J Med Syst (2014) 38:67
DOI 10.1007/s10916-014-0067-4

SYSTEMS-LEVEL QUALITY IMPROVEMENT

An Ontological Case Base Engineering Methodology for Diabetes Management

Shaker H. El-Sappagh · Samir El-Masri · Mohammed Elmogy · A. M. Riad · Basema Saddik

Received: 23 October 2013 / Accepted: 28 May 2014 / Published online: 24 June 2014

Abstract Ontology engineering covers issues related to ontology development and use. In Case Based Reasoning (CBR) system, ontology plays two main roles; the first as case base and the second as domain ontology. However, the ontology engineering literature does not provide adequate guidance on how to build, evaluate, and maintain ontologies. This paper proposes an ontology engineering methodology to generate case bases in the medical domain. It mainly focuses on the research of case representation in the form of ontology to support the case semantic retrieval and enhance all knowledge intensive CBR processes. A case study on diabetes diagnosis case base will be provided to evaluate the proposed methodology.

Keywords Case based reasoning · Ontology engineering · Case representation · Knowledge management and clinical decision support system

This article is part of the Topical Collection on *Systems-Level Quality Improvement*

S. H. El-Sappagh
Department of Mathematics, College of Science, King Saud University, Riyadh, KSA, Saudi Arabia

S. El-Masri (✉)
Department of Information Systems, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia
e-mail: selmasri@ksu.edu.sa

S. H. El-Sappagh · M. Elmogy · A. M. Riad
Faculty of Computers & Information, Mansoura University, Mansoura, Egypt

B. Saddik
College of Public Health and Health Informatics, University of King Saud Bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia

Introduction

Ontology is a formal and explicit specification of a shared conceptualization [1]. Ontology defines a common vocabulary for researchers who need to share information in a domain. It includes machine-interpre definitions of basic concepts (classes) in the domain, properties of each concept describing various features and attributes of the concept (slots, relationships) and restrictions on slots (facets or role restrictions) [2]. Ontology together with a set of individual instances of classes constitutes a knowledge base. Ontology engineering methodology may be understood as an organized set of procedures and guidelines aiding and guiding the development of ontology during its lifecycle, or parts of it [3] and development can be top-down, bottom-up or middle-out.

A number of approaches have been reported for developing ontologies. Some of which describe how to build ontology from scratch or by reusing other ontologies [2–10]. All these methodologies are considered incomplete compared to the complete life cycle methodology in software engineering. Moreover, each methodology is specialized in a specific domain or for a specific purpose. In CBR field, case representation and case retrieval are the most important and completely related steps. Case representation defines the case base conceptual model, which can be physically stored in any format (i.e. XML, database, or even text file). There are many formats for case base such as feature vector and object oriented which shorten case retrieval process. In this paper, we will present how cases can be represented and stored in ontology. We will develop an ontology engineering methodology specialized in CBR case base creation. This case base can be instantiated in the medical domain. The combination of ontologies for domain and case base will achieve the Knowledge Intensive KI-CBR [11]. As shown in Fig. 1, case retrieval has enough semantic understanding ability, which can accurately understand query intention and retrieve the desired case. The figure main steps are concept standardization with domain ontology, and case retrieval from case base ontology. Ontology has not

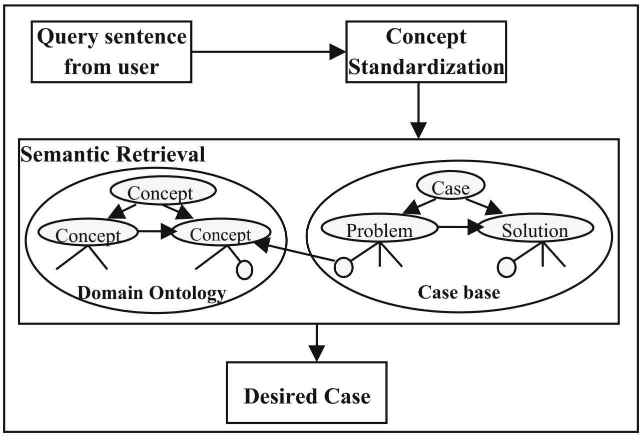


Fig. 1 Semantic case retrieval process

only powerful ability of knowledge acquisition and representation, but also good semantic understanding. The purpose of this research is to develop high quality case base ontology engineering methodology which takes into account the medical domain features, temporal, uncertainty, context, default values, and how to deal with Electronic Health Record (EHR) and Clinical Practice Guideline (CPG). The paper is organized as follows: Section 2 outlines literature on related works, Section 3 lists the contribution point, Section 4 illustrates the proposed methodology, Section 5 evaluates the proposed methodology, Section 6 discusses future work and finally section 7 is the conclusion.

Related work

Ontology plays an important role in each step of KI-CBR systems [12]. It can be used to represent domain knowledge, case base, or both [13]. The linkage of the two sources has not been handled before, and our research will proposed a solution. Especially in medical domain, the background domain ontology for CBR systems can be standardized using any of standard ontologies as SNOMED CT, GALEN, UMLS, LOINC, ICD, disease ontology, glucose metabolic disorder and others [14–16]. We have created a diabetes diagnosis domain ontology from SNOMED CT in other work [17]. Building medical case base ontology for CBR is more complicated process, and a methodology for engineering the process is required from requirement specification to operation and evolution [18]. Existing ontologies are incomplete, do not support accurate semantic retrieval, do not have effective plans for case base updates and most importantly are not specialized in the medical domain. Ontology engineering methodologies, tools and languages have been discussed in the literature, [19], however all of these methodologies do not provide a complete ontology lifecycle as in the software engineering case. Moreover, each methodology is specialized in a specific domain or for a specific purpose [20]. These methodologies are ordered chronologically as CYC [4], TOVE [21], Uschold et al. [5], KACTUS [22], METHONTOLOGY [6], SENSUS [23], On-To-Knowledge [7], Ontology development 101 [2], Termontography [24], DOGMA [25], UPON [26], SMOL [27] and NeOn [28]. These methods in general have similar characteristics and steps. Despite the large number of proposed ontology engineering methodologies, still the field lacks widely accepted and mature methodologies [29]. As a result, application of these models will require enhancements. For example, [30] has proposed OntoDocMan methodology which merged the methodologies [6] and [7]. Haghighi et al. [31] has asserted that ontology development for medical domain requires extensions to any of existing methodologies.

In medical environment, ontology engineering methodologies varies according to its purpose and domain. Jaya [18] has proposed a methodology for creating diabetes diagnosis symptom ontology. Zhao and Passi [32] have provides an ontology methodology for Cancer patients follow-up care. Rahimi et al. [33] has proposed a methodology for the development of an ontology to assess data quality in diabetes management. However, these used methodologies are very abstract and cannot be followed to create complex ontologies. Forbes et al. [34] has asserted that diabetes ontology can be created using terms from clinical practice guidelines and has used NeOn methodology to create a patient-practitioner communication ontology for diabetes type 2. Selecting the appropriate methodology for specific purpose or domain affect the quality of resulting ontology. For example, Sutton et al. [35] has extended the methodology in [2] to be suitable for building ontology for specific domain (type 1 diabetes management) and for specific purpose (mobile applications). However, this ontology only collect the diabetes used terms in an OWL DL ontology to support interoperability, and building case base ontology must handle specific aspects regarding CBR as a clinical decision support system.

Ontology plays a critical role in diabetes diagnosis CBR [36]. However, all of the used ontologies in medical domain, especially in chronic diseases as diabetes, do not build case base ontologies. Development of ontology for CBR is not yet an engineering process [40]. However, in some domains as engineering, there are ontology engineering methodologies specific for case base construction [37]. There is no equivalence in medical domain. Moreover, non-medical domain methodologies are not suitable for medical domain. As a result, there are no studies for the methodologies for case base ontology construction for medical domain. In this paper, we will propose a new and complete framework for OWL 2 case base ontology engineering. Moreover, diabetes diagnosis will be taken as a case study, and the ontology will generate the ontology TBOX part of the ontology.

