents probably in the same domain, the similarity between these two posted documents can be identified. This will be very essential in case of use of a different technology called ontology merging. Ontology merging is a subfield of knowledge representation and this can be used for merging two similar documents posted by many doctors. However, ontology merging is not discussed in this paper. The proposed framework is restricted to the computation of similarity between two documents alone.

Step 6: The generic users using this framework can obtain many relevant documents based on the user's input document. For obtaining the relevant documents, the similarity computation is very essential and for the retrieval simple if-then rules classifier is used. The metadata of all the relevant documents are retrieved to the authenticated cloud user. For instance, a doctor can post a single document and can obtain multiple related documents' metadata which can be further used for analysis or documentation.

The remainder of this paper is structured as follows. Section 2 presents a quick survey of the related works. Section 3 gives a detailed description of the proposed framework. Section 4 discusses performance analysis of the proposed framework. The final section gives the concluding remarks and a few directions of the future work.

2. Literature review

Many graphical models are present for ontology construction and similarity computation measure [3]. Unified Modeling Language (UML) based Object Constrained Language (OCL) is one such technique. OCL is used as a graphical model for ontology representation. UML is suitable for representing explicit taxonomical information instead of implicit (hidden) non-taxonomic relationship [12]. The description of semantic relationships among existing objects can be identified using Semantic Link Network (SLN) technique. The property of semantic richness is given importance in SLN helps rather than semantic correctness [11].

Measuring ontologies based on ontology measures is called Ontology measurement and the existing ontology measures uses only the explicit knowledge exhibited by ontologies to compare similarity of ontological entities and structures explicitly expressed in ontologies. Cluster-based techniques are used in the literature which combines the minimum path length and the taxonomical depth and defines clusters for each of the branches in the hierarchy with respect to the root node. An ontology-based measure utilizing taxonomical features is proposed in many applications without using the tuning parameters to weight the contribution of potentially scarce semantic features [13].

The relevant super-concepts and sub-concepts of the two different concepts are extracted from the ontology structure and then use a similarity function to determine a similar concept class by a matching process based on semantics [14]. Graph-based ontology terms are used to calculate the similarity of two gene products in their proposed work. Quality measures were introduced to measure and evaluate certain ontology quality properties such as expressivity, cohesion, complexity, richness, degree of similarity [14–18]. However, most of the available system frameworks for handling polymorphic objects of ontology representation are limited, inaccurate and inefficient. In this paper, we define a solution by developing an automated framework for constructing a stable and highly expressive ontology structure that could efficiently handle polymorphism in ontology representation. The framework also aims to estimate the degree of similarity between two different ontology structures for the future ontology reuse.

2.1. Analysis of earlier works

The target dataset in the form of a graphical model must provide the following features given below:

- It should have the power of explicitly expressing the semantic knowledge including the implicit kinship of concepts and non-taxonomic relations. The existing ontology measures must be still applicable to the model.
- The ontology construction framework must be capable of handling non-dominant words and conditionally occurring events with respect to the conditional probability events.
- The problem of polymorphic objects in ontology representation must be addressed in the model to ensure the stability of the ontology structure.
- It must satisfy the fundamental factor for estimating the expressivity of the ontology structure.
- Automatic computation of the degree of similarity value must be feasible and integrated in the automated framework.

However, most of the existing graphical models discussed in the literature survey fail to satisfy the above features of a graphical model for representing the semantics. Hence, it is necessary to devise a new technique for generating a GDR which represents the implicit knowledge hidden and the explicit knowledge. Moreover, it is also essential to devise some algorithms to solve the problem of polymorphism in such explicit and implicit knowledge representation [10].

