International Journal of Intelligent Computing and Information Science

FOORC: A FUZZY ONTOLOGY-BASED REPRESENTATION FOR OBESITY RELATED CANCER KNOWLEDGE

M. A. Elhefny
Information Systems Dept., Faculty of Computers and Information, Mansoura University, Egypt
mohhed20@mans.edu.eg

M. Elmogy
Information Technology Dept., Faculty of Computers and Information, Mansoura University, Egypt
melmogy@mans.edu.eg

A. A. Elfetouh
Information Systems Dept., Faculty of Computers and Information, Mansoura University, Egypt
elfetouh@gmail.com

F. A. Badria
PharmacognosyDept., Faculty of Pharmacy, Mansoura University, Egypt
faridbadria@gmail.com

Abstract- Obesity has a tight relationship with increased risks of different cancer types, such as Colorectal, Ovarian, Female Breast, Gallbladder, Adenocarcinoma, Kidney (Renal-Cell), Liver, and Pancreatic. It can also lead to some other diseases like diabetes and heart diseases. This paperproposes a fuzzy ontology that is based on OWL 2to represent the Obesity Related Cancer (ORC) domain knowledge. The diseases taxonomy isconstructed using the standard Disease Ontology. The presented FuzzyOntology includes more concepts than in crisp one and copes with the domain linguistic variables. It allows the users to query the Fuzzy Dl reasoner for element and get them back the fuzzy ontology for that element. It is expected to be good practice for ontologists and knowledge engineers in medical field aiding them to solve the overlapping concepts, linguistic variables, and reasoning problems by building their fuzzy ontologies. Building FOORC as an open ontology is a first attempt to organize information related to the obesity and cancer diseases in a formalized, structured manner that both physicians and intelligent systems can exploit it in knowledge sharing, reusability, and reasoning.

Keywords: Fuzzy Ontology, Obesity Related Cancer, OWL 2, Knowledge Representation, Disease Ontology.

1. Introduction

Nowadays, we are facing endless needs for the human's expertise in all specialized fields, such as medical/healthcare, education, finance, fault diagnosis, industrial applications, and business. In addition to the need to take the actual decision at the appropriate time based on well formalized and specialized knowledge. Therefore, it is critical to represent the knowledge efficiently using Ontologies via integrating the scattered informational resources. Practically, there are problems while building ontologies like the linguistic variables, overlapping concepts, and the state of uncertainty that exist in the domain. From a medical view, the domain of ORC is a critical topic for research. There is a strong relationship between obesity as an overnutrition disease and different types of cancers. There are plenty of death cases because of cancers and the bad body reaction to cure resulted from the morbid obesity.
The developed Ontology will allow the users to query it for element and get him back the fuzzy ontology for that element by using the Fuzzy Dl reasoner[1]. The developed Ontology was encoded using Protégé 4.3[2] in OWL 2-DL format and thenwas integrated with a pre-developed fuzzificationplug-in[3].

[1]http://nemis.isti.cnr.it/~straccia/software/fuzzyDL/fuzzyDL.html
[2] http://protege.stanford.edu/

The diseases hierarchy was built upon Disease Ontology (DO). We preferred using DO hierarchy [2] as it is a human diseasestandard Ontology that semantically incorporates ailments and therapeutic vocabularies by broad cross mapping of its terms according to ICD, OMIM, MeSH, NCI's thesaurus,and SNOMED.

In this paper, our ontology is called "FOORC" (fuzzy ontology for obesity-related cancer). Itis defined as the representation of knowledge and data relating to obesity and cancer diseases, risk factors, symptoms, diagnosis, and treatment while taking into account the fuzzy aspects (linguistic variables and uncertainties) that may be present in this medical domain.

Therest of the paper is organized as follows.Section1 introduces the preliminaries about Ontologies and the standard web ontology language. Section 2 presents basic information to understand the medical domain, Section 3 shows some of researchers' work in medical ontologies and fuzzy, while the proposed work to construct FOORC within 3 phases is introduced in Section 4.Section 5 discusses the results. Finally, in Section 6, we end with a conclusion and future prospects.

1.1. Basic Definitions

The Semantic Web is a Web extension to enable individuals to share contents beyond the limits of the Websites and applications [3]. It means to transform the present Web (with unstructured and semi-structured documents) into a "Web of data", and its stack expands on the Resource Description Framework (RDF) [4].

Ontology is the main method torepresent, share, and reuse of knowledgeon the Semantic Web. It can be described as a domain conceptualization into a human intelligible, machine-clear form involving axioms, attributes, relationships, and entities [5]. W3C defined Ontologies as formalized vocabularies of terms that cover a particular domain and are shared by a users' community. In the ontology,the definition termis specified by its associations with the other terms [7]. The domain ontology is a format of anacceptable computer representation of knowledge about a part of an abstract or a real world [8].

The Fuzzy Ontology can be described as an extended domain Ontology to overcome the uncertainty, reasoning, and retrieval problems. The Fuzzy Ontologies are qualified to deal with fuzzy knowledge [4], [9].

1.2. Medical Ontologies

Instead of reinventing the wheel and start from scratch, there are different free and open Ontologies and medical projects that can be effectively reused, such as GALEN[4], MeSH[5], SNOM[6], Gene Ontology[7], Bio-Ontology[8], OBO Foundry[9], and DOID [2].

1.3. OWL: The Web Ontology Language

The OWL is used to describe ontologies. It is based on XML, and can be divided into three language levels (OWL DL, OWL Lite, and OWL Full) [5].

[3]http://nemis.isti.cnr.it/~straccia/software/FuzzyOWL
[4]http://www.opengalen.org
[5]http://www.nlm.nih.gov/mesh/meshhome.html
[6]http://www.snomed.org
[7]http://www.geneontology.org
[8]http://www.bioontology.org
[9]http://www.obofoundry.org

The built ontologies by OWL 2 are stored as Semantic Web documentsand support adding properties, classes, individuals, and data values. They are mainly exchanged as RDF documents and can be utilized alongside written information in RDF [7].