With respect to ontology tools, every step in ontology engineering is supported by a tool. For instance, 3Store, OWLIM, Jena and others can manage ontology repositories. Ontology programming interfaces such as OWL API, RDF2Go and SemWeb.NET can provide programming interfaces. Ontology reasoners such as CEL, Cerebra Engine, FaCT ++, fuzzyDL, HermiT, KAON2, MSPASS, Pellet, QuOnto, RacerPro, SHER, SoftFacts and TrOWL take care of reasoning over ontologies. For ontology editors, Protégé, SemTalk, DogmaModeler, FONTE, DODDLE-OWL, Hozo, NeOn, Ontolingua, WebODE and Semantic Works exist. Brownsauce, BrowseRDF, Disco,/facet, Fenfire, Jambalaya, Longwell, mSpace, OINK, Ontosphere 3D, Ontoviz, OWLViz, RDF Gravity, Tabulator, TGVizTab and Welkin are ontology browsers. gOntt is ontology activity scheduling. The Ontology learner acquires knowledge and generates ontologies of a given domain through some kind of (semi) automatic process (e.g., KEA, OntoGen, OntoLearn, Text2Onto and TERMINAE). The Ontology versioner maintains, stores, and manages different versions of ontology (e.g., SemVersion). Ontology population from database includes DB2OWL. In this research paper, we will learn from these methods and develop our methodology specialized for the medical domain (diabetes management) and for CBR. Protégé [38] and OWL2 [39] will be used to build and manage the ontology.

Main contribution

This section provides the main contributions of this paper:

- This is a pioneered work for creating a *case base ontology* in medical domain (e.g. diabetes diagnosis) CBR systems.
- It proposes a novel *ontology engineering methodology* for creating a case base ontologies for medical domains from heterogeneous sources through four phases over a structured flow. It is oriented to medical experts and knowledge engineers. This action has been applied in industry field as in [37, 40], however our methodology is more complete, and it provides a detailed framework that can be followed in any other medical domain.
- A *case study* for the paper has been applied for diabetes diagnosis, and ontology is created and implemented using OWL 2 language and Protégé tool. The methodology creates the ontology TBOX part (i.e. classes and properties), and ABOX (i.e. instances) can be populated from EHR database using any existing tool or API (e.g. OWL API 3, Jena, DB2OWL).
- The methodology handled the *temporal abstraction* dimension appears in the chronic diseases management.
- The methodology handled how to integrate *CPG rules* in the ontology using SWRL.
- The *connection of domain and case base ontology* dramatically increase the intelligence of CBR. In this work, a method has been proposed for connecting standard terminologies as SNOMED CT and ICD with case base ontology. A semantic case retrieval algorithm can be easily developed based on the two ontologies.

Proposed ontology engineering methodology

The proposed methodology will build an ontology using the middle-out strategy. This strategy identifies concepts from the most relevant to the most abstract and most concrete. The process of developing ontology is an iterative, and its steps are manual and semi-automatic [41]. First, the sketch framework of ontology will be constructed, and then it will evolve to reach the final ontology. Each phase has input, output, participants and activities. This methodology is organized into four phases as shown below, see Fig. 2.

Phase 1: Requirement specifications

Ontology developers (ontology engineers and domain experts) firstly specify the requirements that the ontology should fulfill. The output is Ontology Requirements Specification Document (ORSD), which answers the set of Competency Questions (CQs). ORSD includes the purpose, the scope, the implementation language, and the intended uses of the ontology. This phase is achieved manually by ontology developers and its inputs are the problem to solve, the existing ontological and non-ontological resources, and will produce ORSD. It has the following steps:

Step 1: Determine the ontology domain and scope

The ontology domain refers to the specific area of ontology research such as medical diagnosis or medical treatment. This paper discusses ontology building for the medical domain especially for diabetes management [42]. The scope is the CBR context in diabetes detection, diagnosis and prevention. For each case, the ontology specifies the problem description, the solution and the outcome of the solution.

Step 2: Consider reusing existing ontologies

Much ontology is already available in electronic form and can be imported into an ontology-development environment. Moreover, as EHR is one of the main sources of cases in CBR, it provides a temporal ordering of clinical events, i.e. testing Fasting Plasma Glucose (FPG). Time ontology (i.e. OWL time ontology and Semantic Web Rule Language (*SWRL*) *temporal ontology*) is used to model the temporal dimension in medical data analysis [43]. The SWRL provides basic and advanced build-ins to handle temporal reasoning and supports the creation of custom build-ins to be used in

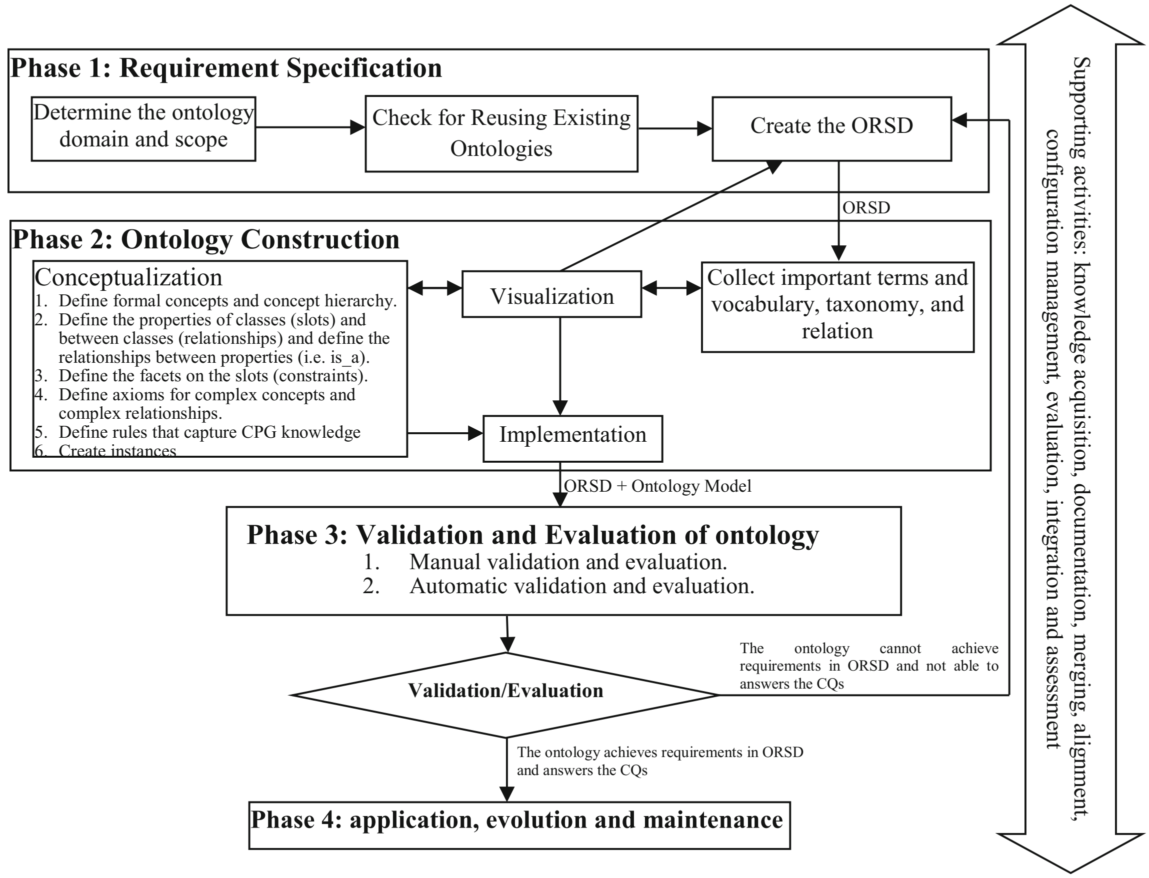


Fig. 2 Methodology for ontology development

rules. Temporal event is represented by many methods such as N-ary relation approach [44] which suffers from data redundancy in the case of inverse and symmetric properties. This research will use the method of Martin et al. [45] for handling temporal events (i.e. <FPG, 127, mg/dL, "1/1/2013", "1/2/2013">). As shown in Fig. 3, any temporal concept is a subclass of *temporal Entity*, but the time independent concept is not and object properties will handle temporal events.