In general, ontology can be expressed using different knowledge

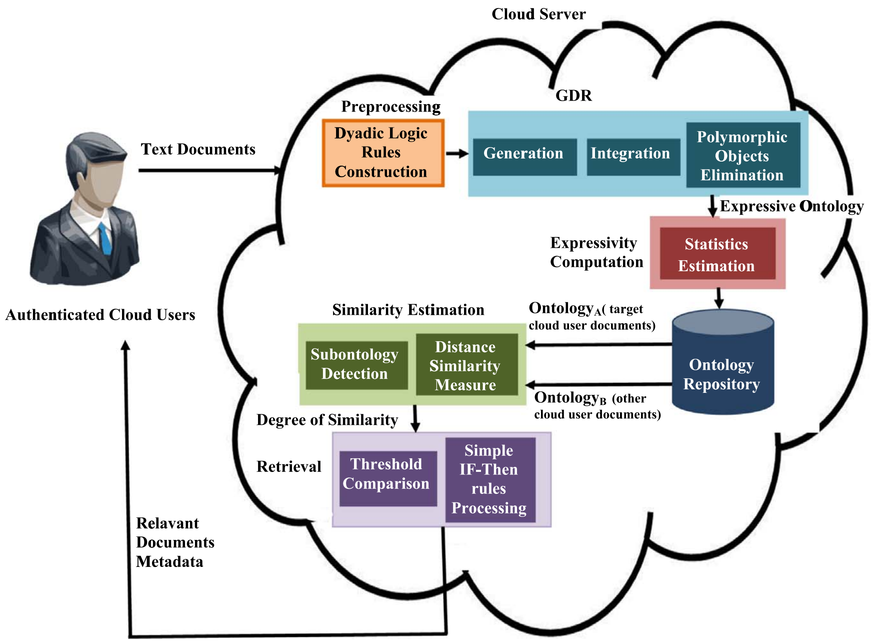
Cloud Server
Text Documents
Preprocessing
Dyadic Logic Rules Construction
GDR
Generation
Integration
Polymorphic Objects Elimination
Expressive Ontology
Expressivity Computation
Statistics Estimation
Authenticated Cloud Users
Similarity Estimation
Ontology$_A$(target cloud user documents)
Subontology Detection
Distance Similarity Measure
Ontology Repository
Ontology$_B$ (other cloud user documents)
Degree of Similarity
Retrieval
Threshold Comparison
Simple IF-Then rules Processing
Relavant Documents Metadata

Fig. 1. Proposed automated framework.

representation languages like logic, frames, semantic nets, etc. Most of the existing works on ontology representation uses logic as the knowledge representation language. The presence of implicit knowledge is attributed by the presence of non-dominant words in the target data set. Moreover, there are some situations wherein the statements not only contain non-dominant words but also contain conditionally probable events. This paper aims to enhance the expressivity by identifying and processing the dominant, non-dominant words and conditionally probable events also. Moreover, in future for the purpose of ontology reuse some measures of similarity computation is essential.

3. Proposed system framework

The authenticated cloud users have the facility of obtaining highly related documents in response to their posted raw documents. GDR provides a graphical model for the semantic descriptions for text documents in this proposed automated framework. The goal of generating GDR's for ontologies is to measure and compare ontologies based on their underlying GDR for stable semantic measurement. It helps to derive and understand the complete structural semantics for the target ontology. On successful generation of a stable and expressive ontology using GDR technique and dyadic rules generation, the expressivity of these constructed ontology structure must be evaluated. This expressivity measure helps in identifying the extent of representing the implicit knowledge of the text document. For the purpose of achieving enhanced expressivity, dyadic deontic logic is used for knowledge representation before transforming into the corresponding GDR.

The proposed automated framework also facilitates to estimate the degree of similarity between different ontologies. This is done by using cosine similarity measure. Subsequently, the threshold value is estimated after performing several experiments on the underlying dataset and the simple rules are used for obtaining the related documents based on the similarity value. Therefore, the proposed automated framework provides feasible solution for constructing semantically rich ontology structures, expressivity measurement, degree of similarity estimation and the retrieval of related documents using simple rules metric. This automated framework can be used in several applications like text information retrieval, domain dictionary construction, information extraction, recommendation systems, etc. The architecture of the proposed automated framework is shown in Fig. 1.

3.1. Cloud user

In order to enjoy the facilities provided by the cloud service provider for related text documents retrieval, the user must be an authenticated user. These users are generally called as an authenticated cloud user since, these users obtain the mentioned services from any of the cloud service provider. For the purpose of authentication, the traditional security metric of using the respective username-password combination is used. This combination metric shall facilitate the users to obtain the services continually and securely from the cloud service providers.

3.2. Dyadic deontic logic representation

The input given to the proposed framework is a text document from the repository. Since, dyadic deontic logic is a powerful knowledge representation tool and deals with the statements such as obligatory, forbidden, permissible, conditional obligations and conditional permissible. Therefore, the sentences in the text document can be transformed to its corresponding dyadic deontic logic format by identifying the clauses of obligatory, forbidden and permissible, conditional obligations and conditional permissible statements added to the standard deontic logic statements. Once these statements are found, they can be represented in dyadic deontic logic using the suitable constructors. Finally the format is converted into the form using the operators such as ◇ (possible),□ (necessary),^ (conjunction),˅ (disjunction), negation (forbidden), if-then (conditional) and A | B (conditional probability) statements.

