Cluster Comput (2015) 18:351–367
DOI 10.1007/s10586-014-0406-8

Investigating an ontology-based approach for Big Data analysis of inter-dependent medical and oral health conditions

Tejal Shah · Fethi Rabhi · Pradeep Ray

Received: 4 September 2014 / Revised: 3 October 2014 / Accepted: 7 October 2014 / Published online: 20 November 2014

Abstract The volume, velocity and variety of data generated today require special techniques and technologies for analysis and inferencing. These challenges are significantly pronounced within healthcare where data is being generated exponentially from biomedical research and electronic patient records. Moreover, with the increasing importance on holistic care, it has become vital to analyse information from all the domains that affect patient health, such as medical and oral conditions. A lot of medical and oral conditions are interdependent and call for collaborative management; however, technical issues such as heterogeneous data collection and storage formats, limited sharing of patient information, and lack of decision support over the shared information among others have seriously limited collaborative patient care. To address the above issues, the following research investigates the development and application of ontology and rules to build an evidence-based, reusable and cross-domain knowledge base. An example implementation of the knowledge base in Protégé is also done to evaluate the effectiveness of the approach for reasoning and decision support of cross-domain patient information.

Keywords Big Data for health · Ontology · OWL · SWRL

T. Shah (✉) · F. Rabhi
School of CSE, The University of New South Wales,
Sydney, NSW, Australia
e-mail: tejals@cse.unsw.edu.au

T. Shah · F. Rabhi · P. Ray
APuHC, The University of New South Wales,
Sydney, NSW, Australia

F. Rabhi
e-mail: f.rabhi@unsw.edu.au

P. Ray
e-mail: p.ray@unsw.edu.au

1 Introduction

The advent of Internet, the World Wide Web, Cloud computing [1–3], has led to data explosion in almost all areas of life. Moreover, this data is complex, heterogeneous, and generated at a rapid rate leading to what is commonly dubbed as the 3Vs of Big Data namely, volume, variety and velocity of data [4]. To handle such data, e-Science [5,6] infrastructure has emerged especially in scientific domains. These characteristics of data, if exploited timely and appropriately, can bring much value in the form of cost savings, improved decision-making and better productivity in diverse fields such as healthcare, finance, commerce, education, national security, emergency management, weather forecasting and so on [4]. However, manual analysis of such complex data is challenging and prohibitive leading to loss of value that could have been derived from the information held within this data. Researchers are therefore increasingly focused on finding ways to handle the data available today. New and efficient approaches are becoming necessary to create, store, manage, access, process, and share the information available.

This fact is especially pronounced in healthcare where the use of electronic health records can generate huge volumes of high velocity data that must often be analysed in real time to make patient-related decisions such as diagnosis, treatment plans, medication prescription, etc. Moreover, with an increasing focus on holistic care approach, healthcare practitioners require patient data from across different health domains to make informed decisions. One such case is between the medical and dental (oral health)[1] domains. Research has repeatedly shown strong associations between medical and oral health conditions and has stressed

[1] For the purpose of this paper, we refer to the terms 'dental', 'dentistry' and 'oral health' interchangeably.

on analysing patient information from both domains while making diagnostic and treatment decisions. However, there is very little technological support that can provide the appropriate computing environment to analyse the data and provide decision support over shared and inter-dependent knowledge from both the medical and oral health domains. For instance, most state-of-the-art systems today concentrate on sharing of information i.e., making patients' medical and dental information available to the respective practitioners but we envision a scenario where the information is seamlessly integrated while retaining the semantics. Consequently, the information can be reasoned over for deriving practical benefits such as reusability, and decision support capabilities including alerts, recommendations, reminders, and explanations. In the absence of efficient and automated analysis, practitioners may fail to exploit the wealth of knowledge that is contained within the information. Therefore, they may overlook the information or may not be able to foresee risks that may arise from interactions of existing conditions. Therefore, it is important to develop techniques and technologies to handle the complexity of Big Data so that usable information can be derived from it in a relevant, timely and accurate manner.

In this paper, which is an extended version of our previously published conference paper [7], we have presented an ontology-based approach to reason over the complex medical and oral health data, draw inferences, and generate new knowledge from it. We envision that this knowledge can be further used to provide better and a more comprehensive healthcare to patients. Rest of the paper is organised as follows. In Sect. 2 we discuss the relation of Big Data and health, specifically how Big Data brings both challenges and advantages for healthcare. In Sect. 3 we discuss the related work followed by a brief description of the semantic web technologies in Sect. 4. Further, in Sect. 5 we present our approach including the ontology development process and examples to explain the advantages of ontology-based Big Data analysis. In Sect. 6 we show where and how ontology fits in the big picture of Big Data analysis and conclude the paper in Sect. 7.

2 Big Data and health

Big Data is defined by Gartner as "high-volume, high-velocity, and/or high-variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization" [8]. In addition to the 3 Vs (volume, velocity, variety), complexity adds to the challenges of Big Data making manual analysis impossible. Therefore, computer aided processing is essential if the vast knowledge that is contained but implicit within this data is to be obtained and used. Accordingly, the term 'Big Data' also encompasses the techniques and technologies that are used to handle the data [4].

In the health sector, the above Big Data characteristics are especially pronounced. For instance, Fig. 1 is a heat map from the Gartner report of July 2012 [4] that shows the various sectors that have the potential to benefit from the different aspects of Big Data. It is evident from the heat map that in healthcare the variety and velocity of data generated is significant and hence the potential to exploit the richness of such data is high [9]. The advent of electronic patient records and biomedical research including genetics research has led to an explosion of the volume of data generated at a very high velocity and which is further required to be analysed in real time. Moreover, health data can be found in various forms including structured (relational databases), semi-structured (XML, other markup languages) unstructured data (clinician notes), and annotations (images). For instance, patient data can now be available to the healthcare professional from various sources such as electronic health records, electronic medical records, electronic dental records, pharmaceutical records, digital images, laboratory tests, and so on. Further, the information contained within them is in different formats and more so from different domains. Thus, the quantity, complexity and heterogeneity of the data bring the potential to discover new knowledge that can improve work practices and produce better outcomes. At the same time, there also arise significant challenges such as accessing, processing, analysing, and distributing the data.

In addition to these inherent challenges is the issue of information silos, which are commonplace in the healthcare sector due to the absence of technology that can seamlessly integrate and distribute the data [10,11]. Such silos also fragment the medical and oral health domains although research has categorically shown that various medical and oral heath problems are related and must be managed collaboratively. For example, as shown in [12], decreased metabolic control in diabetes mellitus type 2 has a negative impact on the periodontal health (which refers to the health of the supporting structures of a tooth) of the patients and aggravates pre-existing periodontitis (which refers to the health of the supporting structures of a tooth). Periodontitis in turn adversely affects the prognosis of other medical conditions such as respiratory and cardiovascular diseases, stress, can affect the immune system leading to conditions associated with immune suppression, cause nutritional compromise and much more. In fact, over the years, research has established definite links between several medical and oral health conditions and in doing so have stressed the importance of a dental professional's role in early diagnosis or influencing the prognosis of medical conditions in a patient. Early diagnosis especially becomes important when considering that more than 120 medical conditions manifest first in the oral cavity [13].

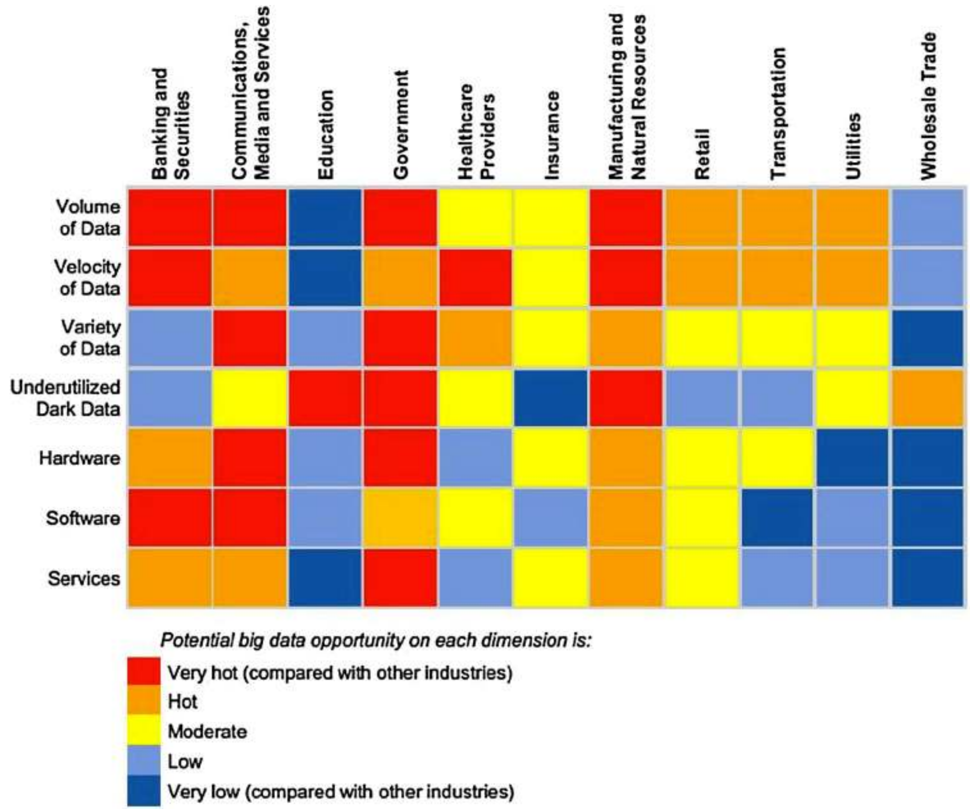


Fig. 1 Big Data heat map

The complexity and variety of Big Data coupled with the information silos in healthcare especially in the medical and oral health domains call for suitable technologies that can help to process, integrate and seamlessly share information among the healthcare providers and achieve continuity of care for the patients. In this direction, semantic web technologies such as ontologies, are being increasingly developed and employed to bridge and integrate diverse and heterogeneous data sources at a semantic level resulting in semantic interoperability. Semantic interoperability is invaluable in situations where information from diverse domains and multiple disciplines need to be analysed such as between medicine and oral health and their subspecialties [14]. The challenges brought forth by Big Data in health have thus stimulated the innovation of techniques and technologies that lead to 'right care', which includes physician communication, clinical decision support, and disease management [10]. The research work and innovation presented in this paper is similarly aimed at 'right care' as represented by physician-dentist communication, clinical decision making in the combined area of oral and systemic healthcare and managing conditions arising from the combined effects of oral and systemic conditions respectively. Specifically, the research investigates the use of semantic web technologies for the development of a novel cross-domain knowledge base that consists of scientifically proven associations between the medical and oral health conditions. The knowledge base helps in Big Data analysis as it can integrate and reason over a large volume of heterogeneous and complex information, perform automated decision support tasks and provide explanations for the outputs. We envision that such a semantic knowledge base can thus help to derive value from Big Data by discovering new and actionable knowledge from it. The contributions of this paper are threefold (a) developing a comprehensive cross-domain knowledge base that is generic enough to be reused by various health decision support applications, (b) presenting an approach to achieve cross-domain communication between two theoretically inter-dependent but practically separate healthcare domains, and (c) demonstrating the application of Semantic Web technologies to access, process, analyse, and share Big Data in health.