FPG (f1)
has_value (f1,127)
hasUnit (f1, mg/dL)
ValidInterval (t1)
hasBegining (t1, "1/1/2013")
hasFinish (t1, "1/2/2013")
hasValidTime (f1,t1)

FPG (f2)
has_value (f2, 128)
hasUnit (f2, mg/dL)
ValidInterval (t2)
hasBegining (t2, "1/2/2013")
hasFinish (t1, "1/3/2013")
hasValidTime (f2,t2)
ValidInterval (t3)
hasBegining (t3, "1/1/2013")
hasFinish (t3, "1/3/2013")
Temporal_trend(a1)
has_value(a1, "increasing")
Trend_in_interval(a1, t3)

FPG(?x) ^ FPG(?y) ^ hasValidTime(?x, ?t1) ^ hasValidTime(?y, ?t2) ^ temporal: after (t1, t2) ^ has_value (?x,?v1) ^ has_value (?y,?v2) ^ SWRLb:greaterThan(?v2, ?v1) ^Temporal_trend (?TR)→ has_trend(?x, ?TR)

The temporal events and temporal abstraction can be represented as follows:

Case will be represented with two types of features such as multi-valued features which take at least two values over time as Glycated Haemoglobin test (Hb_{A1c}) and FPG and one-value features such as sex. Temporal features range from instant features such as "the lab test checked in at 8 PM" or interval feature as "the weight checked in starting from Jan. 10 and ended in Jan. 19". Temporal abstraction will be done only for temporal features. Consistency check of the ontology is done by pellet and by temporal Reasoner as SOWL [46]. Standard medical terminology such as SNOMED CT, are considered domain ontology. The CBR features are divided in to *simple features* such as SEX, which has one meaning, and *concept features* as symptom, disease, test, etc. Each concept feature has two attributes (its CODE and

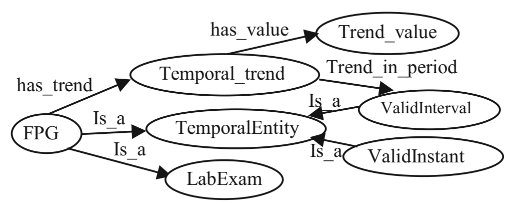


Fig. 3 Handling temporal events using subclass method

CODING_SYSTEM). There should be at least one reference per low level class to a known medical terminology. The Yu Lin and Norihiro Sakamoto ontology for glucose metabolism disorder (OGMD) [47] will be used for the proposed methodology since it contains concepts for diabetes complications and disorders such as Hypertension_in_Diabetes and Diabetes Neuropathy.

Step 3: Create the ORSD

Ontology requirements are represented by ORSD, which answer a set of questions, *CQs*. These questions will serve as the test and validation later. A competency question is a typical query that an expert might want to submit to a knowledge base of its target domain, for a certain task. Every query requested from the CBR for diabetes must be specified informally as a competency question such as:

1. What are the useful features (symptoms, lab tests, vital signs, demographics, drugs…) for diagnosing the patient with diabetes?
2. What are the target types of diabetes (type 1, type 2 and gestational diabetes)?
3. What are the target types of patients (children, adult and pregnant)?
4. How to handle temporal property of diabetes features?
5. How can diabetes be prevented according to current diagnosis and medical history?
6. How screening for diabetes is carried out?
7. How is diabetes diagnosed?
8. How can people with diabetes be managed?
9. How can we build medication, diet and exercise plans to prevent or manage diabetes?

The ontology must contain the necessary axioms to represent and answer these domain oriented questions. Figure 4 shows the relationship between ontology goals, concepts and CQs. Concepts are also covered in the next sections.

After identifying CQs as the set of ontology's set of functional requirements, it must be validated for correctness, completeness, consistency, verifiability, understandability, unambiguity, conciseness, and prioritized. The ORSD can now be created as shown in Table 1.

Phase 2: Ontology construction

Ontology construction is the core phase, which involves the creation of ontology framework. Ontological case base will be

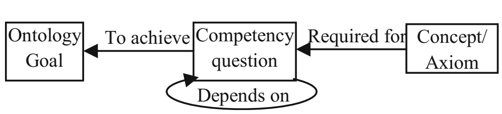


Fig. 4 Relationship between domain concepts, goals, and CQs

Table 1 ORSD

Ontology requirment specification document	
1	Purpose
	The purpose is to build a CBR that enhance the functionality of CBR system.
2	Scope
	As Specified before.
3	Implementation language
	The ontology has to be implemented in OWL.
4	Intended end-users
	User 1: Patient. *User 2*: General practitioner. *User 3*: Diabetes domain expert
5	Intended uses
	The ontology will be used as a case base in diabetes management CBR system.
6	Ontology requirements
	a. Non-functional requirements
	The ontology will support only English language. The ontology will follow the standard naming method as the class name is singular and begin with capital letter, and subclass name will contain parent class as a suffix. The low level concepts in the ontology will be connected to the standardized terminological systems as SNOMED CT.
	b. Functional requirements
	It is the list of prioritized CQs specified previously.

created with dynamic case structure and with multiple similarity measures according to the feature type and value. This phase is semi-automatic and achieved by an ontology engineer with the help of the domain expert. Its input is ORSD, and its output is the ontology model. It has the following steps:

Step 1: Collect important terms and vocabulary, taxonomy, and relations

Initially, it is important to get a comprehensive list of terms without worrying about overlap between concepts they represent, relations among the terms, any properties that the concepts may have, or whether the concepts are classes or slots. Terms are the natural language and informal phrases of the corresponding concepts. There are two methods for determining the domain list of informal terms. Firstly from domain expert brainstorming, text books and research articles, however, the results are very limited and incomplete. The other method reuses existing resources such as medical databases [48, 49], existing ontologies [50], CPGs [51–58] and may be data mining results (association rules, classification and clustering [59]). This will retrieve a comprehensive list of used terms in the medical domain (diabetes), temporal concepts, context concepts and CBR features. Table 2 displays a small sample of the general terms and relations that exist in CBR and diabetes domains. For CBR case base concepts, the

Table 2 Small sample of ontology key terms

Lab Exam	Physical Exam	Blood Pressure
Fasting plasma glucose	Disorder	Vital sign
Case base	Feature	Context
Weight	Case	Diabetes type
Case problem	Case solution	Valid time
Body mass index	Pregnancy	HbA1c

existing case base ontologies are considered as in [50, 60]. For example, the concept of Case Description is utilized in the form of Case description={Context, Problem, Solution and Outcome}. Moreover, the Feature concept is used as described with {Domain, Terminology_Concept, Attribute and Value}. Other concepts, relations, axioms and rules are reused in a customized way in our ontology. The Time concept is derived from these ontologies but not handled in the same way. W3C standard Time Ontology in OWL provides the temporal concepts and relations such as temporal instant, interval and event [44]. It is a general ontology that describes time representation using ontology. To add a time dimension to any diabetic temporal concept, we will use the temporal ontology in [61]. We will add other concepts in our ontological case base regarding context, feature weights, temporal events, temporal abstraction, diabetes management and case status (unconfirmed, confirmed, protected and obsolete), etc.

Step 2: Visualization

Visualization operations overcome the gaps between how people think about a domain and the languages in which ontologies will be formalized. Using graphical and tabular notations, the collected informal terms from previous steps can be converted to conceptual models as in software engineering. Representing the domain problem using collected vocabularies in the form of UML diagrams or Entity Relationship (ER) diagrams will provide advanced understanding of the domain. Concepts, concepts relationships are semi-formally defined. This is an iterative process, and return to the previous phase is acceptable.

The structural aspect of ontology can be modeled by UML *class diagram* (ignoring the methods part of each class) and *object diagram* to represent instances. There will be focus on UML modeling particularly the class diagram, which shows the types being modeled within the system [62]. Figure 5 shows a fragment of the domain class diagram. The Case Base class contains cases, and Case class composed of four parts: The Problem, Solution, Context, and Outcome classes. The problem is described by features in Feature class. The feature can be Test, Characterization and other classes [63].