3.3. Rules for detecting dyadic deontic relationships

Rule 1 - If X is a noun and X is related to Y by attribute or part of relationship and there exists a Determiner relationship between X and Y then OBLIGATORY(X HAS Y).

Rule 2 - If X is a noun and X is related to Y by attribute or part of relationship and there is a Modal relationship between X and Y then.

Rule 2.1 - If the modal relationship is MUST or SHOULD then OBLIGATORY(X HAS Y).

Rule 2.2 - If the modal relationship is CAN or WILL then PERMITTED (X HAS Y).

Rule 3 - If X is a noun and X is related to Y by attribute or part of relationship and there is a Dyadic modal relationship between X and Y then.

Rule 3.1 - If the modal relationship is CONDITIONAL MUST or CONDITIONAL SHOULD then CONDITIONAL OBLIGATORY(X | Y).

Rule 3.2 - If the modal relationship is CONDITIONAL CAN or CONDITIONAL WILL then CONDITIONAL PERMISSIBLE (X | Y).

Rule 4- If X is a noun and X is related to Y by part of or attribute relationship and consists of negative modal relationship.

Rule 4.1 - If the modal relationship is MUST NOT or SHOULD NOT then FORBIDDEN (X HAS Y).

Rule 4.2 - If the modal relationship is CAN NOT or WILL NOT then FORBIDDEN (X HASY).

Rule 5- If X and Y are nouns and are related with property Of relationship OBLIGATORY (X is NOT NULL).

Rule 6- If X and Y are nouns and are related by isA relationship OBLIGATORY (X has attribute TYPE).

Rule 7- If X and Y are nouns and X is related to Y by instance of relationship OBLIGATORY (X has instance Y).

Rule 8- If X and Y are nouns and X is related to Y by contains relationship OBLIGATORY (X HAS Y).

3.4. Mathematical predicate

3.4.1. Predicate calculus for deontic rules

RULE 1 ∀x, ∃ y - > OBLIGATORY(x,y).

RULE 2.1 MUST(x,y) ∨ SHOULD (x,y) - > HAS_OBLIGATORY(x,y).

RULE 2.2 NOUN(x) ∧ NOUN(y) ∧ CAN(x,y)- > HAS_PERMITTED(x,y).

RULE 3.1 NOUN(x) ∧ NOUN(y) ∧MUST(x|y) - > CONDITIONAL_OBLIGATORY(x,y).

RULE 3.2NOUN(x) ∧ NOUN(y) ∧SHOULD(x|y) - > CONDITIONAL_OBLIGATORY(x,y).

RULE 3.3NOUN(x) ∧ NOUN(y) ∧CAN(x|y) - > CONDITIONAL_PERMITTED(x,y).

RULE 3.4NOUN(x) ∧ NOUN(y) ∧SHALL(x|y) - > CONDITIONAL_PERMITTED(x,y).

RULE 4 NOUN(x) ∧ NOUN(y) ∧ MUST_NOT(x,y) ∧ SHOULD_NOT(x,y) - > HAS_FORBIDDEN(x,y).

RULE 5 NOUN(x) ∧ NOUN(y) ∧CAN_NOT(x, y) - > HAS_NOT_PERMITTED(x,y).

RULE 6 NOUN(x) ∧ NOUN(y) ∧ PROPERTY_OF(x,y) - > OBLIGATORY(x, NOTNULL).

RULE 7 NOUN(x) ∧ NOUN(y) ∧OBLIGATORY(x, y) - > HAS_ATTRIBUTE(x, TYPE).

3.5. Graph derivation representation (GDR)

The second working module includes GDR with three major submodules namely GDR Generation, Integration and Elimination of Technical Barriers [22]. This module generates the GDR by identifying the axioms present in the dyadic deontic logic. The graph derivation process is conducted in three phases based on the three mapping functions ρ, λ and η. In the first phase, each axiom and assertion is indexed with positive integers. The $G_O$is originally set to empty, and has no vertex and relation. Then, each axiom/assertion α is examined and the GDR (denoted as Gα) is generated for each α. Once the GDR for each axiom/assertion is generated, the second phase is started, which integrates each GDR into G_O by the integration operation. The integrated (but untreated) GDR for the given ontology is obtained at the end of the second phase. In the third phase, G_O is treated by eliminating cycles of class inheritance and non-direct relationships with transitive property. The final complete GDR is obtained from the second working module. In the proposed framework, the integrated GDR is found to be highly stable by avoiding the polymorphic objects. This is evident from the estimation of the stability factor. The Stability Factor is defined as.,

$$S=\{G_{O1}\ ,\ G_{O2}\ ,\dots\ ,\ G_{On}\} \tag{1}$$

Such that $G_{On}=\{V_{On}, E_{On}, \rho, \lambda, \eta,\beta\}$.