3 Related work

In healthcare, the need for managing, integrating and analysing data from various sources to help informed decision-making and improve patient outcomes has been long accepted and efforts in that direction have been consistently made. Specifically with respect to medical and dental information, large-scale systems in the United States such as VistA, CattailsMDTM, and the Indian Health Service Health Information System have enabled sharing of patient's medical and dental information [13]. There is however no support provided to analyse the shared information and make explicit any implicit knowledge that may result from the interactions of the medical and dental conditions. The practical usability of the information therefore remains limited since how the shared information is used becomes subjective and rests with the individual healthcare professionals who access such information. The analysis and decision support, even if desired, requires semantic knowledge bases that contain information about both medical and dental domains and modeled such that implicit knowledge can be extracted and put to use. In that direction, ontologies are very expressive semantic web technologies and allow for rich modelling of information, which can be reasoned over to derive new knowledge.

The biomedical ontologies available today that are most relevant to medicine and oral health are the Systematized Nomenclature of Medicine—Clinical Terms (SNOMED-CT) [15], Systematized Nomenclature of Dentistry (SNODENT) [16], and International Classification of Diseases (ICD) [17]. Of these, SNOMED-CT and ICD involve concepts from both medical and dental domains while SNODENT is specific to the latter. SNOMED-CT is the largest and most widely used biomedical ontology [18,19]; however, it is primarily an ontology for representation rather than for analyzing and decision making [8]. In fact, the structure of SNOMED-CT is such that reasoning and inferencing over factual data could be error-prone and problematic [20,21]. Most importantly our analysis, in agreement with past analysis on SNOMED and SNODENT [22], concluded that it did not contain all the terms and relationships that are required to model the inter-dependent conditions across the medical and oral health domains. SNODENT is a subset of SNOMED specific to the dental domain and accordingly also contains many of the dental concepts that SNOMED does not. However like SNOMED, SNODENT too is a representation ontology with limited expressivity in terms of Description Logic which restricts its ability to perform inferencing for complex analysis and decision making over large amounts of heterogeneous data. More expressive Description Logic based ontologies can be now modelled to allow inferencing and decision support tasks over heterogeneous data. These ontologies are represented in the Web Ontology Language (OWL), which is a very expressive Description Logic and will be discussed in detail in Sect. 4.2.

Researchers and developers have accordingly employed OWL based ontologies and semantic rules to address semantic interoperability challenges in healthcare [23–29] and to guide healthcare professionals in following clinical guidelines [30]. However, the use of OWL ontology and semantic rules to semantically integrate data from various sources have been restricted to local organizations or within a specific medical domain such as for acute cardiac disorders in [31] and diabetes medication in [32]. To the best of our knowledge, there is no OWL-based ontology enriched with semantic rules to manage and analyse Big Data across two different healthcare domains. Our approach of developing a formal cross-domain, evidence-based knowledge base that can infer over the shared medical and dental information to obtain practical and usable knowledge from the interdependencies of the medical and dental conditions is therefore novel. We have thus shown how information from Big Data can be used to provide a more holistic care to patients.

Health data in general and medical data in particular is commonly stored in databases, and institutions are often reluctant to completely migrate from the databases to ontology based information systems. This is true of other areas as well such as commerce and finance. Accordingly, researchers have developed an approach known as ontology based data access (OBDA) so that the rich semantic expressions of the ontology can be used to query the data stored in various data sources such as databases [33–36]. It is often the intended purpose that the querying can also be done by end-users who need not be programmers or database specialists and may not know how the data is stored and/or structured. One such major project is the Optique project [37], which employs OBDA for visual query formulation such that the end-users, who need not be programmers or database specialists, can also query the database. Mappings are done to connect the ontology to database and in this manner further integration with data from external databases can also be achieved. Thus ontologies are being increasingly used to access, manage and analyse data from disparate sources. The ontology presented in this paper can also be mapped, if and when required, to database(s) for ontology based data access. Unlike databases, which are very specific, ontologies can be used in a generic manner and mapped to other ontologies so as to access several data sources without compromising on the semantics of the data. This is one of the many advantages that formal ontologies bring to Big Data analysis.

4 Semantic web technologies

Ontologies and rules are commonly considered as the 'foundations of knowledge bases'. In addition to ontology and rules, Description Logic reasoners are required to discover implicit information and infer new knowledge. In this section, we discuss these three important semantic web technologies.

4.1 Ontology

Formal ontologies are based on Description Logics, which are knowledge representation formalisms [38]. Studer et al.

[39] define ontology as: "a formal, explicit specification of a shared conceptualization". An ontology ensures retention of meaning and accuracy of the information exchanged since it formally defines the concepts and their relationships so as to remove any heterogeneity and allow for semantic interoperability between different systems [40]. Besides, as opposed to terminologies and classification systems, which are static structures for knowledge reference, ontologies allow domain knowledge reference, reuse and reasoning [41].

4.2 Web Ontology Language

Web Ontology Language version 2 (OWL 2), which is a Web Standard, is an expressive ontology representation language based in Description Logic (DL) for describing the semantics of knowledge [42,43]. However, OWL 2 is not decidable in its full form; therefore, a subset OWL 2 DL, is used for reasoning tasks and to take advantage of the various reasoners available [44]. OWL 2 DL is based on the DL $\mathcal{SROIQ}$, which is decidable [45]. $\mathcal{SROIQ}$ is a very expressive DL and consequently so is OWL 2 DL. However, it also has less expressive profiles namely OWL 2 EL, OWL 2 RL and OWL 2 QL [44]. An OWL 2 ontology primarily consists of [44] (i) axioms—the basic statements in an OWL ontology, (ii) entities—the terms used for representing real world objects, and (iii) expressions—the complex descriptions derived from the combinations of various entities. An OWL ontology is a DL knowledge base, which is typically made up of a Terminological Box or TBox and an Assertional Box or ABox [46]. $\mathcal{SROIQ}$ DL also provides a Role Box or RBox [45,47]. The TBox captures the intensional knowledge of a domain and models it as concepts and binary relations between those concepts, while the ABox contains the extensional knowledge in the form of asserted facts i.e., the ground knowledge about individuals with respect to the TBox, and the RBox describes the role characteristics.

4.3 Semantic Web Rule Language

For the purpose of retaining decidability and classifying in polynomial time, there are several restrictions employed on OWL 2 thereby limiting its expressivity. For example, OWL 2 cannot express the relation *child of married parents* [48], which is basically a relation between individuals with which another individual is related. For such purposes, rules can be used to enhance the expressivity of the underlying ontology language. Further, the rules provide actionable knowledge so that it is possible to develop decision support tasks in the form of alerts, reminders, recommendations, guidelines and diagnosis. However, in order to maintain semantic compatibility of the rules with the ontology, the rule language must be semantically compatible with OWL. The W3C proposal, Semantic Web Rule Language (SWRL) [49], provides a Horn clause rules extension to OWL in a semantically coherent manner. The basic structure of SWRL rule is of the form $antecedent \rightarrow consequent$ that is, if the antecedent or body of the rule is true then it is implied that the consequent or head is true as well and holds. The antecedent and consequent consist of a conjunction of atoms in the form $\mathsf{a}_1 \wedge \cdots \wedge \mathsf{a}_n$. However, a combination of SWRL rules and OWL 2 DL can lead to undecidability; therefore a subset of SWRL rules known as DL-safe rules was developed where the binding of variables is restricted to only known individuals so as to ensure decidability of the resulting rules [50]. Moreover, limiting the rule atoms to named classes and properties within the base OWL ontology also ensures interoperability of the ontological knowledge embedded within the rules with other OWL ontologies, which may or may not support SWRL [49]. Such restrictions also facilitate translation of SWRL rules to other rule systems such as Prolog, production rules and SQL and also improve tractability of the reasoning tasks that are performed over the rules [49]. In this format thus, the previously mentioned relation *child of married parents* can be expressed as:

$$\mathsf{Person(?x)} \wedge \mathsf{hasParent(?x,\ ?y)} \wedge \mathsf{hasParent(?x,\ ?z)} \wedge \mathsf{hasSpouse(?y,\ ?z)} \rightarrow \mathsf{ChildOfMarriedParents(?x)}$$

where Person and $\mathsf{ChildOfMarriedParents}$ are named classes in the underlying OWL ontology; $\mathsf{hasParent}$ and $\mathsf{hasSpouse}$ are named properties; and ?x, ?y, ?z are variables. The rule states that a person whose parents are married is essentially a child of married parents.

4.4 Reasoners

The OWL ontology and SWRL rules together form a knowledge base for a specific domain. This knowledge base consists of implicit information, which can be extracted using DL reasoners or inference engines. Some of the well-known reasoners include Pellet [51], Hermit [52], Fact++ [53], Kaon2 [54], and RacerPro [55] among others. These reasoners for OWL are considerably mature and provide sound and complete inference services [56]. We have selected Pellet for our work, which is based in Java and is available as open source. In addition to being a very efficient reasoner, the newer versions of Pellet provide native support for SWRL, albeit limited, thereby combining the knowledge for the reasoning process to provide more accurate and comprehensive results. Pellet also provides an explanation facility, which justifies the inferencing result by showing the pathways that were used to reach the specific decision.