Figure 1 shows an overview of the OWL 2 language and the relationships among its main building blocks. The centered ellipse can be considered as an RDF graph or an abstract structure. Different particular syntaxes can be used for Ontologies exchange, at the top part. Two semantic specifications define the OWL 2 Ontologies meanings, at the bottom part. In their work with OWL 2, themajority of the users need only one syntax and one semantics [7].

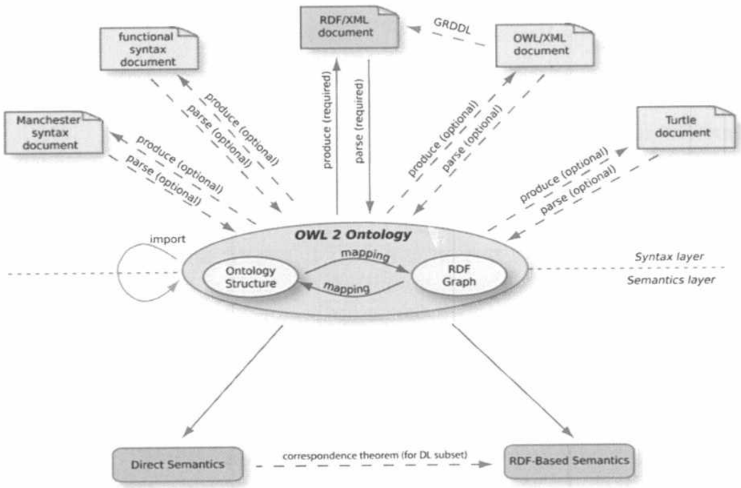


Fig.1: The structure of the OWL 2 [7].

2. ORC Domain

2.1. Obesity and Cancer Risk

Recently, the percentage of the overweighted and obese adults and children has significantly increased. Obesity is a condition in which a human has an abnormally high and unhealthy proportion of body fat. Obese people are more exposed to cancers as well as coronary heart disease, stroke, high blood pressure, diabetes, and some other chronic diseases.The rate of cases ascribed to obesity varied for different types of cancers but recorded 40%of some cancers, specifically Endometrial and Esophageal cancers [12].

For obesity measurement, scientists use the Body Mass Indexthat is computedby dividing a man's weight byhis squaredheight, using kilograms and meters to measure weight and height. The guidelines of NIH[10] considered the 20 years old adults and older with their BMI values into the defined categories, as shown in Table 1.

Calle and Kaaks [13] stated that in the US, about two-thirds of adults were obese or overweight by the year 2000, and 300 million adults had obesity around the world.

[10] http://www.nih.gov/- The National Institutes of Health

Less attention was given to the strong association between the cancertypes and the causing obesity than its cardiovascular effects. In the US, it was assessed that nearly 20% of all cancer deaths can be credited tooverweight and corpulence. There is a defined relation between the obesity and the high levels of Insulin. Table 2 shows the relative risk of different BMI ranges with different cancer types throughout a statistical study made for EU and US populations.

Table 1: The Guidelines of BMI [12].

BMI Categories	BMI
Obese	30.0 and above
Overweight	25.0 to 29.9
Normal	18.5 to 24.9
Underweight	Below 18.5

2.2. Tumor Markers and Reference Ranges

Tumor markers are produced by cancers, different cells of the body because of malignanttumors, or certain benign (noncancerous) conditions. They can be detected in blood, urine, stool, tumor tissue, other tissues, or bodily fluids. They are utilized to help distinguish, analyze, and deal with a few types of cancer. The raised level of a tumor marker may be a diagnostic factor of cancer existence, but alone it is not sufficient to diagnose cancer. Therefore, other tests, such as biopsies, are usually combined with measurements of tumor markers to diagnose cancer [14].

According to National Cancer Institute (NCI) [14], Table 3 summarizes the required tumor markers tests for each cancer type. We focused on the most common cancer types in Mansoura University Hospitals, Mansoura, Egypt.

We asked our experts from Mansoura University Hospitals to determine which required properties and tests for each patient to diagnose if he is cancerous or not, especially in association with being obese. They specified (Age, Gender, BMI, Glucose tests [2hPG, FPG, HbA1C], Lipid profile [Total Cholesterol, TG (triglyceride)]) in addition to other diagnostics that will be mentioned later. Table 4 indicates the reference ranges of tumor markers that physicians use to determine the possibility of cancer.

Tables 5 and 6list the reference ranges for glucose and cholesterol levels, which are used by the Egyptian experts to judge the patient condition, respectively. The standard reference values for these glucose and cholesterol levels can be found in[11, 12]. For cholesterol, we found that the US and some other countries use the same ranges while Canada and most of theEurope use different ranges. Our Egyptian ranges are closer to US ranges. The required tests are as follows:**FPG**(Fasting Plasma Glucose), **2hPG**(2-hour Plasma Glucose), **HbA1C** (Glycated hemoglobin), **AFP**(Alpha-fetoprotein), **CEA**(Carcinoembryonic antigen), **Kras**(KRAS mutation analysis), **ER** (Estrogen receptor), **PR** (Progesterone Receptor), **T:** (Triglyceride), and **T.Chol** (Total Cholesterol).

3. Related Work

In ORC domain, there are some issues with the overlapping concepts/terms, linguistic variables, and the uncertainty circumstances that exist and need to be addressed and accommodated while representing its knowledge. Our work focuses on integrating the Fuzzy logic while building the ORC Ontology using

[11] http://labtestsonline.org/understanding/analytes/glucose/tab/test/

[12] http://www.mayoclinic.org/diseases-conditions/high-blood-cholesterol/in-depth/cholesterol-levels/art-20048245

OWL 2 and Protégé to formalize the ORC domain. It introduces more efficient knowledge semantically representation of the ORC domain andprovides reasoning capabilities. It is useful to physicians, experts or medical researchers, and computer scientists who are interested in this domain of knowledge.

Table 2: The Obesity-related cancer [13].