— V_O is a finite set of vertices, where each vertex is a unique positive integer.

— *$E_O \subseteq V_O \times V_O$is a set of edges.*

— ρ: C → V_O is a mapping function, where C is the set of the defined concepts and individual examples in O.

— λ: A → $E_O \cup V_O$ is a mapping function, where A is the set of axioms/assertions in O.

— η is a labelling function that assigns a set of literal names $\eta(i) \subseteq N_L$ to each vertex $i \in V_O$, and a set of literal names $\eta(i, j) \subseteq N_P$ to each edge $(i, j) \in E_O$, where $N_L = N_C \cup N_I$, and N_C, N_I and N_P are the sets of literal names of concepts, individual examples and role relations, respectively.

3.6. Expressivity measurement

Ontology measures are selected based on the measurement entities such as Fine-grained entities and Coarse-grained entities. Fine-grained are the basic elements of ontologies such as concepts/classes, properties, binary relationships, axioms and examples. Coarse-grained are the other ontological elements such as Fanin and Fanout. However, in the proposed framework only fine-grained ontology elements are analyzed. Considering the coarse-grained elements in the ontology structure like fanin and fanout are analyzed in the future work. The following measures are calculated for expressivity estimation which uses some of the measurement entities such as concepts, individual examples and role relations.

For any ontology, O_i where i=1 to n (and O_i in repository), the following parameters are calculated.

$$\text{NOC}\,(\text{number of classes}){:}\ NOC\,(O) = |SC|,\ \text{where SC} = \text{set of classes.} \tag{2}$$

$$\text{NOP}\,(\text{number of examples}){:}\ NOE\,(O) = |SE|,\ \text{where SP} = \text{set of examples} \tag{3}$$

$$\text{NOA}\,(\text{number of axioms}){:}\ NOA\,(O) = |SA|,\ \text{where SA} = \text{set of axioms} \tag{4}$$

$$\text{NOL}\,(\text{number of path links}){:}\ NOL\,(O) = |SL|,\ \text{where SA} = \text{set of path links} \tag{5}$$

On successful calculation of the number of concepts, examples, axioms and path links from Eqs. (2–5), the expressivity measure of an particular ontology for a dataset is given by,

$$E\,(O_i)=Stat\,(O_i) \tag{6}$$

Where $Stat\,(O_i)=\Sigma_i\,(|NOC_i| \wedge |NOE_i| \wedge |NOL_i|)$.

Moreover, any two ontologies can be compared by using this E(O) measure in order to find out the their expressiveness factor. Such measure can be estimated recursively using user defined functions or procedures. The Expressiveness (E) of the target ontology O_i is a Boolean metric and it is compared with all the other ontology structures present in the repository. For any two ontologies, $O_i and O_j$ from the repository,

$$E\,(O_i,\,O_j)=\begin{cases} 0, \ \&\ if Stat\,(O_i)\ <\ Stat\,(O_j) \\ \quad 1,\ \&\ Otherwise \end{cases} \tag{7}$$

3.7. Degree of similarity measure computation

This module in the automated framework concentrates on the second objective of estimating the reuse measure [15]. This component includes three sub modules such as Sub-ontology detection, Maximal common subgraph determination and cosine similarity measure [16]. The input to this module is an ontology repository. The effectiveness of technique of knowledge representation using GDR can be computed by aligning the constructed ontologies. Such ontology alignment is based on two aspects namely sub-ontology detection and measuring the semantic cosine similarity between two ontologies.

3.7.1. Sub-ontology detection

Sub-ontology detection is the process of finding whether one ontology is the sub-ontology of the other. From the graphical perspective, Ontology O_i is a sub-ontology of $O_j$iffG_{Oi} is a subgraph of G_{Oj}. G_{Oi} is a subgraph of G_{Oj}, and denoted as $G_{Oi}\subseteq G_{Oj}$, iff there exists an onto function sub:

$V_{Oi}\rightarrow V_{Oj}$such that:

- For any vertex $m\in V_{oi}$, $\eta1$ (m) $\subseteq \eta2$ (sub(m)).
- For any vertex $n\in V_{oi}$,$\eta1$ (n) $\subseteq \eta2$ (sub(n)).
- For any edge(m, n) $\in E_{Oi}$,$\eta1$ (m, n) $\subseteq \eta2$ (sub(m), sub(n)).
- For any path link(m, n) $\in E_{Oi}$,$\eta1$ (m, n) $\subseteq \eta2$ (sub(m), sub(n)).