Step 3: Conceptualization

All created conceptual models (UML and ER) in the previous step are input for this step. The ontology will include specific concepts such as CasualPlasmaGlucose and general concepts such as Problem.

This step consists of the following sub-steps:

Sub-Step 1: Define formal concepts, concept hierarchy, CBR features weights, features context and default values

In this process, all the vocabularies and terms that express the same meaning in the domain are taken as a group into one ontological concept. All vocabularies and terms that belong to the same ontological concept are called synonyms to the concept. There are several approaches to develop the class hierarchy (is_a relationship between concepts) such as top-down, bottom-up and middle-out development approaches. The proposed methodology will follow the last approach, middle-out, as shown in Fig. 6. Weights and default values of features will be estimated by medical experts [64]. The context of cases is used in case indexing where the context is defined by three features with disjoint values: *Diabetes type* (T1DM, T2DM or GDM), *sex* (male or female) and *age* (child or adult). Any case must have specific value for each of these features. Table 3 provides some of concepts (classes) and there sub-concepts (subclasses).

Sub-step 2: Define the properties of classes (data properties) and between classes (relationships or object properties) and define the relationships between properties (i.e. Is_a properties hierarchy)

The classes alone will not provide enough information to answer the CQs from Phase 1. Once we have defined some of the classes, we must describe their internal structure. Figure 7 is a Protégé snapshot to show the built ontology's data type and object properties. Table 4 describes some of these properties in some detail as domain and range of each property.

Sub-step 3: Define the facets on the slots (constraints)

Defining different facets on slots as the value type (i.e. string, numeric, date/time, other concept, Boolean, enumerated …), allowed values, the number of the values (i.e. minimum, maximum and exact cardinality) is required. For instance, *hasProblem exactly 1 Problem* and *hasSolution exactly 1 Solution* say that each case must have exactly one problem and one solution. *solvedBy exactly 1 Solution* specifies one

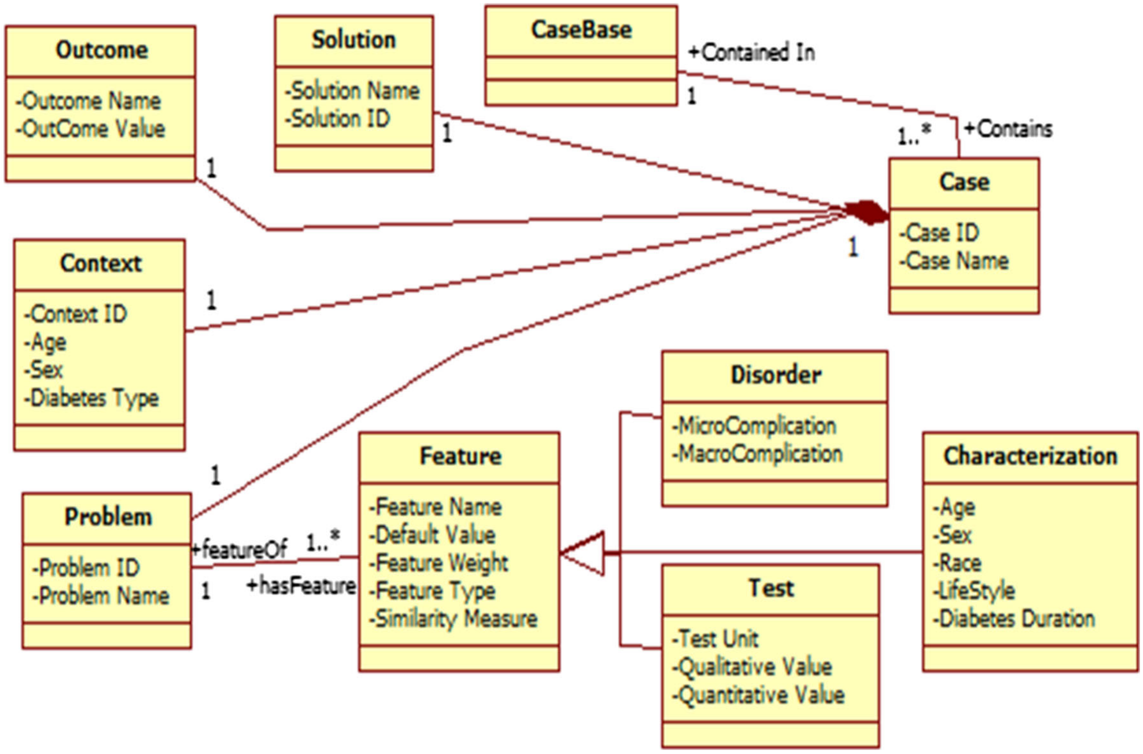


Fig. 5 Part of ontology UML class diagram

solution for a problem. *hasFeature min 16 Feature* specifies that each case is represented with a minimum of 16 features as (FBG, BMI, HbA1c, Age, …). Moreover, Symmetric, transitive, functional, inverse, inverse-functional, and other properties should be defined.

Sub-step 4: Define axioms for complex concepts and complex relationships

Some complex concepts and properties can be created from existing ones by means supported by the ontology language.

Fig. 6 Ontology concepts

Table 3 Classes in the CBR Ontology

Class	Subclasses	Type
Case	–	CBR
Feature	Test, Disorder, Characterization,	Diabetes
ValidTime	ValidInstance, ValidInterval	Temporal

For instance, in OWL 2, we can define new concepts and relationships by using class and property expressions as IntersectionOf, UnionOf, ComplementOf, OneOf, SomeValuesFrom, AllValuesFrom, EquivalentTo, DisjointWith, DisjointUnion and so on. For example, *BloodPressure≡Systolic or Diastolic* and *LabExam DisjointWith PhysicalExam* specifies that the two classes are separated.

Sub-step 5: Define rules to add complex axioms to OWL ontology and capture CPG knowledge

CPGs in diabetes management contain actionable medical knowledge. OWL cannot represent deductive knowledge (i.e. rules). This step constructs rules for CPGs as a second layer over ontology. These rules extend and enhance the semantic and inference capabilities of the ontology. Rules has many purposes in case base ontology. For example, it can enrich the domain by inferring new concepts and relations. It can check for consistency between features values and derive the dependency between patient attributes for predicting some features based on given ones. It can solve some case adaptation problems (by providing adaptation rules) and combine CBR with Rule Based

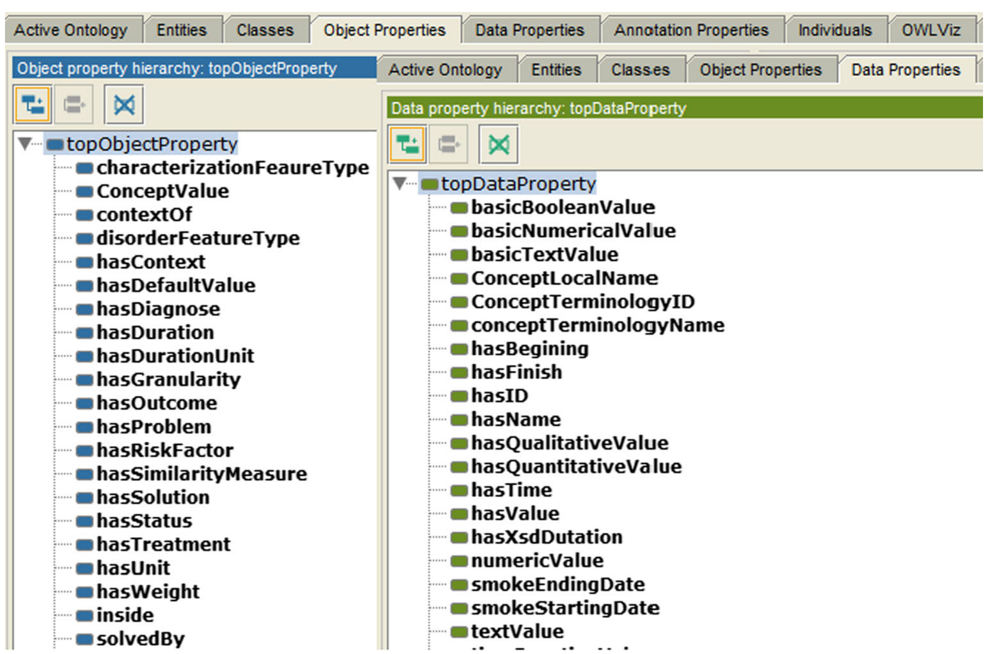
Table 4 Object and data properties

Property	Type	Domain	Range
hasFeatureType	object	Feature	FeatureType
hasQnValue	datatype	Test	string
hasProblem	object	Case	Problem
hasDiagnosis	object	Solution	Diagnose
hasTime	datatype	Validinstant	dateTime

Reasoning (RBR) in a hybrid system. SWRL is the rule language that expands OWL with the rule-based inference [43].