Type of cancer	Relative risk* with BMI of ≥ 30 kg/m^2	Relative risk* with BMI of 25-30 kg/m^2	PAF (%) for US population	PAF (%) for EU population
Colorectal (men)	2.0	1.5	35.4	27.5
Colorectal (women)	1.5	1.2	20.8	14.2
Female breast (postmenopausal)	1.5	1.3	22.6	16.7
Endometrial	3.5	2.0	56.8	45.2
Kidney (renal-cell)	2.5	1.5	24.5	31.1
Oesophageal (adenocarcinoma)	3.0	2.0	52.4	24.7
Pancreatic	1.7	1.3	26.9	19.3
Liver	1.5-4.0	ND	ND	ND
Gallbladder	2.0	1.5	35.5	27.1
Gastric cardia (adenocarcinoma)	2.0	1.5	35.5	27.1
* Relative risk estimates are summarized from the literature cited in the main text.				

[13] The two sets of PAFs (population attributable fractions) have been computed using these relative risks.

Parry [6] presented a Fuzzy Ontology technique for medical document retrieval. To enhance any ontology searching tool, he made a mapping between query terms and individualsof an Ontology. In any case, the relative significanceof a specificmapping to an overloaded term might be diverse for various users, and this information is essential for the reasonable fulfillmentofinquiry.For every user or a users' group, the Fuzzy Ontology was used by adding a degree membershipvalue to every "overloaded" term. Then, from Ontology mediated search, the retrieved documents can give the probable information request. Parry's approach addressed the "overloaded" terms (the same terms occur more than once) not the "overlapping" terms (the similar concepts in meaning that have different degrees of usage), but it was a starting point to ensure the concept of fuzzy use of medical ontologies.

Chen et al. [5] introduced Fuzzy rulesbased anti-diabetic drugs recommendation system, Fuzzy reasoning techniques, and the Ontologyof anti-diabetic drugs for medicine recommendation. Their experimental results showed that the drugs selection achieveda good performance. They used the clinical practice data of the American Association of Clinical Endocrinologists Medical Guidelines for 20 patients, and according to six attributes/tests. They used tools like Protégé, OWL DL, Joseki server software, and SPARQL as a query language. They used the old version of OWL, and they integrated the fuzzy logic into the reasoning system not in building the ontology itself.

Table 3: The cancer types and their related tumor markers.

[13]http://www.medscape.com/viewarticle/487381_4

Cancer Type	Required tumor markers tests
Ovarian	CA-125, * HE4, * 5-Protein signature (Ova1)
Colorectal	* BRAF mutation V600E, Carcinoembryonic antigen (CEA), KRAS mutation analysis
Pancreatic	CA19-9
Liver	AFP
Breast	CA15-3/CA27.29, Estrogen receptor (ER)/progesterone receptor (PR),Carcinoembryonic antigen (CEA), HER2/neu, * Plasminogen activator inhibitor (PAI-1) and Urokinase plasminogen activator (uPA), * 21-Gene signature (Oncotype DX), * 70-Gene signature (Mammaprint)
* This tumor marker test is not applied in Egypt.	

Table 4: Tumor Markers Reference Ranges [15].

CA-125	AFP	CEA	Kras	CA19-9	CA15-3	ER PR	HER2/neu
0-35 U/ml	Low levels present in both men & non-pregnant women (0-15 IU/ml); generally results >400 are caused by cancer (Half-life 4-6 days)	<2.5 ng/ml in non-smokers <5 ng/ml in smokers Generally, > 100 signifies metastatic cancer	1 % is the cut-off level between nonmutant and mutant Kras	٣٧ >U/ml is normal > 120 U/ml is generally caused by tumor	<31 U/ml (30% of patients have an elevated CA 15-3 for 30-90 days after treatment, so hold up 2-3 months after beginning new treatment to check)	1 % staining is the cut-off point Above this, it is positive	**10 % staining is the cut-off point**

Table 5: The Diabetes reference ranges.

	Oral Glucose Tolerance Test (mg/dL)	FPG (mg/dL)	HbA1C
Diabetic	200 or above	126 or above	6.5 or above
Prediabetic	144- 199	100 - 125	5.7-6.4
Normal	139 or below	99 or below	About 5.0

Table 6: The Lipid Profile reference ranges.

Triglycerides (mg/dL)	60 - 160
Total cholesterol (mg/dL)	0 - 200

Alfonse et al. [16] introduced a developing method to build anOntology for liver cancer using Protégé and OWL-DL format to encode their Ontology. Their Ontology was expected to benefit experts or medical researchers who need such knowledge be semantically represented. The data was acquiredfrom cancer.gov, medicinenet.com, and cancer.net. They did not make integrate of Fuzzy logic in their Ontology, and they used the older version of OWL (OWL 1).

Moawad et al. [17] presented building a Viral Hepatitis Ontology based on OBR framework. A three stages methodology; Acquisition, Validation, and OWL Representation was used. In designing the ontology, the bottom-up approach was used and in implementation, they had used Protégé.

Abdel-Badeeh and Hisham [18] introduced a five steps approachfor developing a Web-based Ontology of Knowledge Engineering. We favored to apply this methodology rather than in [17] and coming [19] as it was found simple, clear, logical and more proper for developing the ORC Ontology.

To develop an Ontology for breast cancer domain, Fatimatufaridah et al. [19] used a hybrid approach. They followed a three phases methodology that included three stages: 1- preparation, 2- hybrid Ontology process (a- build global Ontology, b- build local Ontology, c- mapping between global and local Ontologies, d- mapping between data sources and local Ontology), and 3- development of Ontology. Their knowledge resourcesare involved a medical officer as a domain expert, and documentationwastaken from journals, articles, and Websites. They did not use Fuzzy logic in their work.

Torshizi et al. [8] presented a savvy hybrid system based on Fuzzy-ontology that determines the severitylevel and recommends the treatment for Benign Prostatic Hyperplasia (BPH). They used Ontologies for expert's knowledge representation. They used brainstorming procedure among experts. They used Fuzzy logic to make inference on rule bases using Fuzzy variables that are in the form of if-then rules. They did not work directly on the Ontology itself; they needed to transform it into if-then rules to be usedwithin a Fuzzy system.