By testing the inclusion relations between the sets of labels of vertices and edges from the two GDRs, it can easily be concluded that one ontology is a sub-ontology of another ontology (i.e. one graph is a subgraph of another).

3.7.2. Distance similarity

In this module, the final objective of the degree of similarity computation is resolved. This is facilitated by using the cosine distance measure. In order to facilitate this computation, normalized weight values are assigned between the concepts present in the vertices of the GDR representation. These weight values in the edges among the vertices are assigned using some standard metrics [12]. According to the weight assignment, the cosine distance metric between any two graphs G_{oa} and G_{ob}, commonly denoted by *dSim*.

dSim(G_{Oa}, G_{Ob}), is defined as follows.

$$dSim=\frac{\sum_{i,j=1}^{n} V_{i,a}V_{j,a}\cdot \quad \sum_{i,j=1}^{n} V_{i,b}V_{j,b}}{\sqrt{\sum_{i,j=1}^{n} (V_{i,a}V_{j,a})^2}\sqrt{\sum_{i,j=1}^{n} (V_{i,b}V_{j,b})^2}} \tag{8}$$

where$V_{i,a}V_{j,a}$are weight *v*alues from *v*ertex *i to* vertex *j* in graph *Goa* and $V_{i,b}V_{j,b}$are weight values f rom vertex *i* to *v*ertex *j in g*raph *Gob* and

$\forall V_{i,a}V_{j,a}= \quad V_{i,b}V_{j,b}$

The degree of similarity between two ontologies decides the degree of reuse during ontology alignment and is normalized. The similarity values occur between 0 and 1. The reusability is decided by fixing a threshold. In this paper, the threshold value for the reuse is fixed to be 0.6. However, this value has been fixed after various repeated experiments in different domain applications and is not a benchmark value. If an ontology has been detected as sub-ontology of another ontology, they represent the same domain, but the knowledge scopes they cover in the domain are possibly different. Subsequent, to the analysis of subontology detection, the similarity computation is handled. The degree, to which the sub-ontology covers the knowledge scope compared with the ontology, is estimated by the cosine distance similarity between them. The larger the similarity between them is the greater knowledge space they cover. If the cosine distance value between two ontologies is 1.000, then they represent the same semantic knowledge and vice versa for 0.000 similarity value. If the distance similarity between two ontologies is larger than 0.000 but less than 1.000, then the partial semantic knowledge that they carry are overlapped.

3.8. Retrieval using rules metric

Once, the degree of similarity component is estimated using Section 3.7.2, the final stage of retrieving and providing the related documents to the authenticated cloud users is easier. In this module, the threshold value is estimated which helps in restricting the number of related documents to be retrieved. In this paper, the threshold value is determined to be 0.85 resulted after the execution of various experiments on benchmark medical datasets. This similarity value is very high, since this paper deal with the domain ontology construction. In this paper, medical ontology construction is discussed and experimented. Subsequently, simple if-then rules are employed to determine the documents to be retrieved and sent back to the authenticated cloud users. These cloud users can then employ the obtained documents to design a website for their own company, or construct a domain dictionary viz. a medical dictionary, or information extraction from a set of obtained related documents in order to prepare a consolidated information rich documents. The pseudo code for this module is given below.

```
Pseudo code: Retrieval
Inputs:
Ontology A (called as base ontology) – constructed
  ontology for the user's input document
Ontology B (called as repository ontology) – constructed
  ontology's from the already existing other
  text documents (other users).
Output:
Document metadata of the retrieved relevant documents
Algorithm:
N = Number of Ontology present in the repository
C = Counter
P= 0 (index of relevant documents)
Q=0 (index of irrelevant documents)
Ontology A = Input ontology of authenticated cloud
  user document
Ontology B = Repository ontology of other existing
  documents (other authenticated cloud users)
B = 1, 2, 3..... N
For Loop C= 1 to N where N is the number of Ontology
  present in the repository
If
The similarity value of Ontology A and Ontology B is
  between 0.85 and 1.00, then the metadata of document
  B corresponding to Ontology B is returned and relevancy
  index incremented;
P= P+1;
Else
Metadata of the irrelevant documents are not returned
  and irrelevancy index incremented;
Q=Q+1;
Increment Counter C =C+1
End Loop
```