With the Semantic Web technologies gaining increased maturity and with the increasing need for health systems

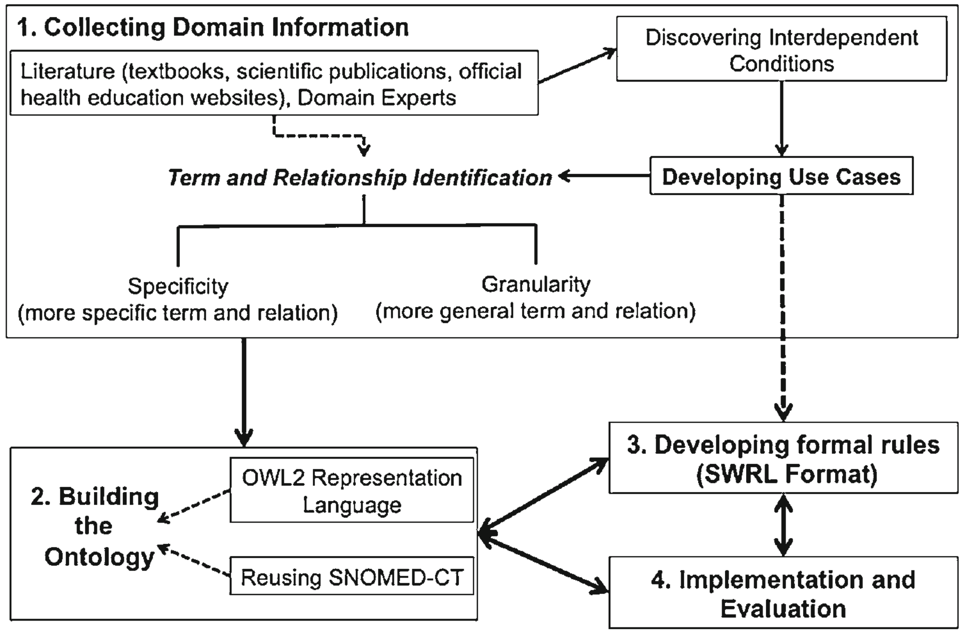


Fig. 2 Developmental framework

to be interoperable, several biomedical ontologies such as SNOMED-CT, Foundational Model of Anatomy (FMA), Medical Subject Headings (MeSH) and National Cancer Institute (NCI) Thesaurus have been converted to OWL [57]. However, none of these ontologies use SWRL rules to extend the knowledge represented in the ontology and form a comprehensive knowledge base representing their respective domains. In the next section, we discuss how we employ the above technologies and develop the knowledge base for Big Data analysis in the medical and dental domains.

5 Approach

In this section, we explain the approach adopted by us in developing the cross-domain knowledge base. We also demonstrate how this knowledge base is applied to reason over complex and varied cross-domain data that is generated from disparate data sources. Figure 2 shows our developmental framework, which involves four basic steps starting with discovering and collecting information of medical and dental conditions via use cases. The same use cases are then used to identify relevant terms for the ontology and associations between them. The next step involves building the formal ontology from the information collected in the previous step followed by developing formal rules. The rules are vital to the goal of achieving decision support over the medical and dental conditions. Finally, the formal ontology and rules developed in the previous step are implemented and evaluated. As part of the evaluation process, the use cases from step 1 are reused to evaluate the knowledge base for accuracy, consistency and comprehensiveness. The above steps are discussed in detail below.

5.1 Step 1: collecting domain information

The first task involved scoping that is, identifying the domain and its boundaries for representation. The domain for this research lays at the intersection of the medical and oral health domains and represents the inter-dependent conditions from both the domains. These conditions were obtained from various sources including scientific literature and domain experts' knowledge. Three domain experts were involved in our development process—a general practitioner and two dental surgeons. They were consulted to verify the correctness of the conditions, associations and the rules formed from them. The conditions were developed into use cases, which then formed our reference for discovering the terms to be modelled. The relationships between the terms were converted into object properties within the formal ontology in the later stages. Table 1 lists some example use cases that we extracted from literature.

The use cases are a combination of complex information since they are derived from several subspecialties within medicine and dentistry and as a result the analysis of interdependent conditions becomes challenging and manually impossible. Figure 3 shows the number of medicine and dental subspecialties involved in each of the above use cases. The subspecialties were considered as per the classification provided by the Medical and Dental Boards of Australia respectively [77,78].

Table 1 Example use cases of inter-dependent Medical and Oral Health conditions

	Use cases (made succinct here)
1	Candidiasis and Oral Hairy Leukoplakia are early indicators of the presence of Human Immunodeficiency Virus (HIV) [58,59]. Both the former conditions occur in the oral cavity and this puts the dental practitioner at a very important position where he/she can diagnose underlying systemic conditions in the early stages
2	The presence of periodontitis in pregnant women has been associated with the birth of low birth weight infants [60]. Therefore, maintenance of good oral hygiene or providing periodontal treatment is essential during pregnancy
3	The progress of Diabetes Mellitus (DM) is adversely affected by periodontal disease. Conversely, poorly controlled DM exacerbates periodontal disease [61]. Therefore, a patient with either of these conditions must be managed collaboratively by the medical and oral health practitioners
4	An untreated periodontal abscess can lead to the development of Ludwig's Angina, which if left untreated, can cause fatal complications such as asphyxia [62]. Therefore, it becomes important for the dental practitioner to understand the systemic implications of Ludwig's Angina, which is essentially an oral condition
5	If a patient with any form of congenital heart disease (CHD) and poor oral hygiene undergoes surgical dental extraction, then the resulting transient bacteraemia will most likely react with the underlying CHD and put the patient at risk of a bacterial endocarditis. Therefore, the patient should be given antibiotic prophylaxis to prevent the occurrence of endocarditis in such cases [60,63]
6	The presence of oral mucosal papillomatosis, which affects the oral cavity, and acral keratosis in a patient is diagnostic of Cowden syndrome, which is a very complex medical condition that involves several body systems and correspondingly several sub-specialties [64,65]. This rule is based on the domain knowledge that oral papillomatosis is one of the main symptoms of Cowden syndrome [66]
7	Management of intensive care unit (ICU) patients with any form of intubation should involve a comprehensive oral hygiene program to prevent the occurrence of ventilator-associated pneumonia. Due to minimal or no involvement of dentists in managing intensive care patients, this guideline is often missed thereby endangering patient safety [67]
8	A patient receiving bisphosphonate therapy is at risk of developing bisphosphonate associated osteonecrosis of jaw (BON) if he/she undergoes an invasive oral procedure such as tooth extraction. BON is the sudden death of the bone of the oral cavity and could prove to be fatal [68]. Therefore, anyone undergoing bisphosphonate therapy should not undergo invasive oral procedure. In this case therefore, prevention is the key, which is possible only if the medical and dental carers of the patient have access to the relevant information and understanding of the interaction between the conditions
9	It is important that HIV patients with very low platelet count be treated conservatively by dentists to prevent excessive and uncontrolled bleeding [69,70]. Therefore, a dental practitioner must be aware of the patient's medical history and understand how it may affect the dental procedures to prevent any untoward incidences
10	Several oral conditions and procedures can lead to transient bacteraemia including dental extractions, periodontal procedures, endodontic instrumentation, dental prophylaxis, dental implant placement, initial placement of orthodontic bands and many more [63,71,72]. Therefore, appropriate precautions should be taken by dental practitioners for patients who may have an underlying systemic condition that could deteriorate as a result of bacteraemia
11	The oral radiology manifestation of osteoporosis includes thinning of mandibular cortex i.e., thinning of the bone of the lower jaw. Generalized osteoporosis is the radiographic manifestation of sickle cell anaemia in the jaws [73]
12	Hypophosphatasia is a metabolic bone disease and could be fatal. Delayed eruption of permanent dentition is often the first clinical sign of hypophosphatasia. The radiographic features of the jaw include—generalized radiolucency of the mandible and maxilla, thinning of the cortical bone and lamina dura and poor calcification of the alveolar bone. Associated radiographic changes of the teeth include—thin enamel layer, and large pulp chambers and root canals [73]
13	Niacin deficiency results in a debilitating systemic condition known as pellagra and the oral manifestations include glossitis and generalized stomatitis. The latter may be the earliest clinical signs of niacin deficiency while the most frequent oral finding is acute necrotizing ulcerative gingivitis [70,74]. Several organs including brain, skin and the intestine are affected in pellagra [75]
14	The oral changes in parathyroid hypersecretion (also known as osteitis fibrosa cystica or von Recklinghausen's bone disease), which arises due to hormonal disturbances include—malocclusion and tooth mobility. The oral radiographic changes include: alveolar osteoporosis, widening of periodontal space, absence of lamina dura, and presence of bone cysts [73]. The alertness and knowledge of a dental practitioner can thus help in early detection and treatment of the condition
15	Clenching of teeth or bruxism could be a manifestation of underlying psychological factors such as chronic stress [74,76], and this once again puts the dental practitioner in an ideal position of diagnosing underlying systemic issues based on the presence and intensity of oral conditions

Terms and relationships were then identified from the use cases, which were used to model classes and properties respectively in the final ontology. Example terms and relationships from the first five use cases are shown in Table 2. It would be apt to note the semantics of various languages at this stage. The word 'terms' refers to real-world concepts or

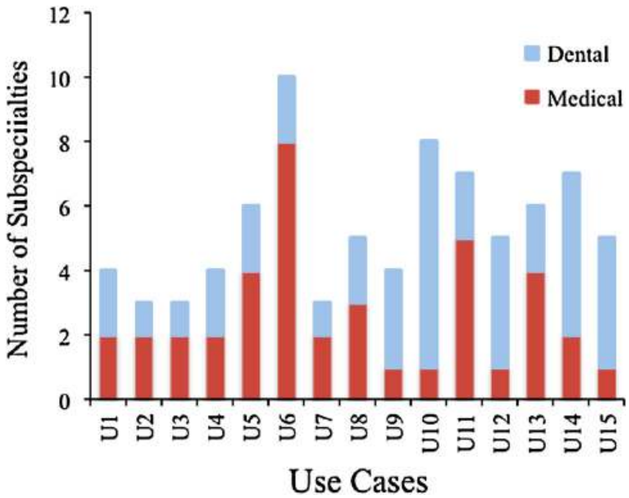


Fig. 3 Number of subspecialties involved in each use case

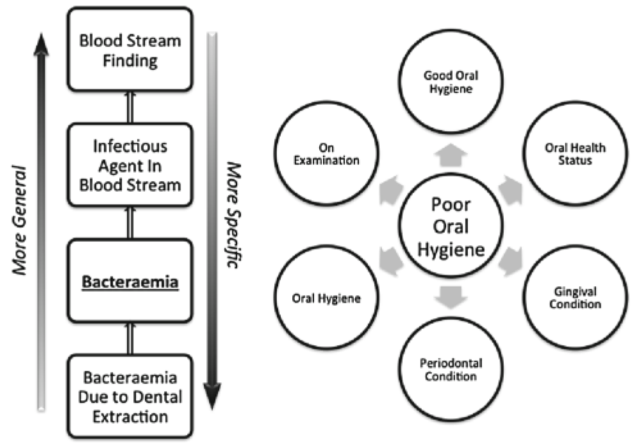


Fig. 4 Hierarchy showing more general and specific terms for 'bacteraemia' and neighbourhood terms for 'poor oral hygiene'

objects and the 'relationships' indicate how the concepts are related to each other. In OWL, these terms are modelled as classes and the relationships become properties whereas in Description Logic, the terms are referred to as concepts and relationships as roles [79]. In the discussion ahead, we will refer to classes-concepts and properties-roles interchangeably.