Elhefny et al. [20] introduced building a crisp Ontology for representing the Obesity-Related Cancer domain that included Diseases, Diagnosis, Treatment, and Symptoms Classes. They had constructed their Ontology usinga simple methodology within Protégé building environment; itwasformattedin OWL 2-DLsyntax. It was helpful to reuse their ontology as the core of the first phase of developing FOORC. We extended their Ontology by adding more concepts and terms, instances, and properties. Then, we integrated Fuzzy logic to handle the overlapped concepts, linguistic variables, and uncertainty circumstances to get a more efficient representation of the domain knowledge.

4. The Proposed Obesity Related Cancer (ORC) Fuzzy Ontology

It is important to represent linguistic variables and overlapped concepts of Semantic Web Languages in a standard way. It can be performed by either developing the current Semantic Web Languages or by a procedure representing such information within current standard languages and tools.In our suggested framework, we utilized the last approach withinOWL 2 to represent the ORC Ontology to meet the mentioned needs.Figure 2 shows the block diagram of the proposed framework to build the FOORC.

We constructed our work in three main phases. First, we built the typical Ontology by using Protégé 4.3 to allow reasoning using standard Ontology reasoners, validation, and evaluation. The output of this phase is a validated Ontology for obesity-related cancer with no fuzzy. Second, we represented the Fuzzy data types and overlapping concepts by using OWL 2 and Fuzzy annotation properties through FuzzyOWL 2 plug-in [1]. Finally, we reasoned with/ queried the constructed Fuzzy Ontology using fuzzyDL.

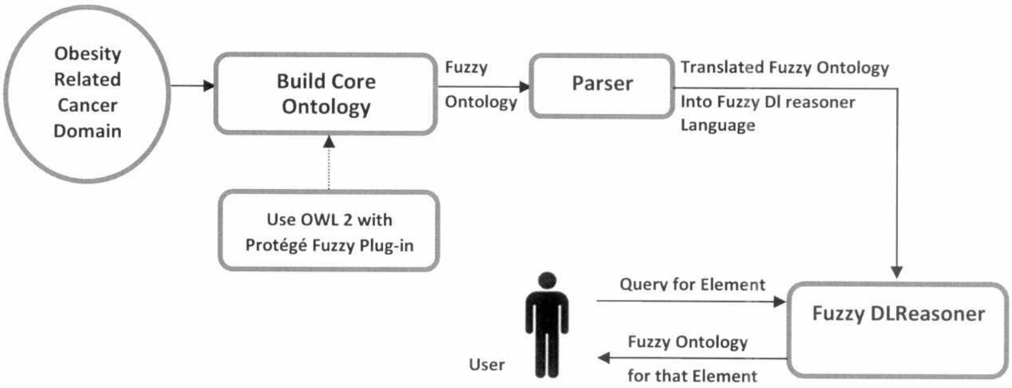


Figure 2: The proposed framework for building the FOORC.

4.1. Phase 1: Building Obesity Related Cancer Ontology

4.1.1. Building Methodology

We began building our Ontology by analyzing the vocabularies of Obesity-related cancer domain. Then, we identified the most commonly used terminologies by physicians from Mansoura University Hospitals. Several official sources were used like [12], [13], [14], [15], and [21]. In "disease" superclass, we made our class based on "DO" [2] hierarchy and terms. We followed a methodology with five processes used in [20], as shown in Figure 3, and pre-built ORC Ontology as the core to start adding the new items to it. We consulted the domain experts to validate the classification trees, edit terminologies, add other classes, and determine concepts synonyms. Our first layout of building Ontology is shown in Figure 4 with classes, properties, and relationship.

4.1.2. Ontology Structure

Figure 5 displays the "DO/DOID" visualized hierarchy for the Obesity disease class (as an example). After applied the "DO" terms and hierarchy to Obesity and Cancer disease classes, we represented them using protégé.

"DO" stated that the Ovarian Cancer has synonyms such as Malignant Ovarian tumor, neoplasm of ovary (disorder), ovarian neoplasm, ovary cancer, atumor of the ovary, ... etc.To reduce the time and effort, we considered the term "Ovarian Cancer" (that is a subclass of Female_reproductive_organ_cancer) as a member (individual) of its superclass Cancer, without taking into account the very detailed synonyms and sub-items, as the experts recommended, so we also did for some other diseases.

ORC Ontology consists of **five superclasses**:Disease, Medical Intervention, References, Patient, and Country [20]. **Medical Intervention class** includes both Diagnosis that in turn involves Cancer Diagnosis and Obesity Diagnosis subclasses. Treatmentclass includes Cancer Treatment and Obesity Treatment subclasses. **References class** includes Risk Factors subclass that in turn involves Cancer Risk Factors and Obesity Risk Factors subclasses and Symptoms subclass that in turn involves Cancer Symptoms and Obesity Symptoms subclasses. **Patient class** involves Male Patient and Female Patient subclasses. **Country class** involves Egypt. **Our used relationships are**is_a, has_Disease, IsLocatedIn, ResultsIn, hasCauses, hasSymptoms, diagnosedBy, treatedBy, see Table 7. Table 8 displays the classes' individuals. The hierarchy of the full ORC Ontology classes is shown in Figure 6 throughout protégé environment. Figure 8shows an excerpt ofkclasses, data and object properties in the ontology (partial).