4. Performance evaluation and result discussions

4.1. Experiment methodology

The proposed framework is tested for various domain ontologies available in the UCI repository [http://archive.ics.uci.edu/ml/]. The framework is tested for medical domain initially. However, this automated framework can be utilized for other major domains like education, business, marketing, military and other applications. In the repository, the underlying text documents are pre-processed to convert the statements into a suitable format [19,20]. Once an expressive ontology is produced, it is compared with ontology taken from repository in order to determine the reuse measure of ontologies. The ontological elements of diabetes can be reused by some other

Table 1
Ontology Comparison –UML-GM, GDR-DL, GDR-DEOL and GDR-DYDL (PROPOSED).

Dataset	No. of classes (NOC)				No. of instance examples (NOE)				No. of axioms (NOA)				No. of path links (NOP)			
	UML-GM	GDR-DL	GDR-DEOL	GDR-DYDL	UML-GM	GDR-DL	GDR-DEOL	GDR-DYDL	UML-GM	GDR-DL	GDR-DEOL	GDR-DYDL	UML-GM	GDR-DL	GDR-DEOL	GDR-DYDL
BC	286	302	330	342	9	9	12	15	250	280	283	285	128	130	132	138
BT	106	170	210	228	10	10	15	18	125	150	157	159	54	55	57	65
CT	212	249	270	283	23	23	27	32	230	260	280	283	108	110	112	122
DT	102	196	260	271	20	20	25	28	140	140	253	260	57	59	60	70
HD	303	415	450	466	3	3	7	11	219	396	467	469	164	166	170	175
IR	150	272	300	318	4	4	8	13	120	302	293	297	79	80	82	93

ontologies. This is possible by determining the similarity of diabetes with the other ontologies such as Breast Cancer (BC), Breast Tissue (BT), Cardiotocography (CT), Heart Disease (HD), Iris (IR) etc.

4.2. Stability measurement

The GDRs obtained for the given text document are said to stable when the issues of cyclic inheritance and non-direct relationships due to transitive property resolution as discussed in Section 3.1 [17]. Stability is determined by combining the integration and treatment of the GDRs. Integration (I)of GDRs can be done by employing the following equation:

$$G_o = \sum_{i=1}^{n} G_{\alpha i} \tag{9}$$

Table 1 provides Ontology measurement values for stability estimation based on UML-GM (Unified Modeling Language- Graphical Model), GDR-DL (Graph Derivation Representation-Description Logic)using DL (Description Logic) , GDR-DEOL (Graph Derivation Representation-DEOntic Logic) and GDR-DYDL (Graph Derivation Representation-DYaDic Logic)using Dyadic deontic Logic (Proposed). The analysis of the Table 1, is that the GDRs generated using Dyadic deontic logic produces more number of classes, instance examples, axioms and path links compared to the other three models namely UML-GM (Unified Modeling Language- Graphical Model), GDR-DL (Graph Derivation Representation-Description Logic) using DL and GDR-DEOL Graph Derivation Representation-DEOntic Logic. The reason for producing more number of classes is that, since dyadic deontic logic is highly expressive in nature which considers not only the dominant words and non-dominant words but also the conditional dependence events occurring in the document also. If the input dataset contains more non-dominant words and conditional dependence events, the other models as per the literature survey cannot produce increased numbers of concepts, instance examples, axioms and path links. Therefore, the proposed automated framework consisting of GDR using Dyadic deontic Logic are useful for determining stable and semantic rich ontologies.

The corresponding graphical result of Table 1 is shown in Fig. 2. The above graph represents the expressivity and stability of the exemplar ontologies based on the results computed using UML-GM, GDR-DL, GDR-DEOL and the proposed GDR-DYDL. From the graphical results, it is evident that GDR-DYDL provides the maximum expressivity computed based on the number of classes, the number of instance examples, number of axioms and the number of meaningful path links in the target ontologies [18].