The identified terms were analysed for lexical and semantic similarities and differences. Each term was further refined to derive a more general and a more specific term. This way, more terms were discovered and an initial hierarchy obtained. Neighbourhood terms were further identified from the associations obtained from literature. The main research challenge in this step was to discover all possible terms to represent comprehensiveness of both the domains and at the same time, all the terms must be relevant to both the domains. This dual requirement ensured that only necessary terms were modelled in the ontology so as to prevent cluttering from unwanted and irrelevant information. Figure 4 shows the more specific and general terms for 'bacteraemia' and the neighbourhood terms for 'poor oral hygiene' from use case 5 above.

5.2 Step 2: building the cross-domain ontology

One of the OBO principles states that where possible, existing ontologies must be reused instead of building a new ontology from scratch [80]. This approach helps to restrict rapid and uncontrolled proliferation of ontologies and simultaneously stimulates reuse of the terms already available in the existing ontologies. Moreover, the reuse approach enables semantic interoperability across systems that do not use the same ontologies by making it easier to develop mappings and alignments between ontologies. Accordingly, we reused SNOMED-CT to build the ontology since it has the broadest coverage of concepts amongst all the biomedical ontologies available. We refer to our ontology as Oral-

Table 2 Terms and relationships identified from the use cases

Use case no.	Terms	Relationships
1	Pseudomembranous Candidiasis, Oral Hairy Leukoplakia, HIV Infection	Early Indicator, Diagnostic
2	Pregnant, Low Birth Weight Infant, Periodontitis, Periodontal Therapy, Good Oral Hygiene	Patient At Risk, To Maintain, Recommended Therapy, Preventive Measure
3	Diabetes Mellitus, Diabetes Mellitus Type 2, Periodontal Disease, Periodontitis	Affects, Influences, Interacts With
4	Periodontal Abscess, Untreated/No Treatment, Ludwig's Angina, Asphyxia	Causes, Leads To, Has Complication
5	Congenital Heart Disease, Poor Oral Hygiene, Surgical Extraction of Tooth, Bacteraemia, Bacterial Endocarditis, Antibiotic, Prophylaxis, Antibiotic Prophylaxis	Has Condition, Undergoes Procedure, Causes, Patient At Risk, Preventive Recommendation

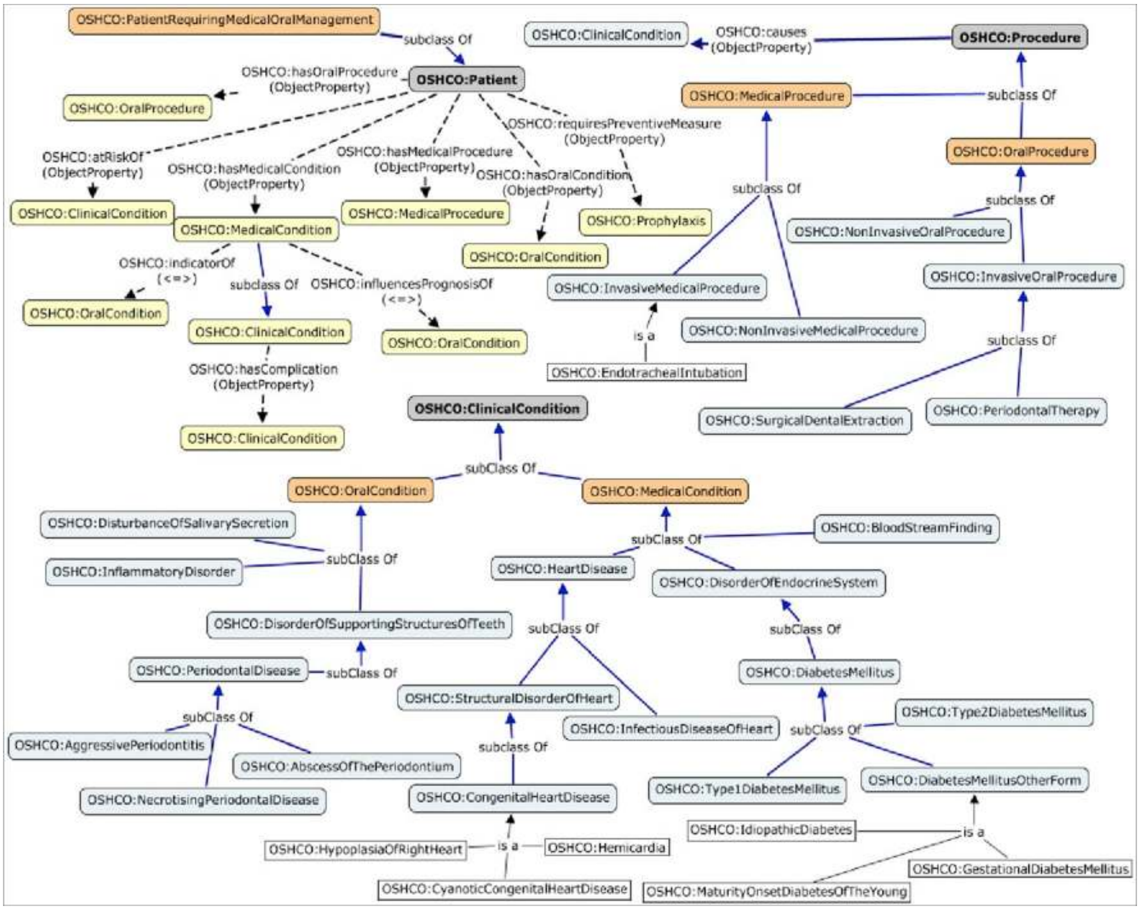


Fig. 5 Screenshot showing a section of OSHCO

Systemic[2] Cross-domain Ontology (OSHCO) as it models cross-domain medical-oral health knowledge.

In this step, the identified terms and relationships that formed the initial hierarchy in step 1 were matched with the corresponding concepts and properties in SNOMED-CT and the relevant hierarchical structure in SNOMED identified. However, not all the required terms and relations were available in SNOMED-CT. When that happened, a more general or specific term was identified and the hierarchy with the closest match to the context of the required term was selected and the term was added to it. The reader is referred to our other work [81] for a detailed description of reusing SNOMED-CT. As discussed previously, we selected OWL 2 as the representation language for OSHCO. This was done to ensure semantic interoperability between the systems using our ontology with other systems that use the corresponding Web Standards. A significant research challenge at this step was to ensure that in modelling OSHCO as closely as possible to SNOMED-CT, the structural and modelling pitfalls of the latter were not replicated in the resulting ontology. This is important because OSHCO being a cross-domain ontology contains a rich density of relationships to represent the various use cases correctly and in doing so it is extremely easy to convert into a heavy ontology with a large number of concepts and properties. This could make OSHCO practically inconvenient to run and reason over on local machines for real-time decision support tasks, which is one of the major issues with SNOMED-CT [82].

We used Protégé 4.2 [83], an open source ontology editor to build and validate OSHCO. Figure 5 is a screenshot depicting a section of the ontology including classes, subclasses, named individuals and object properties as built in Protégé and exported into CMap [84]. Three main classes namely patient, procedure and clinical condition with some of their subclasses, as well as few relationships (referred to as object properties in OWL) for the patient class can be seen. The properties have been modelled according to the use cases to connect the inter-dependent conditions thereby linking the medical and oral health domains within the ontology. Moreover, as we will discuss in the next Sect. 5.3, the relationships of patient class to the other classes help in deriving actionable knowledge from the asserted facts.

In OSHCO, we incorporated the metamodelling feature known as punning [85]. Accordingly, the same name can

[2] Medical conditions are also referred to as Systemic conditions.

be used for a class and individual or instances. This effectively means that the same thing can be treated as a class and an individual or instance depending on the context of the situation. For example, 'hospital' is a class represented as Hospital $\sqsubseteq$ Location as well as an individual represented as Location(Hospital). Depending on the context and the detail of the information that is required to be modelled, 'hospital' can be treated in either way. For instance, if individual wards of the hospital are required to be represented, then 'hospital' can be treated as a class with the various wards being its subclasses or individuals. Punning not only increases the expressivity of the ontology by incorporating contextual modeling but also helps to improve interoperability with other ontologies that may have modelled the same thing differently [86].

Moreover, the reason to model the world is to express information and derive knowledge of the individuals from it. The classes and properties in an ontology express this information and describe the individuals. Without the individuals therefore, validating the classes and properties remains incomplete. SNOMED, however, does not include individuals in its structure and this is one aspect where OSHCO considerably differs from SNOMED. Another aspect where OSHCO differs is in the DL expressivity. SNOMED-CT, which is originally represented in Ontylog [21], can now be converted into OWL and corresponds to the OWL EL profile of OWL 2 [87]. The DL expressivity of SNOMED converted to OWL is $\mathcal{ALER}$ [87] while the expressivity of OSHCO is $\mathcal{SROIF}(\mathcal{D})$ making it more expressive than the former due to the presence of additional concept and role constructors, and RBox axioms. The $\mathcal{AL}$ in $\mathcal{ALER}$ represents Attributive language that allows for: atomic negation, concept intersection, universal restrictions and limited existential quantification; the $\mathcal{E}$ allows full existential quantification and $\mathcal{R}$ represents the presence of complex role inclusion axioms [47,83]. In $\mathcal{SROIF}(\mathcal{D})$, the $\mathcal{S}$ is an abbreviation for $\mathcal{AL}$ in addition to $\mathcal{C}$ which includes complex concept negation; $\mathcal{O}$ refers to enumerated classes or nominals; $\mathcal{I}$ refers to the presence of inverse properties; $\mathcal{F}$ represents functional properties; and $(\mathcal{D})$ refers to datatypes, which are a part of OWL but not of the underlying DL $\mathcal{SROIQ}$. Lastly, $\mathcal{SROIF}(\mathcal{D})$ is contained by $\mathcal{SROIQ}$ and hence is decidable [45].

