Informatics in Medicine Unlocked 8 (2017) 66–73

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
Available online 18 December 2016

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.