Figure 8 shows the complete picture of case base ontology and other participating ontologies such as patient ontology which is used to standardize the new case representation. Here we concentrate only on the ontologies, rules, jess engine and CBR. The case retrieval, adaptation, and retention are outside the scope of the paper. Case base ontology will be built using data and knowledge from many sources as EHR, domain expert, etc. The CPG rules are combined with ontology in the form of SWRL. We have two types of reasoning: ontology reasoners (as Pellet) which check the consistency and integrity on the ontology, and inference engines (IE) (as Jess) which do the inference using knowledge from ontology instances and rules. The ontology and SWRL rules are transformed to Jess facts and rules using Jess Bridge. The CBR system will be implemented as a knowledge base system in java using any of APIs as Protégé API, Jena API or OWL API and any of the query languages such as SQWRL or SPARQL. New cases will be standardized and entered to Jess IE to infer some hidden knowledge and the inferred knowledge and the

Fig. 7 object and data properties



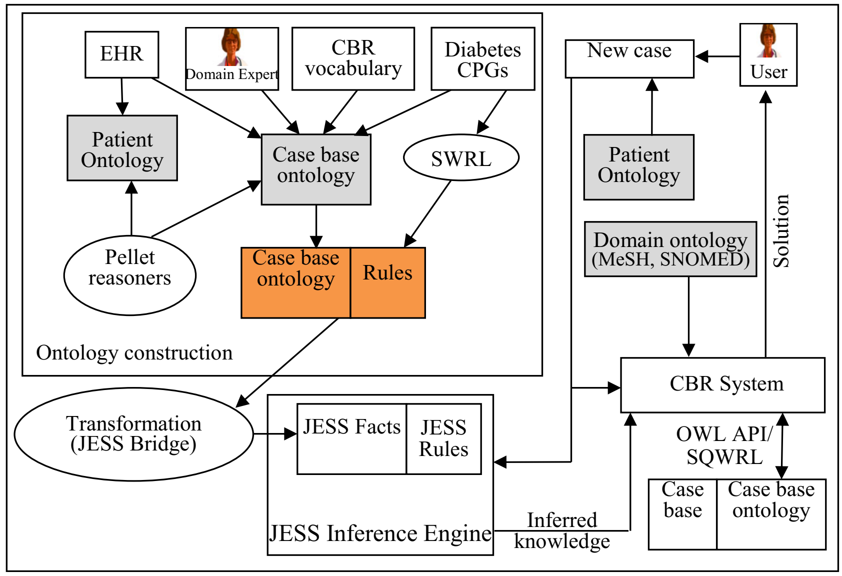


Fig. 8 Ontology integration with rules and role of jess in enriching the new case

standardized cases are entered to the CBR system to discover the solution. Some rules are shown in the following:

(1) We will benefit from SWRL build-ins to implement Allen temporal operators as equals, before, after, meets, metBy, overlaps, overlappedBy, contains, during, starts, startedBy, finishes and finishedBy and its negations.

(2) For checking the validity and consistency of values, we can add rules to check values of features. For instance:

Pregnant(?*x*)→*Female*(?*x*), which is if ?x is pregnant then it must be a female.

Age(?*x*) ^ *SWRLb*:*greaterThan*(?*x*, *18*)→*Adult*(?*x*), which defines the Adult concept.

(3) From diabetes CPG we can extract many rules, for instance:

Symptomatic(?*x*) ^ *FPG*(?*y*) ^ *hasValue*(?*y*, ?*v*) ^ *hasUnit*(?*y*, "*mmol*/*L*") ^ *SWRLb*:*greaterThan*(?*v*, *7.0*) ^ *Solution* (?*s*) ^ *Diagnose* (?*d*) ^ *hasDiagnose* (?*s*, ?*d*)→ *Diabetes*(?*d*), which defines a rule for diabetic patient.

Sub-step 6: Create instances

The last step is to create individual instances of classes which is knowledge about the real world. In our case, the individuals collectively represent cases which create the case base. Ontology is not affected by a change on the instance level as long as the corresponding concept and conceptual relationships are not modified. Ontology represents domain structure and does not contain instances [65]. It can be described as a "fat" domain model compared to database schema [66].

Ontology population (semantic annotation) will be postponed as it will be an automated process by generating SQL statements to extract data from EHR to the ontology and to extract rules from CPGs to ontology [67]. Some tools such as Data Master [68] can also automate the process. SQL statements may perform some calculations as temporal abstractions of patient data before being formulated into case. Several ontological APIs can be utilized in this step as Jena [69], as shown in Fig. 9. Figure 10 illustrates the relationship between ontological case base, EHR database and the physical instantiation on ontology in a case base. Diabetes case base ontology can be converted to database. This database's data fields will be linked to the EHR data fields for the ontology population. The interoperability between the EHR and case base databases will be preserved by the standardized terminological systems as SNOMED.

Step 8: Implementation

Here the implementation language and tool are selected, and the formal ontology will be generated. OWL 2 will be used as the language and Protégé 4.2 will be used as the tool for creating the formal ontology.

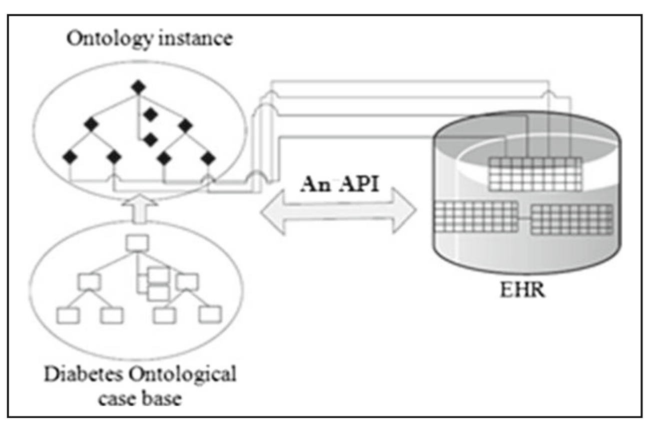


Fig. 9 Mapping EHR data to ontology instances

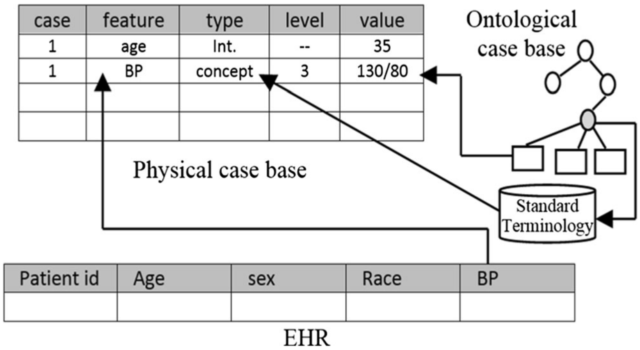


Fig. 10 Mapping between EHR, case base ontology and cases

Phase 3: Validation and evaluation of ontology

Ontology evaluation is an evolutionary process that goes on until the evolved ontology satisfies the design requirements. It is the technical judgment of the content of the ontology with respect to a frame of reference (i.e. CQs and specifications) during every phase, between phases and upon finishing their lifecycle [70]. It is a semi-automatic process conducted by domain experts, users and ontology engineers. The inputs are the resulting ontology and ORSD document. Its outputs are approval to use the ontology or return to previous steps and ontology evaluation document.