5.3 Step 3: developing formal rules

The use cases from Sect. 5.1 step 1, and the terms and properties (relationships) that were used to develop the ontology in Sect. 5.2 step 2 served as the blueprint for writing formal rules in SWRL at this stage. By having used only the named ontology classes, we ensure that our rules can be translated to different rule formats, are interoperable with other OWL based ontologies that may or may not support SWRL, and that the rules remain decidable. We developed rules in two situations—where actionable knowledge is required and where conditions cannot be expressed in OWL. For example, with respect to use case 3 described in Table 1, it is possible to express in our OWL based ontology OSHCO that if a patient has some form of DM and periodontal disease then the patient should be automatically classified into a new class of patients who require collaborative (medical-oral) management:

$$\mathsf{PatientRequiringMedicalOralManagement} \equiv \mathsf{Patient} \sqcap (\exists \mathsf{hasMedicalCondition.DiabetesMellitus} \sqcap \exists \mathsf{hasOralCondition.PeriodontalDisease})$$

However, it is not possible to express that a patient should be classified into the class of patients who require collaborative management only if he/she has those medical and oral conditions that are interdependent or in other words, influence each other's prognosis but it can be expressed in SWRL as shown in Table 3, rule complex 1. This is an example that shows how rules can add to the expressivity of the ontology. Rule complex 2 on the other hand shows how rules add actionable knowledge to the ontology. This process is demonstrated diagrammatically in Fig. 6 where if x, y, z are names of individuals, C is a concept and R a role then the concept assertion is represented as C(x) and role assertion as R(y, z) where z is the filler for y with respect to role R. We consider the original ABox as $\mathrm{A_0}$Box; when the first rule in the complex is executed, it adds assertions to the $\mathrm{A_0}$Box to form the new $\mathrm{A_1}$Box. The $\mathrm{A_1}$Box now becomes the original ABox for the second rule, which adds assertions to it to form $\mathrm{A_2}$Box. Lastly, the third rule updates this $\mathrm{A_2}$Box to form the final $\mathrm{A_3}$Box. The new assertions thus result into actionable knowledge namely the preventive measure required for the patient Sam.

Figure 7 is a screenshot of sections of Protégé showing rule complexes 1 (for example patient 'Tim') and 2 (for example patient 'Sam') and the corresponding inferences derived from these rules. The expressions coloured in yellow and within the dotted lines are the new inferences obtained after reasoning while the rest are asserted statements and facts. The new inferences for the patient, Sam, indicate what preventive measure he requires and which conditions is he at risk of developing. As mentioned before, Pellet also provides a justification for the output by showing the path that led to that specific output. For example, Fig. 7 shows one of four justification paths traversed by Pellet for

Table 3 Rule Complexes represented in SWRL format

Rule Complex 1	MedicalCondition(?y) ∧ OralCondition(?z) ∧ influencesPrognosisOf(?y,?z) → hasInterdependency(?y, ?z)
	Patient(?x) ∧ hasMedicalCondition(?x, ?y) ∧ hasOralCondition(?x, ?z) ∧ MedicalCondition(?y) ∧ OralCondition(?z) ∧ hasInterdependency(?y, ?z) → PatientRequiringMedicalOralManagement(?x)
Rule Complex 2	Patient(?x) ∧ hasOralCondition(?x, PoorOralHygiene) ∧ hasOralProcedure(?x, ?y) ∧ SurgicalDentalExtraction(?y) → atRiskOf(?x, BacteraemiaDueToSurgicalDentalProcedure)
	Patient(?x) ∧ atRiskOf(?x, ?z) ∧ hasMedicalCondition(?x, ?y) ∧ Bacteraemia(?z) ∧ CongenitalHeartDisease(?y) → atRiskOf(?x, BacterialEndocarditis)
	Patient(?x) ∧ atRiskOf(?x, BacterialEndocarditis) → requiresPreventiveMeasure(?x, AntibioticProphylaxis)

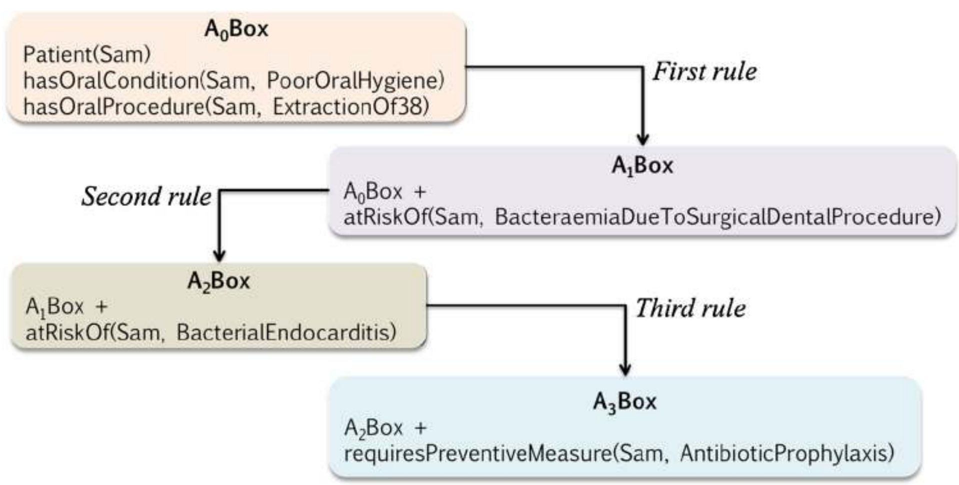


Fig. 6 Diagrammatic representation of changes in ABox resulting in actionable knowledge

rule complex 1 where patient Tim has been classified into a new class. 'PatientRequiringMedicalOralManagement'. The justification module is especially important since any errors in the output can be traced to their source by referring to the path traversed by the reasoner and changes can be made during the development process itself.

5.4 Step 4: implementation and evaluation

The approach we adopted was iterative in nature. Accordingly, we performed regular validation using Pellet to check for ontology consistency, concept satisfiability, classification, and realisation. Moreover, the domain experts were consulted regularly to check the correctness of the represented concepts and rules. In addition to validation, the ontology was implemented and queried over for evaluation purpose. A mix of simple and complex queries was selected to determine if the ontology could answer them and if not, then the ontology was appropriately updated where possible. The queries were also used to test the variety of questions that can be answered with the ontology and how the outcomes can help in Big Data analysis. Table 4 lists some of the useful outcomes that can be derived by employing ontology-based approach to analyse complex and diverse information. Example scenarios are mentioned against some of the use cases from Table 1 to explain the resulting analysis outcomes. The use of ontology-based analysis for research is discussed in detail in Sect. 5.4.1 In addition to the uses listed in Table 4, ontologies can also be used for complex query answering and improving queries to provide more comprehensive answers as exemplified in Sects. 5.4.2 and 5.4.3 respectively.

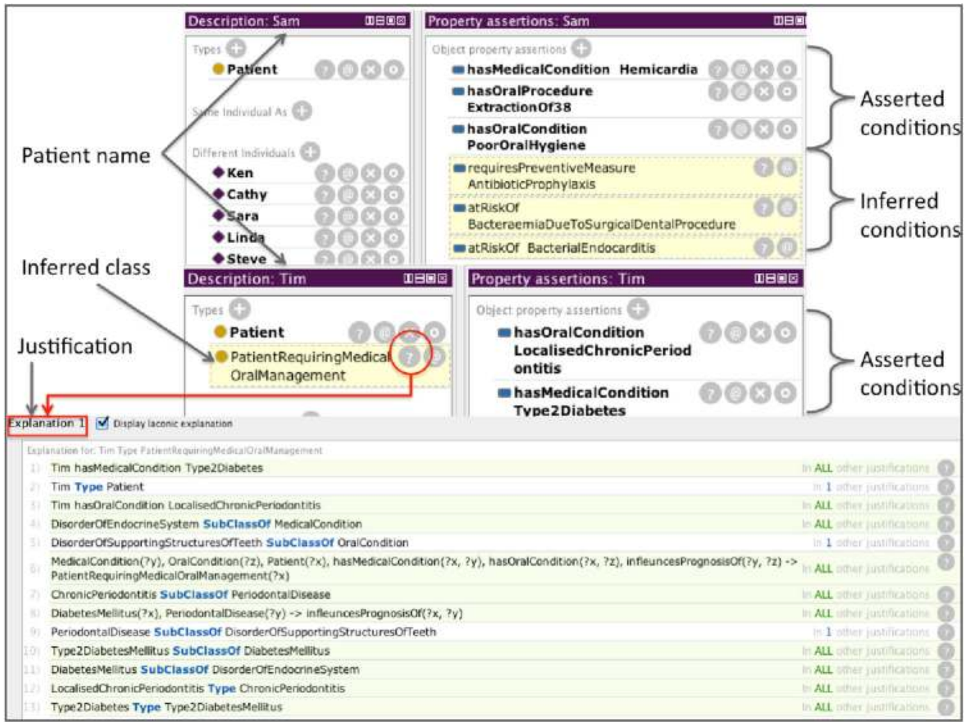


Fig. 7 Screenshot of example inferencing outputs and reasoning justification in Protégé

5.4.1 Research

With respect to use case 3, the knowledge base can be queried to obtain the age group of all the patients who require combined medical and oral health care management. The outcome of such a question can be used to analyse trends regarding most vulnerable age groups who are more likely to suffer from inter-dependent medical and oral health conditions. The results can in turn be used to develop guidelines requiring increased observation and provide precautionary measures to people within those age groups. Figure 8 shows the SPARQL [88] query over OSHCO using the OWL2Query [89] plugin in Protégé and the results obtained from de-identified patient data. In addition to rule complex 1 shown in Table 3, the SWRL rules involved in answering the query include:

5.4.2 Complex query answering

Further, more complex and general questions involving both medical and oral health information of patients can also be asked. Following is an example complex question: identify patients who have both a medical condition and oral condition, with the latter being a type of periodontal disease, such that the existing medical condition influences the prognosis of the periodontal disease. Also identify the type of periodontal disease, the specific medical condition, and the corresponding preventive measure that is recommended for such patients. The SPARQL query for the above question and results based on simulated patient cases are shown in Fig. 9. The evidence-based cross-domain knowledge and association rules modeled in OSHCO can thus be used to obtain consolidated information from even a large amount of patient data such as (i) patient numbers along with the specific types of periodontal disease and medical conditions they have; and (ii) actionable knowledge such as the recommended prophylaxis.

$$\mathrm{Patient}(?x) \wedge \mathrm{dateTime}(?date) \wedge \mathrm{bornOnDate}(?x, ?date) \wedge \mathrm{date}(?date, ?year, ?month, ?day, ?time)$$
$$\rightarrow \mathrm{bornInYear}(?x, ?year)$$
$$\mathrm{Patient}(?x) \wedge \mathrm{bornInYear}(?x, ?year) \wedge \mathrm{presentYear}(?now) \wedge \mathrm{subtract}(?age, ?now, ?year)$$
$$\rightarrow \mathrm{hasAge}(?x, ?age)$$

Moreover, since ontologies can be mapped across several datasets and with other ontologies, a wide variety and large volumes of local and external data can be accessed. The aggregation and analysis of data in this manner provides for a big sample size for public health research such as clinical studies, and disease monitoring for epidemics.