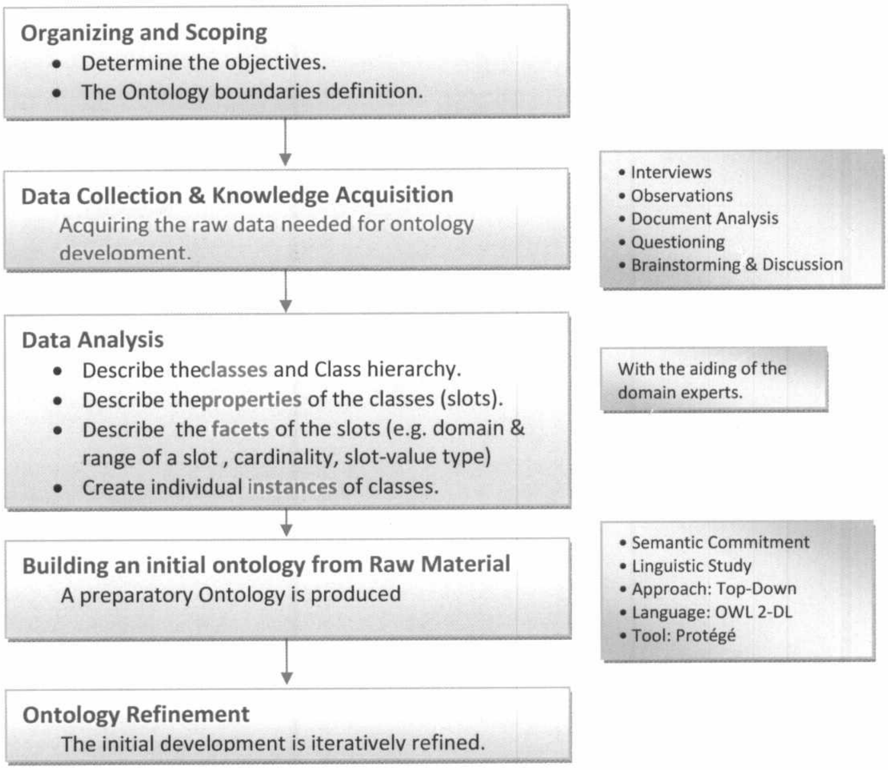


Fig.3: ORC Ontology building methodology [20].

4.1.3. Some Domain Considerations

Initially, we thought to add cancer "Staging" Class, and then we found that most of the staging[16] tests were involved in the "Diagnosis" Class, as shown in Figure 7.In Tumor_markers_tests class, we considered the Tumor markers for our concerned cancer types (Pancreatic, Colorectal, Ovarian, Liver, and Breast cancers) according to NCI [14]. Then, we refined them to fix what are done in Egypt as in Table 3. We considered adding Diabetes as a strong relationship between Obesity and Diabetes Mellitus exists. It is observed the higher BMI (Morbid Obesity) leads to higher blood glucose levels (Type 2 Diabetes). For cancer, we selected the most common cancers that exist in Mansoura University Hospitals as mentioned in [13]. The Mansoura University Hospitals serve patients from all the Egyptian cities.We found some terms were so close to each other, such as (Colon Cancer, **Colorectal**, Colon and rectum, Colon Adenocarcinoma), (**Liver**, Hepatocellular), and (**Tumor**, Neoplasm) that might be treated as overlapping concepts. Our information resources included other Websites like[14,15,16] for Obesity, Cancer, and Cancer Staging, respectively.

---

[14]http://www.nhlbi.nih.gov/health/health-topics/topics/obe/causes.html

[15]http://www.cancerresearchuk.org/cancer-info/cancerandresearch/all-about-cancer/what-is-cancer/what-causes-cancer

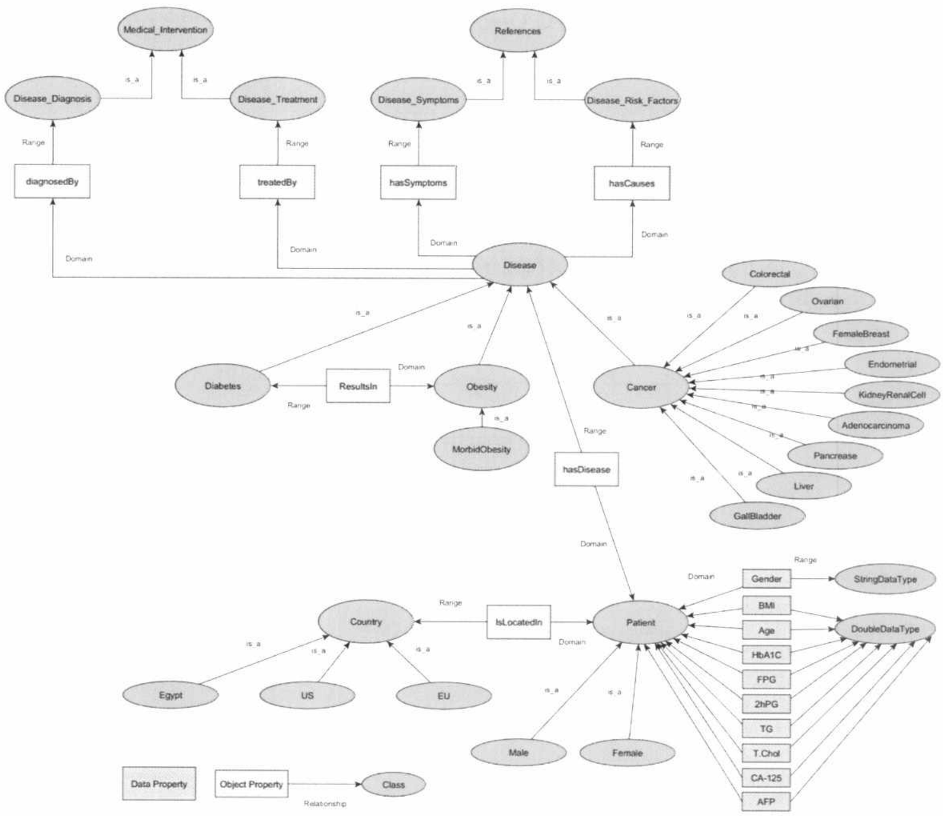


Figure 4: The Ontology building initial layout.

4.1.4. Ontology Validation and Metrics:

We tested the Ontology consistency by using Protégé built-in reasoner(s), and its validation by the experts' review. Statistics and format validation were made using the online tool provided by Manchester University, "Ontology Metrics"[17, 18] for statistics and[19] for format validation. Figure 9 indicates the general metrics of core (crisp) Ontology using the online tool. The metrics for both the typical (core) Ontology [20] and our modified one are displayed in Figure 10 using the built-in metrics tool in Protégé. The syntax validation of our Ontology to OWL 2-DL is reported in Figure 11 using Manchester University Validation Tool.