4.3. Degree of similarity measure

The ontological elements of diabetes can be reused by some other domain ontologies if any. This is possible by determining the similarity of diabetes with the other ontologies such as Breast Cancer (BC), Breast

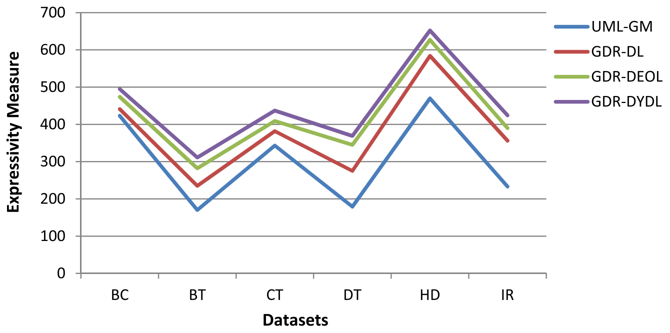

Fig. 2. Performance evaluation – stability measurement.

Tissue (BT), Cardiotocography (CT), Heart Disease (HD), Iris (IR) etc. The degree of similarity is estimated as described in Section 3.6.Table 2 determines the similarity results computed for various ontologies. If the distance similarity between two ontologies is zero (i.e., 0.000), then they represent same semantic knowledge (Fig. 3).

The above graph shows the similarity measures computed using the cosine similarity measure. From the graphical results it is evident that the proposed framework using dyadic deontic logic for semantic stability, expressive measurement and the degree of similarity computation has achieved better results for the target medical diabetes dataset.

5. Conclusion

The problem of constructing semantically stable ontology can be generated using the technique of removing the polymorphic objects, the degree of similarity computation between two ontologies and retrieval of highly relevant documents for various purposes like information retrieval, domain dictionary construction, information extraction has always been very challenging issues. The proposed combo ontology framework for generating the semantically stable ontology, computing the expressivity using statistics of the ontology and the degree of similarity computation in this paper utilizes a highly expressive knowledge representation language called dyadic deontic logic. On applying such logic for processing input dataset, the implicit and conditional dependence knowledge is also identified in addition to the explicit knowledge. Moreover, the extent of similarity is also addressed in this paper using the cosine distance similarity measure. The relevancy and irrelevancy is checked against simple if-then rules generation. However, in the further research this has been extended to make use of fuzzy rules rather than ordinary simple if-then rules. The further extensions can also be done in the work of expressivity and reusability which shall focus on using a different logic based knowledge representation language for even heterogeneous datasets.

Table 2.
Ontology similarity comparison based on GDR-DYDL.

Dataset	BC	BT	CT	DT	HD	IR
BC	1	0.950	0.528	0.307	0.287	0
BT	0.950	1	0.637	0.208	0.302	0
CT	0.528	0.457	1	0.328	0.748	0
DT	0.307	0.208	0.328	1	0.328	0
HD	0.287	0.302	0.748	0.328	1	0
IR	0	0	0	0	0	1

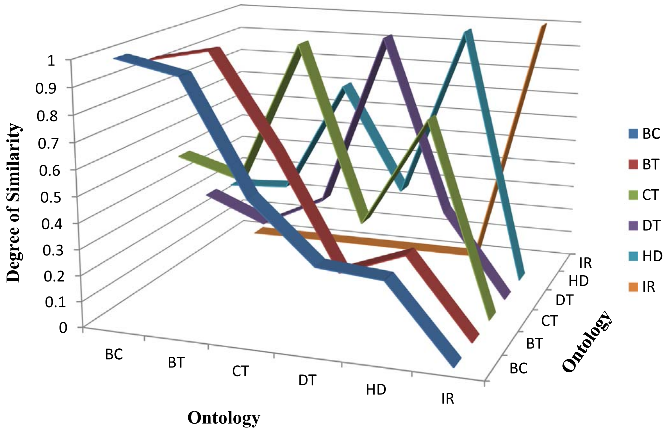

Fig. 3. Performance evaluation – degree of similarity measure.