Table 4 Example outcomes of ontology-based Big Data analysis for inter-dependent Medical and Oral Health data

Analysis Outcomes	Use case	Example
Research (identifying trends)	U3	Identification of the age group of patients who most commonly require collaborative care
	U6	Identification of patients with oral papillomatosis who were later diagnosed with Cowden syndrome to analyse the statistical significance of the association between oral papillomatosis and Cowden syndrome
Diagnosis (decision support)	U1	Presence of OHL and candidiasis could indicate the presence of HIV as the underlying condition thereby prompting further tests and diagnosis
Prevention (decision support)	U7	With timely alert from the health information system regarding recommended oral program for intubated patients in the ICU, the appropriate care can be provided thereby preventing likely complications such as pneumonia
Alert (decision support)	U9	Alerting the oral health practitioner towards any HIV positive patient who has very low platelet count so that appropriate actions can be taken before the patient undergoes any major oral procedure that could cause severe bleeding
Collaboration (collaborative patient care provided by the medical and oral health practitioners)	U2	Recommending pregnant women with gingival and/or periodontal conditions to the oral health practitioner
Data quality (missing data/information)	U13	Assessing the completeness of the patient history with respect to all the signs and symptoms of pellagra manifested by the patient
Practice evaluation (following of guidelines)	U5	Evaluating if the clinical guidelines were followed with respect to antibiotic prophylaxis before the surgical dental procedure was carried out on a patient suffering from CHD

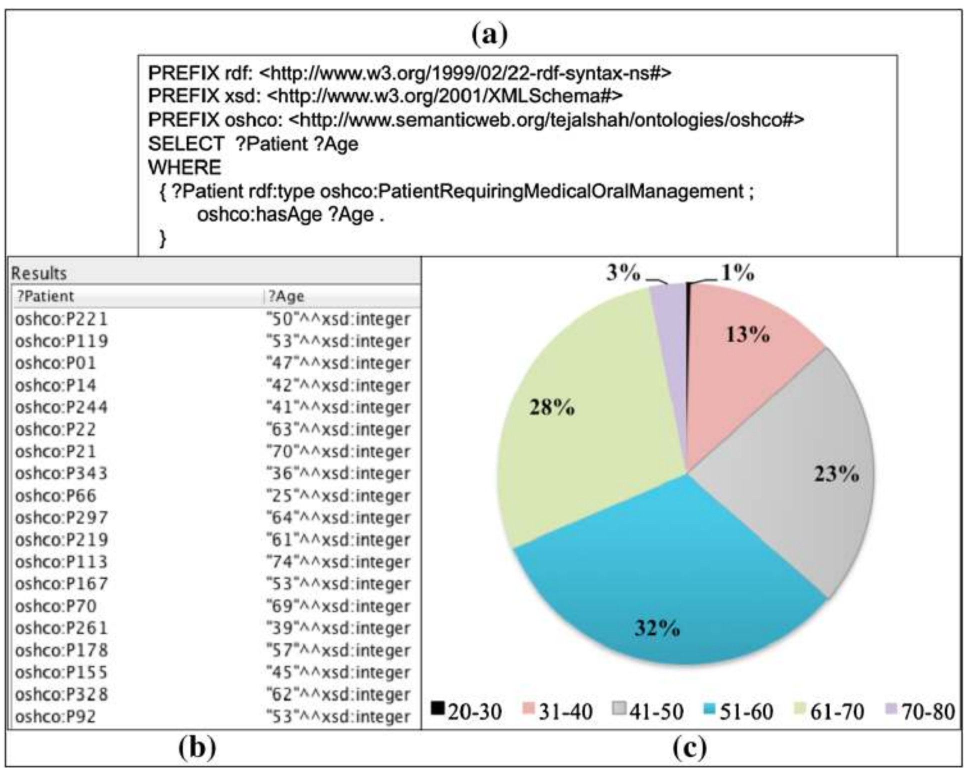
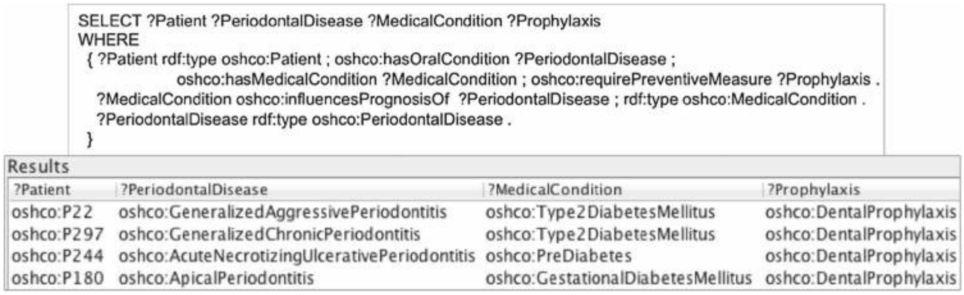


Fig. 8 **a** SPARQL query to identify the age of patients who require combined Medical-Oral Health management, **b** Query results, **c** *Pie chart* showing the distribution of the various age groups

5.4.3 Query enhancement

Considering the fact that the medical and oral health data are often sourced from several subspecialties, querying the data can be difficult. The ontology helps in query enhancement by providing specific answers even when the questions asked are very general. For instance, OSHCO contains a complex role inclusion axiom `affects` $\circ$ `properPartOf` $\sqsubseteq$ `affects`, which is of the form $R \circ S \sqsubseteq R$ where R and S are object properties. The axiom effectively means that if a condition (x) affects a



```
SELECT ?Patient ?PeriodontalDisease ?MedicalCondition ?Prophylaxis
WHERE
 { ?Patient rdf:type oshco:Patient ; oshco:hasOralCondition ?PeriodontalDisease ;
          oshco:hasMedicalCondition ?MedicalCondition ; oshco:requirePreventiveMeasure ?Prophylaxis .
  ?MedicalCondition oshco:influencesPrognosisOf ?PeriodontalDisease ; rdf:type oshco:MedicalCondition .
  ?PeriodontalDisease rdf:type oshco:PeriodontalDisease .
 }
```

Results

?Patient	?PeriodontalDisease	?MedicalCondition	?Prophylaxis
oshco:P22	oshco:GeneralizedAggressivePeriodontitis	oshco:Type2DiabetesMellitus	oshco:DentalProphylaxis
oshco:P297	oshco:GeneralizedChronicPeriodontitis	oshco:Type2DiabetesMellitus	oshco:DentalProphylaxis
oshco:P244	oshco:AcuteNecrotizingUlcerativePeriodontitis	oshco:PreDiabetes	oshco:DentalProphylaxis
oshco:P180	oshco:ApicalPeriodontitis	oshco:GestationalDiabetesMellitus	oshco:DentalProphylaxis

Fig. 9 SPARQL query and the output of the query

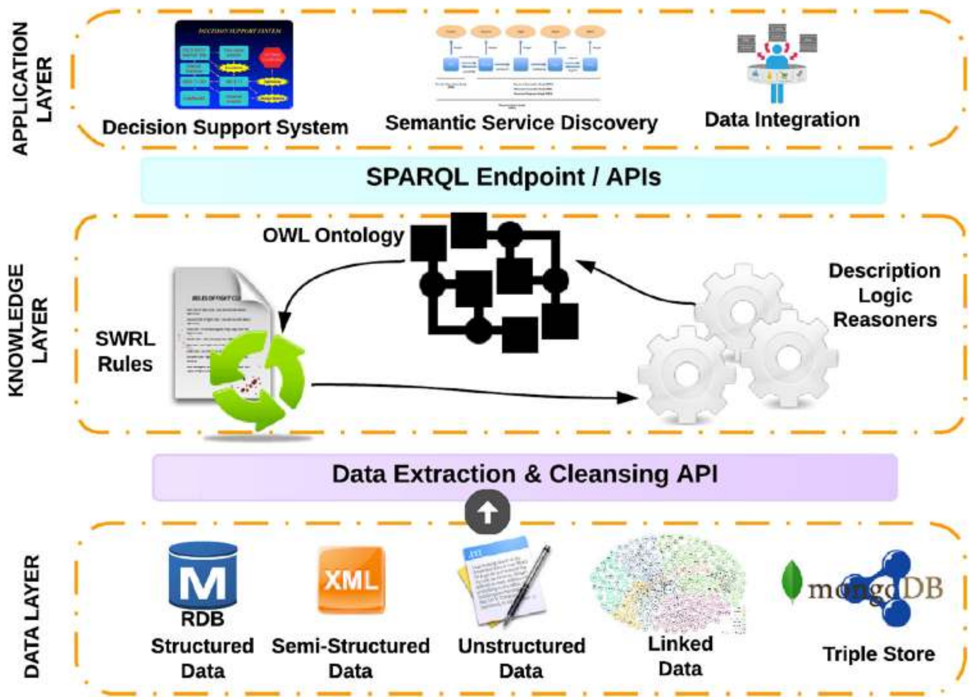


Fig. 10 Ontology and rules in the big picture of Big Data analysis

body structure (y), which is a proper part of structure (z) then the condition (x) also affects the structure (z). For instance,

$$\mathsf{Mandible} \sqsubseteq \exists \mathsf{properPartOf.Jaw}$$
$$\textit{Hence}, \exists \mathsf{affects.Mandible} \sqsubseteq \exists \mathsf{affects.Jaw}$$

I.e., Mandible (commonly known as lower jaw) is a proper part of Jaw; therefore, from the above axiom it follows that if a condition affects the Mandible, then it affects the Jaw as well. Ameloblastoma is a condition that affects the mandible and even if only this specific fact is modelled in the ontology, the axiom will cause the reasoner to automatically classify ameloblastoma as affecting the jaw as well. Thus, when the ontology is queried for the conditions that affect the jaw, ameloblastoma will be one of the answers obtained. Without the role inclusion axiom, only a very specific query targeting the mandible will yield ameloblastoma as its answer. The running example thus shows how querying complex data with the help of ontology is helpful in obtaining a comprehensive set of answers.