4.2. Phase 2: ORC Ontology Extension - Adding the Fuzzy Part Using Annotation Properties

We fuzzified our ORC Ontology to accommodate the linguistic variables (e.g. BMI data types; Underweight, Normal, Overweight, and Obese) and overlapping concepts (e.g. Colorectal Cancer,

---

[16] http://www.cancer.gov/cancertopics/factsheet/detection/staging
[17] http://mowl-power.cs.man.ac.uk:8080/metrics
[18] http://www.w3.org/2001/sw/wiki/Ontology_Metrics
[19] http://mowl-power.cs.man.ac.uk:8080/validator/

Colon Cancer, and Colon Adenocarcinoma). In our regular Ontology, we could not do that, as there are no sharping edges among concepts. In addition, linguistic variables have different ranges of values.

Our fuzzificationprocess was based on OWL 2 and Fuzzy annotation properties that could be done within the **Protégé 4.3**. **The FuzzyOWL2 plug-in**, made by Bobillo and Straccia [1] is publicly available on the Web. It enables defining Fuzzy elements to the typical Ontology (including Fuzzy datatypes, weighted sum concepts, weighted concepts, Fuzzy nominals, and others), specifying the Fuzzy Logic wanted to be used (either Zadeh [20] or Lukasiewicz [21] logics). We used Zadeh Fuzzy logic.The process output is a Fuzzy Ontology.

Eventually, the constructed Fuzzy Ontology uses **fuzzyDL**toreason with/query the processed Ontology. The plug-in is integrated with **fuzzyDL** reasoner [2], translates the annotated OWL 2 Ontology into fuzzyDL syntax, calls fuzzyDL, and makes it possible to submit queries. For the moment, such queries must be expressed using the particular syntax supported by fuzzyDL.

The Fuzzy Ontology can be printed on the screen or saved to a text file. The FuzzyOWL2 plugin installation included gurobi optimization tool[22] installation to use the query panel of the plug-in. All installation instructions were included in their plug-in documentation.

Table 7: The Object Properties of the Obesity-related Cancers Ontology.

Domain[a]	**Range**[b]	**Property**
Patient	Disease	hasDisease
Patient	Country	IsLocatedIn
Obesity	Obesity_Risk_Factors	hasCauses
Obesity	Obesity_Symptoms	hasSymptoms
Obesity	Obesity_Diagnosis	diagnosedBy
Obesity	Obesity_Treatment	treatedBy
Obesity	Diabetes	ResultsIn
Obesity	Diabetes	ResultsIn
Cancer	Cancer_Risk_Factors	hasCauses
Cancer	Cancer_Symptoms	hasSymptoms
Cancer	Cancer_Diagnosis	diagnosedBy
Cancer	Cancer_Treatment	treatedBy

[a] Domain is a built-in property that links a property to a class description.

[b] The range is a built-in property that links a property to either a class description or a data range.

[20] http://en.wikipedia.org/wiki/Fuzzy_logic

[21] http://en.wikipedia.org/wiki/%C5%81ukasiewicz_logic

[22] http://www.gurobi.com

Figure 5: The Obesity Class Visualization in DO [2].

Table 8: The Instances (Individuals) of ORC Ontology classes.

Class	Instances
Obesity_Risk_Factors	Gender, Age,Lipids,Genes_and_family_history, Diabetes, Lifestyle, Hormone_problems, Certain_medicines, Lack_of_sleep, Emotional_factors,Smoking_stopping, Pregnancy, Lack_of_energy_balance_over_time
Obesity_Symptoms	Clothes_feeling_tight, Having_extra_fat_around_the_waist, Greater_scale_measure, A_Higher_than_normal_BMI, Having_higher_waist_circumference
Obesity_Diagnosis	Gender, BMI, Blood_glucose_level_tests (Fasting_plasma_glucose, HbA1c, Oral_gulcose_tolerance), Lipid_profile_tests (Triglyceride, Total_cholestrol), Genetic_factors , Waist_measurement, Retrospective_studies_in_community
Obesity_Treatment	Weight_loss, Lifestyle_change (Cutting_back_on_calories, Healthy_eating_plan, Physical_Activity), Medicines, Surgery
Cancer_Risk_Factors	Age, Gender, Morbid_obesity, Inherited_gene_faults, Lifestyle, Smoking, DNA_damage, Viruses, Problems_with_the_immune_system
Cancer_Symptoms	Feeling_ill_without_obvious_cause, Pernicious Anemia Tumour_mutations
Cancer_Diagnosis	Imaging (X-ray, CT_scan, MRI_scan, PET_scan, Ultrasound), Tumor_markers_tests (CA-125, AFP,…), Biopsy, Endoscopy,Physical_examination
Cancer_Treatment	Radiotherapy, Surgery, Chemotherapy, Hormone_therapy, Immunotherapy, Gene_therapy