The evaluation process can be either (1) manually, where the development team checks the coverage of ontology to the CQs and requirements (completeness), or (2) computerized, where tools like ONTOClean [71] or reasoners like Pellet can check for inconsistency and redundancy errors. We will use the set of evaluation techniques and guidelines surveyed by Mariano Fernández-López [72]. The result of validation shows that our ontology is complete (all concepts, relations, and axioms are considered) and consistent (e.g. no loops).

Phase 4: Evolution, application and maintenance

In this phase, the ontology is used in the CBR system. Since the medical domain is changing very frequently, the EHR database changes rapidly and the development of CPG is continuous, therefore, maintenance of the ontology is critical. This phase updates and modifies the ontology model. As new concepts, rules or relations are added, changed or deleted, these processes must be done carefully to avoid semantic errors. While the execution of the ontology lifecycle is in process, there are ontology supporting activities such as knowledge acquisition, documentation, merging, alignment, configuration management, evaluation, integration and assessment.

Our final ontology for the diabetes case base is shown in Fig. 11. The resulting ontology is the accumulated result of the proposed ontology engineering methodology. The top level concepts are Case base, Case, Problem, Solution, Context and Outcome. Problems are described by Feature concept. Features can be Test, Characterization and Disorder. Each feature has Feature Type and Similarity Measure which describe the types (basic, concept) and the similarity measurement methodology of features respectively. Feature may have a temporal dimension using the Temporal Trend concept. When going down in the ontology, more details are found and the picture will be completed. This ontology is applicable to people with diabetes as well as those at risk of developing diabetes. The diagnosis of diabetes is the main point of this ontology. Creation of appropriate care plan for diet, medication, exercise and education will need some modifications in the solution part, which will be the focus of our future work.

Methodology evaluation

There are no medical ontology engineering methodologies for case-base ontology construction. Moreover, all the existing medical ontology engineering methodologies are very abstract to be compared with our methodology and not related to CBR [18, 32–35]. As a result, we have two choices. The first choice is to compare our method with a non-medical case-base methods as [37]. This choice is not realistic because the two fields have different features. For example, our method must handle CPGs, EHR database structure, temporal aspects, standard ontologies as ICD, and so on. The second choice is to compare ours with the general purpose methodologies as [2–10]. The evaluation will be done using the second choice, and methods in survey papers [19, 29] will be used.

First of all, our methodology has been applied in diabetes diagnosis as a case study, and the resulting OWL 2 case base ontology, created by Protégé, has been tested using some Protégé's plugin reasoners as Pellet, RacerPro and FaCT++. Our resulting ontology is logically consistent. Many evaluation approaches and criteria have been proposed [31]. Iqbal [29] has identified eight aspects to measure the maturity of a methodology and apply it on existing 15 methodologies. As shown in Table 5, by applying the criterion on our method and comparing it with Iqbal's results, we can find that our model outweigh other methods. The last row in Table 5 shows the number of methodologies support our selected criterion. For example, 2 out of 15 methodologies have used the Stage-based development, 2 out of 15 support collaborative construction, and so on.

Changrui and Yan [73] have identified nine criteria for ontology evaluation which overlap with Iqbal [84] list. They have concentrated on project management dimension, and they have applied this criteria on 3 methods. By evaluating ours with results from [73], we can find that our method is not less than other methods in this dimension.

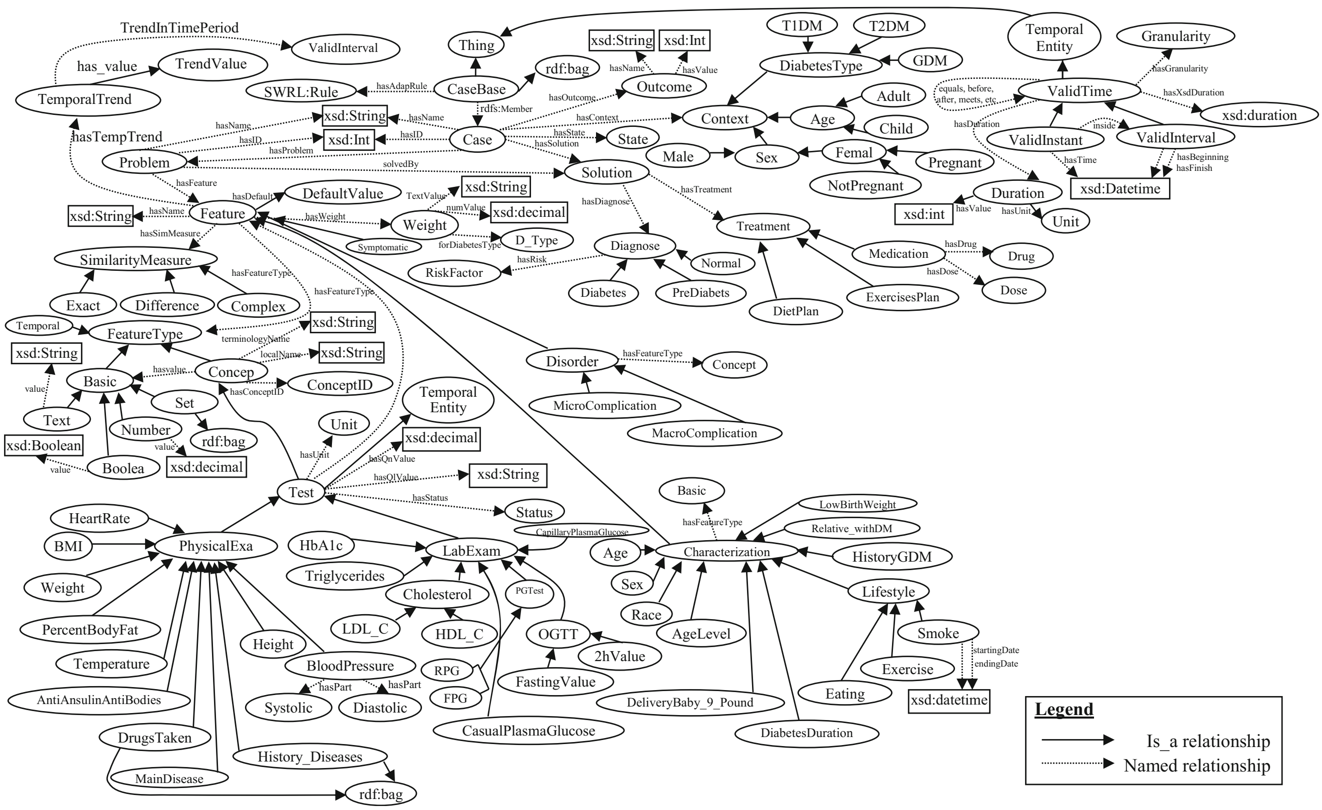


Fig. 11 Diabetes ontological case base

Another evaluation is based on Gómez-Pérez [74] who has established five criteria (consistency, completeness, conciseness, expandability and sensitiveness) for evaluating ontology. The ontology satisfied all criteria. *Firstly*, we have tried to obtain contradictory conclusions from valid inputs and we did not obtain contradictory results, so the ontology is Consistent. *Secondly*, depending on diabetes and CBR experts and based on competency questions, the Completeness has been proved by checking the completeness of the classes, class hierarchies and domains and ranges of classes. *Thirdly*, we have reviewed if the ontology has redundancies among classes or relations, but we could not find a redundancy. Therefore, the ontology satisfies the Conciseness property. *Fourthly*, in our methodology, the ontology can be extended without altering the set of well-defined properties. As a result, the ontology built is Expandable. *Finally*, we can make small changes in the definition of a class and these changes do not have a negative impact on the rest of the ontology, therefore, the ontology has low Sensitiveness.