6 The big picture

One of the primary goals of Big Data analytics system in supporting health applications is to enable better decision-making via (i) simultaneous processing of new and old data. For example, before prescribing patients with new drugs, it is vital that their past and present medical records are assessed. These records could hold information that prohibits the prescription of the new drug such as allergies, and possible drug-drug interactions; and (ii) simultaneous processing of patient records from different specialties. For example, processing of patients' medical and dental records together helps to identify inter-dependent conditions that must be treated simultaneously for better prognosis. Formal ontologies that are built in a standard language such as OWL and are interoperable with other ontologies, can bridge the semantic gap that often exists between records. By doing so, ontologies enable information sharing and assessment between old and new patient records along with records from across different specialties. Figure 10 depicts where the ontology and rules fit in the big picture of Big Data analysis. The picture consists of three basic layers namely the data layer, knowledge layer and the application layer. The ontology in the knowledge layer can be used to access Big Data, which includes a wide variety of heterogeneous and complex data including structured, semi-structured and unstructured. The ontology, rules and reasoners in the knowledge layer process and analyse this data to derive inferences and obtain new knowledge from it. The new knowledge can then be used in several applications such as decision support, semantic service discovery, and data integration. The work presented in this paper deals with the development and implementation of the knowledge layer to analyse patients' medical and oral health information obtained from disparate data sources and diverse subspecialties within the respective domains.

7 Conclusion

In healthcare, like in most other fields, the information derived from Big Data is becoming invaluable for more informed decision-making, better outcomes, and improved collaborative efforts. However, analysing the massive amounts of data for deriving practically usable information is not a trivial task and the healthcare professionals have already been dealing with an overload of information. In such a scenario, loading them with more information without any help to manage, analyse, and use the information defeats the intended aims and purpose of Big Data. In this paper, we have developed a formal ontology to reason over the data and draw inferences from it. Thus, instead of being limiting factors, we have shown how the characteristics of Big Data such as variety and complexity can be leveraged to provide better and a more comprehensive healthcare support to patients. We discussed in detail how the ontology can be developed via a use-case based approach and further enriched with semantic rules to improve its expressivity without compromising on decidability. The approach is generic enough to be employed for developing and implementing ontology across any domains where Big Data analysis can be complicated due to diverse, heterogeneous and complex data.

References

1. Wang, L., Kunze, M., Tao, J., von Laszewski, G.: Towards building a cloud for scientific applications. Adv. Eng. Softw. **42**(9), 714–722 (2011)
2. Wang, L., Von Laszewski, G., Younge, A., He, X., Kunze, M., Tao, J., Fu, C.: Cloud computing: a perspective study. New Gener. Comput. **28**(2), 137–146 (2010)
3. Alhamazani, K., Wang, L., Rabhi, F., Mitra, K., Ranjan, R.: Cloud monitoring for optimizing the QoS of hosted applications. In: Proceedings of the 2012 IEEE 4th International Conference on Cloud Computing Technology and Science (CloudCom), pp. 765–770. IEEE Computer Society (2012)
4. Yan, J.: Big Data, Bigger Opportunities. White Paper, April 9, 2013. www.meritalk.com/pdfs/bdx/bdx-whitepaper-090413.pdf. Accessed 18 July 2014
5. Yang, X., Wang, L., von Laszewski, G.: Recent research advances in e-science. Clust. Comput. **12**(4), 353–356 (2009)
6. Sun, Y., Wang, L., Jie, W., Fu, C.: Research in e-science: current status and future direction. Int. J. Comput. Syst. Sci. Eng. **26**(3), 151–152 (2011)
7. Shah, T., Rabhi, F., Ray, P., Taylor, K.: Enhancing automated decision support across medical and oral health domains with semantic web technologies. In: Proceedings of the 24th Australasian Conference on Information Systems (ACIS) (2013). http://mo.bf.rmit.edu.au/acis2013/382.pdf. Accessed 23 Jan 2014
8. Chen, C.-L., Zhang, C.-Y.: Data-intensive applications, challenges, techniques and technologies: a survey on big data. Inform. Sci. **275**, 314–347 (2014)
9. Horn, W.: AI in medicine on its way from knowledge-intensive to data-intensive systems. Artif. Intell. Med. **23**(1), 5–12 (2001)
10. Groves, P., Kayyali, B., Knott, D., Van Kuiken, S.: The 'big data' revolution in healthcare. McKinsey Quarterly (2013). http://www.mckinsey.com/insights/health_systems_and_services/the_bigdata_revolution_in_us_health_care. Accessed 26 May 2014
11. von Laszewski, G., Dayal, J., Wang, L.: eMOLST: a documentation flow for distributed health informatics. Concurr. Comput. **23**(16), 1857–1867 (2011)
12. Cutler, C.W., Machen, R.L., Jotwani, R., Iacopino, A.M.: Heightened gingival inflammation and attachment loss in type 2 diabetics with hyperlipidemia. J. Periodontol. **70**, 1313–1321 (1999)
13. Rudman, W., Hart-hester, S., Jones, W., Caputo, N., Madison, M.: Integrating medical and dental records: a new frontier in health information management. J. AHIMA **81**, 36–39 (2010)
14. Semantic Web Health Care and Life Sciences (HCLS) Interest group, W3C. http://www.w3.org/blog/hcls/. Accessed 12 Apr 2014
15. International Health Terminology Standards Development Organisation, SNOMED-CT. http://www.ihtsdo.org/snomed-ct/. Accessed 29 Mar 2014
16. American Dental Association, SNODENT. http://www.ada.org/snodent.aspx. Accessed 1 Mar 2014
17. World Health Organization, International Classification of Diseases (ICD), Classifications. http://www.who.int/classifications/icd/en/. Accessed 19 Sept 2013
18. Halland, K., Britz, K., Gerber, A.: Investigations into the use of SNOMED CT to enhance an OpenMRS health information system. South Afr. Comput. J. **47**, 33–45 (2011)
19. Schulz, S., Suntisrivaraporn, B., Baader, F.: SNOMED CT's problem list: ontologists' and logicians' therapy suggestions. Stud. Health Technol. Inform. **129**, 802–806 (2007)
20. Héja, G., Surján, G., Varga, P.: Ontological analysis of SNOMED CT. BMC Med. Inform. Decis. Mak. **8**, S8 (2008)
21. Rector, A.R., Brandt, S.: Why do it the hard way? The case for an expressive description logic for SNOMED. J. Am. Med. Inform. Assoc. **15**, 744–751 (2008)
22. Goldberg, L.J., Ceusters, W., Eisner, J., Smith, B.: The significance of SNODENT. Stud. Health Technol. Inform. **116**, 737–742 (2005)
23. Falkman, G., Gustafsson, M., Jontell, M., Torgersson, O.: SOMWeb: a semantic web-based system for supporting collab-