References

[1] Fensel D. Ontology-based knowledge management. IEEE Comput 2002;35(11):56–9, (November).
[2] Deborah LJegatha, Baskaran R, Kannan A. Deontic logic based ontology alignment technique for E-learning. Int J Intell Inf Technol 2012;8(3):56–78.
[3] Chen L, Shadbolt NR, Goble CA. A semantic web-based approach to knowledge management for grid applications. IEEE Trans Knowl Data Eng 2007;19(2):283–96, (February).
[4] Razmerita L. An ontology-based framework for modelling user behavior-A case study in knowledge management. IEEE Trans Syst, Man Cybern A Syst, Hum 2011;41(4):772–83, (July).
[5] Shadbolt N, Berners-Lee T, Hall W. The semantic web revisited. IEEE Intell Syst 2006;21(3):96–101, (January/February).
[6] Philippi S, Kohler J. Using XML technology for the ontology-based semantic integration of life science databases. IEEE Trans Info Tech Biomed 2004;8(2):154–60, (June).
[7] S. Kraines, W. Guo, B. Kemper, Y. Nakamura, EKOSS: A knowledge-user centered to knowledge sharing, discovery, and integration on the semantic Web. In Proceedings of the 5th ISWC, Athens, GA, USA; 2006:833–846.
[8] D. Vallet, M. Fernandez, P. Castells, An ontology-based information retrieval model. In Proceedings of the 2nd ESWC, Heraklion, Greece; 2005: 455–470.
[9] Maguitman A, Menczer F, Roinestad H, Vespignani A. Algorithmic detection of semantic similarity, In Proceedings of the 14th International Conference WWW, pp. 107–116; 2005.
[10] Deborah LJegatha, Karthika R, Audithan S, Bala BKiran, Enhanced expressivity using deontic logic and reuse measure of ontologies. In Proceedings of the eleventh international conference on data mining and warehousing, Elsevier – Procedia Computer Science; 2015:84:318–326.
[11] Qu Y, Cheng G. Falcons concept search: a practical search engine for web ontologies. IEEE Trans Syst Man Cybern A Syst Hum 2011;41(4):810–6, (July).
[12] Maedche A, Staab S. Measuring similarity between ontologies. In Proceedings of the 13th International Conference EKAW, Sigüenza, Spain; 2002: 251–263.
[13] Popescu M, Keller JM, Mitchell JA. Fuzzy measures on the gene ontology for gene product similarity. IEEE/ACMTrans Comput Bio Bioinfo 2006;3(3):263–74, (July/September).
[14] Al-Mubaid H, Nguyen H. Measuring semantic similarity between biomedical concepts within multiple ontologies. IEEE Trans Syst, Man Cybern C Appl Rev 2009;39(4):389–98, (July).
[15] Sanchez D, Batet M, Isern D, Valls A. Ontology-based semantic similarity: a new feature-based approach. Expert Syst Appl 2012;39(9):7718–28.
[16] Rodriguez A, Egenhofer M. Determining semantic similarity among entity classes from different ontologies. IEEE Trans Knowl Data Eng 2003;15(2):442–56, (March/April).
[17] Tartir S, Arpinar IB, Moore M, Sheth AP, Aleman-Meza B. OntoQA: Metric-based ontology quality analysis. In Proceedings IEEE ICDM workshop KADASH; 2005.
[18] Gangemi A, Catenacci C, Ciaramita M. and Jos. Lehmann, A theoretical framework for ontology evaluation and validation. InProceedings SWAP; 2005.
[19] Lozano-Tello A, Gomez-Perez A. OntoMetric: a method to choose the appropriate ontology. J Database Manag 2004;15(2):1–18.
[20] Wache H, Vogele T, Visser U, Stuckenschmidt H, Schuster G,Neumann H, Hubner S. Ontology-based integration of information-A survey of existing approaches. In Proceedings of the international joint conference on artificial intelligence workshop ontologiesand information sharing,pp. 108–117; 2001.
[21] Giunchiglia F, Yatskevich M, Shvaiko P. Semantic matching: algorithms and implementation. J Data Semantics 2007;9:1–38.
[22] Nagy M, Vargas-Vera M. Multiagent ontology mapping framework for the semantic web. IEEE Trans Syst Man Cybern A Syst Hum 2011;41(4):693–704, (July).
[23] Aha D. Tolerating noisy, irrelevant and novel attributes in instance-based learning algorithms. Int J Man-Mach Stud 1992;36:267–87.
[24] Allan J. Relevance feedback with too much data. In Proceedings of the ACM SIGIR Conference on Research and Development inInformation Retrieval, (Seattle, Washington, USA), pp. 337–343; 1995.
[25] Aone C, Bennett SW, Gorlinsky J. Multi-media fusion through application of machine learning and NLP. In Proceedings of the AAAI Symposium on Machine Learning in Information Access, (Stanford, CA, USA); 1996.
[26] Apte C, Dameru F, Weiss SM. Automated learning of decision rules for text categorization. ACM Trans Inf Syst (TOIS) 1994;12(3):233–50.
[27] Balabanovic M, Shoham Y, Yun Y. An adaptive agent for automated web browsing. Technical Report No. SIDL-WP-1995–0023. California: University of Stanford; 1995.