Another method of evaluation has been followed by Casellas [3]. He has compared the existing methodologies number and types of activities to check the coverage and completeness of each method. He has divided the steps into Preparatory step, Development step, Evaluation step and Tool support. For each of these activities, our methodology is more detailed and more complete than existing methods because the number of activities is larger and more detailed. Because of space restriction we will not provide the detailed comparison table.

Future research directions

The main goal of our work is to build a complete CBR system. This preliminary work only builds the case base without

Table 5 Measurement of our methodology based on [84] established criterion

Type of development	Collaborative construction	Reusability support	Degree of application dependency	Life cycle recommendation	Strategies for identifying concepts	Methodology details	Interoperability support
Stage based	Yes	Yes	Medical-CBR	Yes	Middle out	Sufficient	Yes
2/15	2/15	10/15	5/15	3/15	11/15	1/15	3/15

handling uncertainty (i.e. fuzzy and statistics) and qualitative temporal data as these features are normal in medical domains [75].

Our future work will focus on completing the CBR cycle from case retrieval by introducing a new semantic case retrieval algorithm, to case adaptation and retention by introducing case adaptation algorithm and case indexing structures respectively [76].

Conclusion

In this paper, we have proposed an ontology engineering methodology for the CBR environment, and have applied it to diabetes management. This methodology has addressed the limitations of previous methodology by adding and modifying various steps. It has taken into account the diabetes management CPGs and EHR contents and resulted in an OWL ontology implemented in Protégé. This ontology for CBR case-base concentrates on the diagnosis of diabetes and provides the basic concepts for treatment as diet plan and medication plan. The case-base contains context, default and temporal concepts to be applicable for most chronic diseases. SWRL rules are added to enhance the semantic of case base and OWL, and to support the evidence based medicine of CPGs. Future work will concentrate on qualitative and uncertain temporal data processing and representation.

Acknowledgment This project was supported by King Saud University, Deanship of Scientific Research, College of Sciences Research Centre.

References

1. M. Sicilia (2014) Handbook of Metadata, Semantics and Ontologies, World Scientific Publishing Company.
2. F. Noy, D. McGuinness (2001) Ontology development 101: A guide to creating your first ontology, Technical Report SMI-2001-0880, Stanford University School of Medicine
3. Casellas, N., Methodologies. *Tools and Languages for Ontology Design, Legal Ontology Engineering* 3:57–107, 2011.
4. Lenat, D., and Guha, R., Building large knowledge-based systems: Representation and inference in the CYC project. *Artificial Intelligence* 61(1):41–52, 1993.
5. Uschold, M., and King, M., *Towards a methodology for building ontologies*. Artificial Intelligence Applications Institute, University of Edinburgh, pp. 15–30, 1995.
6. M. Fernández et al. (1997) METHONTOLOGY: From Ontological Art Towards Ontological Engineering, Symposium on Ontological Engineering of AAAI, 33–40
7. Y. Sure, R. Studer (2004) On-To-Knowledge Methodology (OTKM), Handbook on Ontologies (part 2), pp117–132
8. Chikofsky, E., Reverse Engineering and design recovery: A taxonomy. *IEEE Software Magazine* 7(1):13–17, 1990. doi:10.1109/52.43044.
9. A. Gómez-Pérez, M. Rojas, (1999) Ontological Reengineering and Reuse, Proceedings of 11th European Knowledge Acquisition Workshop (EKAW), 1621: 139–156
10. D. Calvanese, M. Lenzerini (2001) A Framework for Ontology Integration, Proceedings of Int. Semantic Web Working Symposium (SWWS), pp 663–667
11. Recio-García, J., et al., jcolibri2: A framework for building Case-based reasoning systems. *Science of Computer Programming* 79: 126–145, 2014.
12. Dufour-Lussier, V., et al., Improving Case Retrieval by Enrichment of the Domain Ontology. *Case-Based Reasoning Research and Development* 6880:62–76, 2011.
13. AbouAssali, A., et al., Heterogeneity in Ontological CBR Systems. *Successful Case-based Reasoning Applications I, Studies in Computational Intelligence* 305:97–116, 2010.
14. Subirats, L., and Ceccaroni, L., An Ontology for Computer-Based Decision Support in Rehabilitation. *Advances in Artificial Intelligence* 7094:549–559, 2011.
15. Disease Ontology, http://disease-ontology.org. Accessed April 2014.
16. Ontology of Glucose Metabolism Disorder, http://rgd.mcw.edu/rgdweb/ontology/view.html? acc_id=RDO:0001421. Accessed April 2014.
17. El-Sappagh, S., Elmogy, M., El-Masri, S., and Riad, A., *A Diabetes Diagnostic Domain Ontology for CBR System from the Conceptual Model of SNOMED CT.* Proceeding of the second IEEE Int. Con. on Engineering and Technology (ICET 2014), Cairo, 2014. Accepted.
18. Jaya, A., A Standard Methodology for the Construction of Symptoms Ontology for Diabetes Diagnosis. *International Journal of Computer Applications* 14(1):47–51, 2011.
19. R. Subhashini, J. Akilandeswari (2011) A survey on ontology construction methodologies, Int. J. of Enterprise Computing and Business Systems, 1(1), www.ijecbs.com
20. Gawich, M., Badr, A., Hegazy, A., and Ismail, H., A Methodology for Ontology Building. *International Journal of Computer Applications* 56(2):0975–8887, 2012.
21. M. Gr¨uninger, M. Fox (1995) Methodology for the design and evaluation of ontologies, International Joint Conference on Artificial Intelligence (IJCAI95), Workshop on Basic Ontological Issues in Knowledge Sharing
22. J. Correa (1996) Building and Reusing Ontologies for Electrical Network Applications, Proceeding of the 12th European Conference on Artificial Intelligence (ECAI96), 298–302
23. Valente, A., Building and (Re)Using an Ontology of Air Campaign Planning. *IEEE Intelligent Systems and their Applications* 14(1):27–36, 1999.
24. K. Kerremans, R. Temmerman, J. Tummers (2003) Representing multilingual and culture specific knowledge in a vat regulatory ontology: Support from the termontography methodology, Proceeding of the 1st International Workshop on Regulatory Ontologies and the Modeling of Complaint Regulations, 2889:662–674
25. Jarrar, M., and Meersman, R., Ontology Engineering – The DOGMA Approach. *Advances in Web Semantics I* 4891:7–34, 2009.
26. F. Sclano, P. Velardi (2008) Termextractor: A web application to learn the common terminology of interest groups and research communities, Proceeding of the 9th Conference on Terminology and Artificial Intelligence (TIA'07).
27. Gil, R., and Martin-Bautista, M., SMOL: a systemic methodology for ontology learning from heterogeneous sources. *J Intell Inf Syst*, 2014. doi:10.1007/s10844-013-0296-x.
28. M. Suárez-Figueroa, A. Gómez-Pérez, M. Fernández-López (2012) The NeOn Methodology for Ontology Engineering, Ontology Engineering in a Networked World, pp 9–34
29. Iqbal, R., An Analysis of Ontology Engineering Methodologies: A Literature Review. *Research Journal of Applied Sciences Engineering and Technology* 6(16):2993–3000, 2013.