oration of distributed medical communities of practice. J. Med. Internet Res. **10**, e25 (2008)
24. Ying, W., Wimalasiri, J.S., Ray, P., Chattopadhyay, S., Wilson, C.S.: An ontology driven multi-agent approach to integrated e-health systems. Int. J. E-Health Med. Commun. **1**, 28–40 (2010)
25. Pérez-Rey, D., Maojo, V., García-Remesal, M., Alonso-Calvo, R., Billhardt, H., Martin-Sánchez, F., Sousa, A.: ONTOFUSION: ontology-based integration of genomic and clinical databases. Comput. Biol. Med. **36**(7–8), 712–730 (2006)
26. Kuziemsky, C.E., Lau, F.: A four stage approach for ontology-based health information system design. Artif. Intell. Med. **50**(3), 133–148 (2010)
27. Staab, S., Studer, R.: Handbook on Ontologies. In: Bernus, P., Blaźewicz, J., Schmidt, G., Shaw, M. (eds.) International Handbooks on Information Systems. Springer, Berlin (2004)
28. Turner, J.A., Mejino, J.L.V., Brinkley, J.F., Detwiler, L.T., Lee, H.J., Martone, M.E., Rubin, D.L.: Application of neuroanatomical ontologies for neuroimaging data annotation. Frontiers Neuroinform. **4**(10), 1–12 (2010)
29. Ganguly, P., Ray, P., Parameswaran, N.: Semantic interoperability in telemedicine through ontology-driven services. Telemed. e-Health **11**(3), 405–412 (2005)
30. Yao, W., Kumar, A.: CONFlexFlow: integrating flexible clinical pathways into clinical decision support systems using context and rules. Decis. Support Syst. **55**, 499–515 (2013)
31. Martínez-Romero, M., Vázquez-Naya, J.M., Pereira, J., Pereira, M., Pazos, A., Baños, G.: The iOSC3 System: using ontologies and SWRL rules for intelligent supervision and care of patients with acute cardiac disorders. Comput. Math. Methods Med. **2013**, 1–13 (2013)
32. Chen, R.-C., Huang, Y.-H., Bau, C.-T., Chen, S.-M.: A recommendation system based on domain ontology and SWRL for anti-diabetic drugs selection. Expert Syst. Appl. **39**, 3995–4006 (2012)
33. Rodrıguez-Muro, M., Calvanese, D.: Quest, an OWL 2 QL reasoner for ontology-based data access. In: Proceedings of the 9th International Workshop on OWL: Experiences and Directions (OWLED 2012). Proceedings of CEUR Electronic Workshop, vol **849** (2012)
34. Decker, S., Erdmann, M., Fensel, D., Studer, R.: Ontobroker: Ontology Based Access to Distributed and Semi-structured Information. Springer, Berlin (1999)
35. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Poggi, A., Rodriguez-Muro, M., Rosati, R., Ruzzi, M., Savo, D.F.: The Mastro system for ontology-based data access. Semant. Web **2**, 43–53 (2011)
36. Bienvenu, M., Ten Cate, B., Lutz, C., Wolter, F.: Ontology-based data access: a study through disjunctive datalog, CSP, and MMSNP. In: Proceedings of the 32nd symposium on Principles of database systems, pp. 213–224. ACM (2013)
37. Soylu, A., Skjæveland, M.G., Giese, M., Horrocks, I., Jimenez-Ruiz, E., Kharlamov, E., Zheleznyakov, D.: A preliminary approach on ontology-based visual query formulation for big data. In Metadata and Semantics Research, pp. 201–212. Springer, Berlin (2013)
38. Baader, F.: The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, Cambridge (2003)
39. Studer, R., Benjamins, V.R., Fensel, D.: Knowledge engineering: principles and methods. Data Knowl. Eng. **25**, 161–197 (1998)
40. Cruz, I.F., Xiao, H.: Ontology driven data integration in heterogeneous networks. Complex Systems in Knowledge-based Environments: Theory, Models and Applications, pp. 75–98. Springer, Berlin (2009)
41. Noy, N.F., McGuinness, D.L.: Ontology development 101: a guide to creating your first ontology. Knowledge Systems Laboratory Stanford University (2001)
42. w3schools: Introduction to OWL. RDF Tutorial, vol. 2012. w3schools (2012)
43. W3C: OWL Web Ontology Language Current Status, Standards, W3C 2013. http://www.w3.org/standards/techs/owl#w3c_all. Accessed 10 Feb 2013
44. Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P.F., Rudolph, S.: OWL 2 web ontology language primer. W3C Recomm. **27**, 1–123 (2009)
45. Horrocks, I., Kutz, O., Sattler, U.: The Even More Irresistible SROIQ. KR **6**, 57–67 (2006)
46. Krötzsch, M.: OWL 2 profiles: an introduction to lightweight ontology languages. In: Eiter, T., Krennwallner, T. (eds.) Reasoning Web, pp. 112–183. Springer, Vienna (2012)
47. Rudolph, S.: Foundations of description logics. Reasoning Web. Semantic Technologies for the Web of Data, pp. 76–136. Springer, Berlin (2011)
48. Kuba, M.: OWL 2 and SWRL Tutorial. http://dior.ics.muni.cz/~makub/owl/ (2012). Accessed 1 Jan 2013
49. Horrocks, I., Patel-Schneider, P.F., Boley, H., Tabet, S., Grosof, B., Dean, M.: SWRL: a semantic web rule language combining OWL and RuleML. W3C Member submission, **21**, 79 (2004). http://www.w3.org/Submission/SWRL/. Accessed 2 Nov 2013
50. Motik, B., Sattler, U., Studer, R.: Query answering for OWL-DL with rules. Web Semant. **3**, 41–60 (2005)
51. Sirin, E., Parsia, B., Grau, B., Kalyanpur, A., Katz, Y.: Pellet: a practical OWL-DL reasoner. Web Semant. **5**, 51–53 (2007)
52. Shearer, R., Motik, B., Horrocks, I.: HermiT: a highly-efficient OWL reasoner. In: Proceedings of the 5th International Workshop on OWL: Experiences and Directions (OWLED 2008 EU), Karlsruhe (2008)
53. Tsarkov, D., Horrocks, I.: FaCT++ description logic reasoner: system description. In: Proceedings of the Third international joint conference on Automated Reasoning, pp. 292–297. Springer, Seattle (2006)
54. Motik, B., Studer, R.: KAON2-A scalable reasoning tool for the semantic web. In: Proceedings of the 2nd European Semantic Web Conference (ESWC'05), Heraklion (2005)
55. Haarslev, V., Müller, R.: RACER system description. Automated Reasoning, pp. 701–705. Springer, Berlin (2001)
56. Konstantinou, N., Spanos, D.-E.: NikolasMitrou: Ontology and database mapping: a survey of current implementations and future directions. J. Web Eng. **7**, 001–024 (2008)
57. Golbreich, C., Horrocks, I.: The OBO to OWL mapping, GO to OWL 1.1!. In: Golbreich, C., Kalyanpur, A., Parsia, B. (eds.) Proceedings of the OWLED 2007 Workshop on OWL: Experiences and Directions. Innsbruck (2007)
58. Chattopadhyay, A., Caplan, D.J., Slade, G.D., Shugars, D.C., Tien, H.C., Patton, L.L.: Risk indicators for oral candidiasis and oral hairy leukoplakia in HIV-infected adults. Community Dentistry Oral Epidemiol. **33**, 35–44 (2005)
59. Laskaris, G., Hadjivassiliou, M., Stratigos, J.: Oral signs and symptoms in 160 Greek HIV-infected patients. J. Oral Pathol. Med. **21**, 120–123 (1992)
60. Li, X., Kolltveit, K.M., Tronstad, L., Olsen, I.: Systemic diseases caused by oral infection. Clin. Microbiol. Rev. **13**, 547–558 (2000)
61. Mealey, B.L.: Periodontal disease and diabetes a two-way street. J. Am. Dental Assoc. **137**, 26S–31S (2006)
62. Marcus, B.J., Kaplan, J., Collins, K.A.: A case of Ludwig angina: a case report and review of the literature. Am. J. Forensic Med. Pathol. **29**, 255–259 (2008)
63. Roda, R.P., Jiménez, Y., Carbonell, E., Gavaldá, C., Muñoz, M.M., Pérez, G.S.: Bacteremia originating in the oral cavity. A review. Med. Oral Patol. Oral Cir. Bucal. **13**, E355–E362 (2008)
64. Pilarski, R., Eng, C.: Will the real Cowden syndrome please stand up (again)? Expanding mutational and clinical spectraof the PTEN hamartoma tumour syndrome. J. Med. Genet. **41**, 323–326 (2004)

65. Chippagiri, P., Banavar Ravi, S., Patwa, N.: Multiple hamartoma syndrome with characteristic oral and cutaneous manifestations. Case reports in dentistry **2013**, article ID 315109 (2013)
66. Fistarol, S.K., Anliker, M.D., Itin, P.H.: Cowden disease or multiple harmartoma syndrome-cutaneous clue to internal malignancy. EJD **12**, 411–421 (2002)
67. Cason, C.L., Tyner, T., Saunders, S., Broome, L.: Nurses' implementation of guidelines for ventilator-associated pneumonia from the centers for disease control and prevention. Am. J. Crit. Care **16**, 28–36 (2007)
68. Migliorati, C.A., Siegel, M.A., Elting, L.S.: Bisphosphonate-associated osteonecrosis: a long-term complication of bisphosphonate treatment. Lancet Oncol. **7**, 508-514 (2006)
69. Davies, R., Bedi, R., Scully, C.: ABC of oral health: oral health care for patients with special needs. British Med. J. **321**, 495 (2000)
70. Newman, M.G., Takei, H., Klokkevold, P.R., Carranza, F.A.: Carranza's clinical periodontology. Elsevier Health Sciences, Amsterdam (2011)
71. Dajani, A.S., Taubert, K.A., Wilson, W., Bolger, A.F., Bayer, A., Ferrieri, P., Gewitz, M.H., Shulman, S.T., Nouri, S., Newburger, J.W.: Prevention of bacterial endocarditis recommendations by the American Heart Association. Circulation **96**, 358–366 (1997)
72. Shay, K.: Infectious complications of dental and periodontal diseases in the elderly population. Clin. Infect. Dis. **34**, 1215–1223 (2002)
73. White, S.C., Pharoah, M.J.: Oral radiology: principles and interpretation. Elsevier Health Sciences, Amsterdam (2013)
74. Ridgeway, E.E.: Periodontal disease: diagnosis and management. J. Am. Acad. Nurse Pract. **12**, 79–84 (2000)
75. Brown, T.M.: Pellagra: an old enemy of timeless importance. Psychosomatics **51**, 93–97 (2010)
76. Lurie, O., Zadik, Y., Einy, S., Tarrasch, R., Raviv, G., Goldstein, L.: Bruxism in military pilots and non-pilots: tooth wear and psychological stress. Aviat. Space Environ. Med. **78**, 137–139 (2007)
77. Medical Board of Australia. http://www.medicalboard.gov.au/Registration/Types/Specialist-Registration/MedicalSpecialties-and-Specialty-Fields.aspx. Accessed 5 Mar 2014
78. Dental Board of Australia. http://www.dentalboard.gov.au/Registration/Specialist-Registration.aspx. Accessed 1 Apr 2014
79. Krötzsch, M., Simancik, F., Horrocks, I.: A description logic primer. arXiv:1201.4089 (2012)
80. Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J.: The OBO foundry: coordinated evolution of ontologies to support biomedical data integration. Nat. Biotechnol. **25**, 1251–1255 (2007)
81. Shah, T., Rabhi, F., Ray, P., Taylor, K.: A Guiding framework for ontology reuse in the biomedical domain. In: Proceedings of the 2014 47th Hawaii International Conference on System Sciences (HICSS), pp. 2878–2887. IEEE (2014)
82. Dentler, K., Cornet, R., Teije, At, Keizer, Nd: Comparison of reasoners for large ontologies in the OWL 2 EL profile. Semant. Web J. **2**, 71–87 (2011)
83. Stanford Center for Biomedical Informatics Research: Protégé Ontology Editor. http://protege.stanford.edu. Accessed 11 June 2013
84. Institute for Human and Machine Cognition (IHMC): CmapTools. http://cmap.ihmc.us/. Accessed 30 June 2014
85. Golbreich, C., Wallace, E.K.: OWL 2 web ontology language: new features and rationale. W3C Recomm. (2012). http://www.w3.org/TR/owl2-new-features/. Accessed 5 May 2014
86. Bergman, M.K.: Metamodeling in Domain Ontologies. AI3 (2014). http://www.mkbergman.com/913/metamodeling-in-domain-ontologies/. Accessed 2 Dec 2013
87. International Health Terminology Standards Development Organisation: SNOMED CT ® Technical Implementation Guide (2014)
88. Harris, S., Seaborne, A.: SPARQL 1.1 query language. W3C Recomm. (2013). http://www.w3.org/TR/sparql11-query/. Accessed 31 Jan 2014
89. Křemen, P., Kostov, B.: OWL2 Query plugin (2011). http://krizik.felk.cvut.cz/km/owl2query/. Accessed 8 May 2014

Tejal Shah is a Ph.D. student in the School of Computer Science and Engineering. Her main research focus is in developing semantic technologies that mitigate the interoperability problem across medical and dental domains. Her research work has appeared in international conferences including Healthcom 2013, ACIS 2013 and HICSS 2014. She has also published one ISI journal paper.

Fethi Rabhi is a Professor in the School of Computer Science and Engineering at the University of New South Wales, Australia. The main theme of his research over the past 25 years has been the application of innovative software engineering techniques in areas as diverse as physics, electronics, health, economics and finance. This has now been consolidated into the application of Service-Oriented Computing (SOC) in the eResearch domain in general, finance in particular. Besides his academic role, he is actively engaged in facilitating University-industry collaborations particularly in the context of multi-disciplinary applied research.

Pradeep Ray is a Professor in the School of Business at the University of New South Wales, Australia. He is the founder of the International Initiative for ubiquitous Healthcare (u-Health) that is setting up a chain of u-Health research centers called Asia-Pacific ubiquitous Healthcare Research Centre (APuHC) with nodes in Australia, Taiwan, Singapore, South Korea and India. This u-Health initiative, launched during the Med-e-Tel 2006 in Luxembourg in April 2006, is aimed at exploiting the mobile broadband communication technologies for healthcare. APuHC is now designated (Aug 2013) as a WHO Collaborating Centre for eHealth to provide support to WHO programs in e/mHealth in Asia Pacific region.