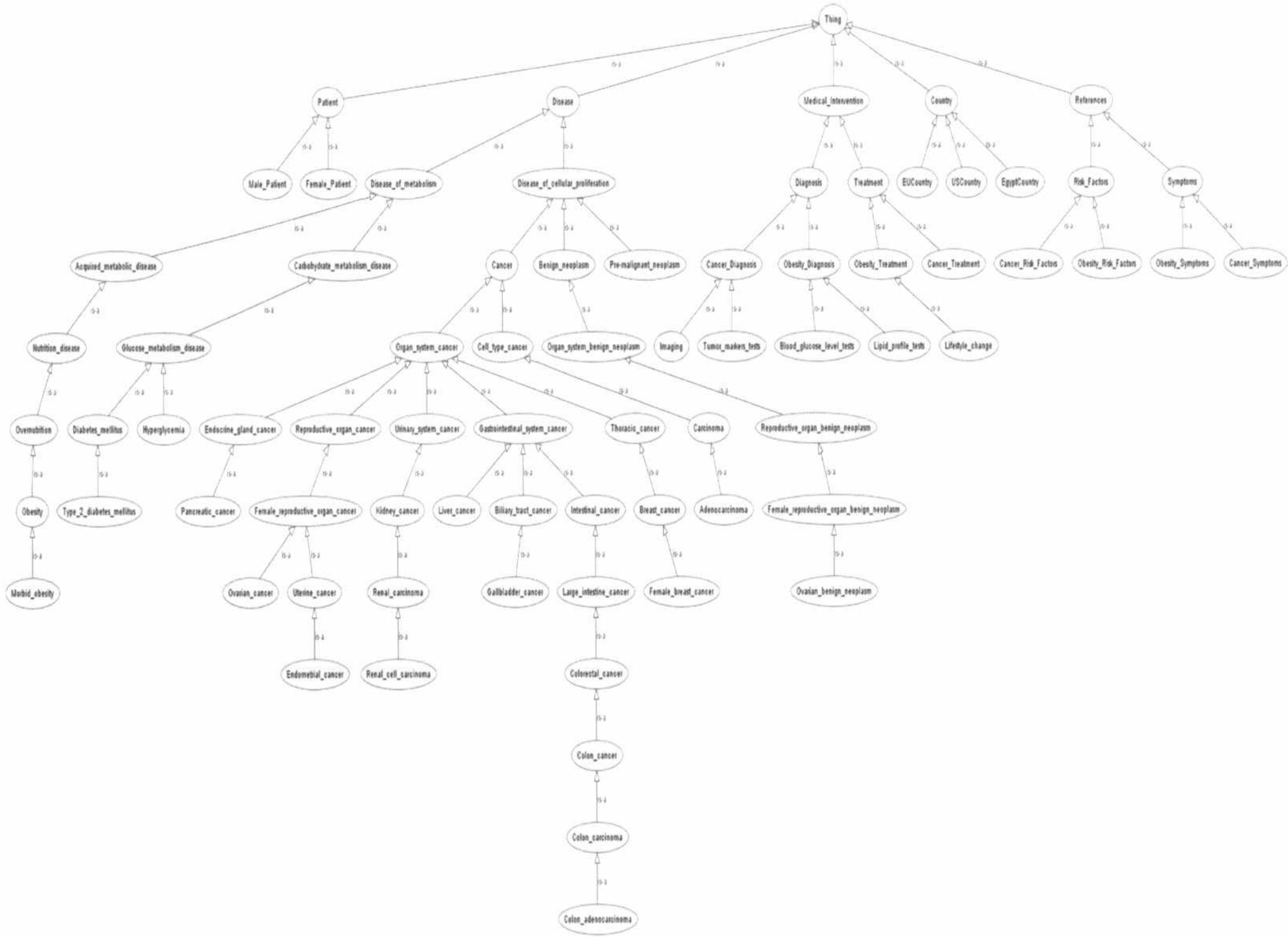


Figure 6: The ORC Full Ontology Classes.

Figure 7: Diagnosis Class and its members represented in ORC Ontology.

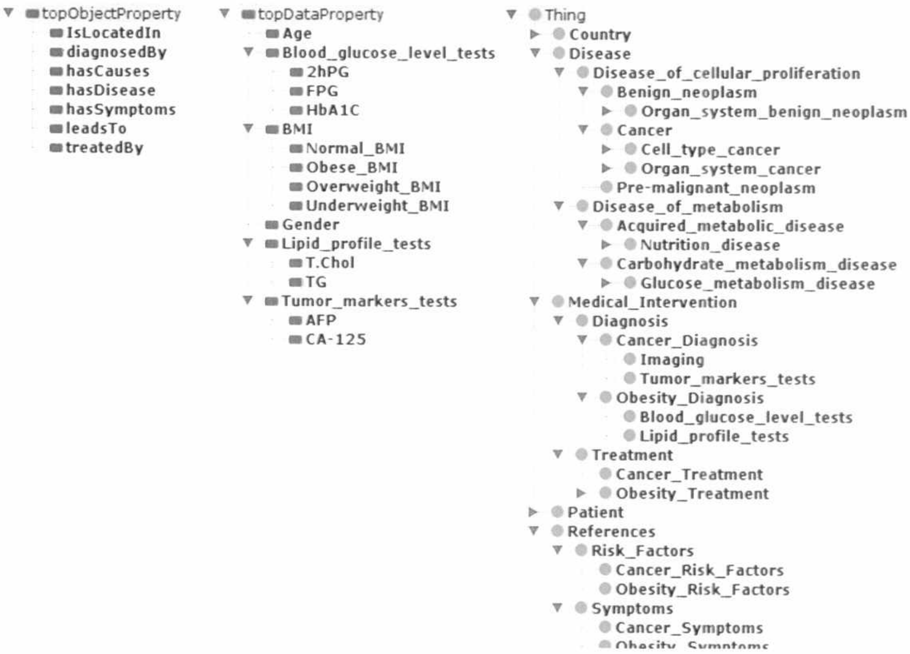


Figure 8: Selection of classes, data and object properties in the ontology (partial).

MANCHESTER 1824

The University of Manchester

Ontology Metrics

http://www.semanticweb.org/mohhed/ontologies/2014/6/1/obesity-related-cancer

Ontology

DL expressivity	ALC(D)
Class count	61
Object property count	4
Data property count	3
Individual count	62
Logical axiom count	199
Ontology annotation axioms count	0
Entity annotation axioms count	128

Class axioms

Logical axiom count	199
SubClass axioms count	57
Equivalent classes axioms count	0
Disjoint classes axioms count	3
Disjoint union axioms count	0

Object property axioms

Sub object property axioms count	0
Equivalent object properties axioms count	0
Disjoint object properties axioms count	0
Inverse object properties axioms count	0
Object property domain axioms count	4
Object property range axioms count	4
Functional object property axioms count	0
Inverse functional object property axioms count	0
Symmetric object property axioms count	0
Anti-symmetric object property axioms count	0
Reflecive object property axioms count	0
Irrefexive object property axioms count	0
Transitive object property axioms count	0

Data property axioms

Sub data property axioms count	0
Equivalent data properties axioms count	0
Disjoint data properties axioms count	0
Object property domain axioms count	4
Object property range axioms count	4
Functional data property axioms count	0

Individual axioms

Class assertion axioms count	125

Figure 9: The typical ORC OntologyMetrics using Manchester University Validation Tool [20].