30. Castillo-Barrera, F., et al., A method for building ontology-based electronic document management systems for quality standards—the case study of the ISO/TS 16949:2002 automotive standard. *J. of Applied Intelligence* 38:99–113, 2013. doi:10.1007/s10489-012-0360-1.
31. Haghighi, P., et al., Development and evaluation of ontology for intelligent decision support in medical emergency management for mass gatherings. *Decision Support Systems* 54:1192–1204, 2013.
32. Zhao, H., and Passi, K., Semantic Web and Ontology Engineering for the Colorectal Cancer Follow-Up Clinical Practice Guidelines. *Health Information Science* 7798:53–64, 2013.
33. A. Rahimi, et al. (2012) Developing an Ontology for Data Quality in Chronic Disease Management, 24th International Conference of the European Federation for Medical Informatics, Quality of Life through Quality of Information (MIE2012).
34. D. Forbes, P. Wongthongtham, J. Singh (2012) Development of Patient-Practitioner Assistive Communications (PPAC) Ontology for Type 2 Diabetes Management, CIHealth 2012: Proceeding of the Workshop on New Trends of Computational Intelligence in Health Applications, pp 43–54
35. D. Sutton, A. Aldea, C. Martin (2011) An ontology of diabetes self-management. Proc. of the first int. workshop on Managing interoperability and complexity in health systems, pp 83–86.
36. Chen, J., Su, S., and Chang, C., Diabetes Care Decision Support System. *International Conference on Industrial and Information Systems (IIS)* 1:323–326, 2010.
37. Guo, Y., Hu, J., and Peng, Y., A CBR system for injection mould design based on ontology: A case study. *Computer-Aided Design* 44:496–508, 2012.
38. The Protégé Ontology Editor and Knowledge Acquisition System (2013). http://protege.stanford.edu.Accessed 10 May 2013
39. OWL 2 Web Ontology Language (2013). http://www.w3.org/TR/owl2-overview.Accessed 30 May 2013
40. Yang, S., Developing an energy-saving and case-based reasoning information agent with Web service and ontology techniques. *Expert Systems with Applications* 40(9):3351–3369, 2013.
41. Siorpaes, K., and Simperl, E., Human Intelligence in the Process of Semantic Content Creation. *Journal World Wide Web* 13(1–2):33–59, 2010. doi:10.1007/s11280-009-0078-0.
42. Canadian Journal of Diabetes, Canadian Diabetes Association 2013 Clinical Practice Guidelines (2013), http://guidelines.diabetes.ca, Accessed 20 May 2013
43. I. Horrocks, et al. (2004) SWRL: A semantic Web Rule Language Combining OWL and RuleML, W3C Member Submission, www.w3.org/Submission/SWRL, last seen June 2013.
44. W3C, Time Ontology in OWL (2013) http://www.w3.org/TR/owl-time.Accessed 11 May 2013.
45. O'Connor, M., and Das, A., *A Method for Representing and Querying Temporal Information in OWL, third International Joint Conference*. BIOSTECValencia, Spain, pp. 97–110, 2011. doi:10.1007/978-3-642-18472-7_8.
46. Batsakis, S., SOWL: A Framework for Handling Spatio-Temporal Information in OWL, Ph.D. dissertation, Technical Univ. of Crete (TUC), Rule-Based Reasoning, Programming, and Applications. *Lecture Notes in Computer Science* 6826:242–249, 2011. doi:10.1007/978-3-642-22546-8_19.
47. Lin, Y., and Sakamoto, N., Ontology Driven Modeling for the Knowledge of Genetic Susceptibility to Disease. *Kobe J. Med. Sci.* 54(6):E290–E303, 2008.
48. El-Sappagh, S., El-Masri, S., Riad, A., and Elmogy, M., Electronic Health Record Data Model Optimized for Knowledge Discovery. *IJCSI International Journal of Computer Science Issues* 9(5):329–338, 2012.
49. Gu, D., et al., Intelligent Technique for Knowledge Reuse of Dental Medical Records Based on Case-Based Reasoning. *Journal of medical systems* 34(2):213–222, 2010.
50. N. Kings, J. Davies (2009) Semantic Web for Knowledge Sharing, Semantic Knowledge Management, pp 103–111. doi: 10.1007/978-3-540-88845-1_8
51. Up to Date Clinical Practice Guideline (2014) http://www.uptodate.com. Accessed 2 April 2014
52. National Guideline Clearinghouse (2014), www.guideline.gov, Accessed 1 April 2014
53. National Institute for Health and Care Excellence (2014) http://www.nice.org.uk.Accessed 2 April 2014
54. Canadian Journal of Diabetes, Canadian Diabetes Association 2013 Clinical Practice Guidelines, http://guidelines.diabetes.ca. Accessed 20 May 2013
55. Ministry of health in Malaysia, Clinical practice guideline, management of type 2 diabetes mellitus, 4th edition, http://www.diabetes.org.my, 2009. Accessed 20 May 2013
56. American diabetes association, standards of medical care in diabetes, diabetes care, volume 36, supplement 1, January 2013, http://care.diabetesjournals.org. Accessed 20 May 2013
57. VA/DOD Evidence-based Practice, VA/DOD clinical practice guideline management of diabetes mellitus (2010), http://www.healthquality.va.gov. Accessed May 2013
58. AACE Diabetes care plan Guidelines, AACE Task Force for Developing a Diabetes Comprehensive Care Plan, Endocrine practice volume 17, suppl (2), 2011
59. A. AlJarullah (2011) Decision Tree Discovery for the Diagnosis of Type II Diabetes, IEEE international conference of innovations in information technology, pp 303–307
60. J. Juarez, et al. (2007) Case representation ontology for case retrieval systems in medical domains, ACM Proceedings of the 25th IASTED International Multi-Conference: artificial intelligence and applications, pp 168–173
61. Pan, F., and Hobbs, J., *Time in OWL-S, In Proceedings of the AAAI Spring Symposium on Semantic Web Services*. Stanford University, CA, pp. 29–36, 2004.
62. Aissam, B., et al., From UML class diagrams to OWL ontologies: a Graph transformation based Approach. *ICWIT, CEUR Workshop Proceedings* 867:330–335, 2012.
63. Dragan Gašević, et al. (2004) Converting UML to OWL Ontologies, ACM Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters (WWW Alt. '04), pp 488–489. doi:10.1145/1013367.1013539
64. Mukesh, J., et al., Diabetes Detection and Care Applying CBR Techniques. *International Journal of Soft Computing and Engineering (IJSCE)* 2(6):132–137, 2013.
65. Meersman, R., *The use of lexicons and other computer-linguistic tools in semantics, design and cooperation of database systems, Proceedings of the Second International Symposium on Cooperative Database Systems for Advanced Applications (CODAS99)*. Springer, Verlag, pp. 1–14, 1999.
66. P. Spyns (2005) Object role modelling for ontology engineering in the DOGMA framework, OTM'05 Proceedings of the 2005 OTM Confederated international conference on the Move to Meaningful Internet Systems, Springer-Verlag, 3762:710–779
67. Branden, M., et al., Integrating case-based reasoning with an electronic patient record system. *J. of Artificial Intelligence in Medicine* 51:117–123, 2011.
68. Data Master (2013), http://protegewiki.stanford.edu/index.php/DataMaster. Accessed 15 June 2013
69. Apache Jena (2014), http://jena.sourceforge.net. Accessed 4 April 2014
70. G´omez-P´erez, A., *Ontological engineering: With examples from the areas of knowledge management, e-commerce and the Semantic Web(Advanced information and knowledge processing)*, First Editionth edition. Springer-Verlag New York, Inc, Secaucus, 2003.

71. OntoClean Ontology (2013) http://protege.stanford.edu/ontologies/ontoClean/ontoCleanOntology.html. Accessed Feb 2013
72. M. Fernández-López (2013) Deliverable 1.4: A survey on methodologies for developing, maintaining, evaluating and reengineering ontologies, IST Project IST-2000-29243 OntoWeb, http://www.kde.cs.uni-kassel.de/stumme/papers/2002/OntoWeb_Del_1-4.pdf
73. Changrui, Y., and Yan, L., Comparative Research on Methodologies for Domain Ontology Development, Advanced Intelligent Computing Theories and Applications. *With Aspects of Artificial Intelligence* 6839:349–356, 2012.
74. S. Bechhofer, G. CA, I. Horrocks (2001) DAML+OIL is not enough, pp 151–159. DBLP:conf/semweb/2001swws, SWWS
75. Ml, H., Yh, H., Wm, L., Rk, L., and Th, W., Usage of case-based reasoning, neural network and adaptive neuro-fuzzy inference system classification techniques in breast cancer dataset classification diagnosis. *J Med Syst.* 36(2):407–14, 2012.
76. Sharaf-El-Deen, D., Moawad, I., and Khalifa, M., A New Hybrid Case-Based Reasoning Approach for Medical Diagnosis Systems. *J Med Syst.* 38(2):9, 2014.