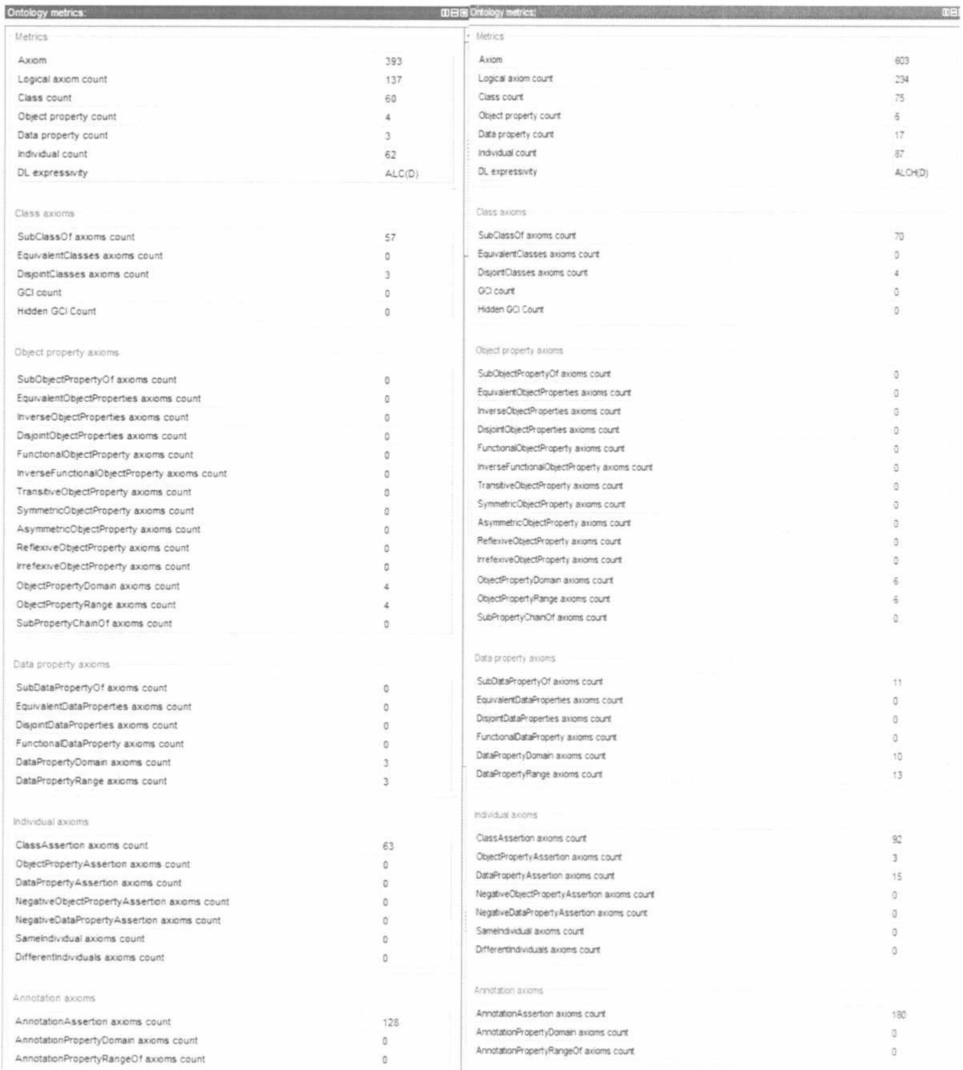
4.2.1. Definition of Fuzzy Sets:

As we showed in Table 1, BMI had four linguistic variables; they are Underweight, Normal Weight, Overweight, and Obese that can be fuzzified denoting the degree of a patient being underweight, normal, overweight or obese and then represented by the fuzzy plug-in using protégé.

We can define the four fuzzy sets of BMI linguistic variables like:

1. Underweight := FUZZY SET (18.5,1), (19.5,0);
2. Normal := FUZZY SET (18.5,0), (19.5,1), (24,1), (25,0);
3. Overweight := FUZZY SET (24,0), (25,1), (29,1), (30,0);
4. Obese := FUZZY SET (29,0), (30,1);

The first three fuzzy sets were defined by Fehre et al. [23] and upon them, we described the fourth one, see Figure 12.



Ontology metrics:

Metrics	
Axiom	393
Logical axiom count	137
Class count	60
Object property count	4
Data property count	3
Individual count	62
DL expressivity	ALC(D)
Class axioms	
SubClassOf axioms count	57
EquivalentClasses axioms count	0
DisjointClasses axioms count	3
GCI count	0
Hidden GCI Count	0
Object property axioms	
SubObjectPropertyOf axioms count	0
EquivalentObjectProperties axioms count	0
InverseObjectProperties axioms count	0
DisjointObjectProperties axioms count	0
FunctionalObjectProperty axioms count	0
InverseFunctionalObjectProperty axioms count	0
TransitiveObjectProperty axioms count	0
SymmetricObjectProperty axioms count	0
AsymmetricObjectProperty axioms count	0
ReflexiveObjectProperty axioms count	0
IrrefexiveObjectProperty axioms count	0
ObjectPropertyDomain axioms count	4
ObjectPropertyRange axioms count	4
SubPropertyChainOf axioms count	0
Data property axioms	
SubDataPropertyOf axioms count	0
EquivalentDataProperties axioms count	0
DisjointDataProperties axioms count	0
FunctionalDataProperty axioms count	0
DataPropertyDomain axioms count	3
DataPropertyRange axioms count	3
Individual axioms	
ClassAssertion axioms count	63
ObjectPropertyAssertion axioms count	0
DataPropertyAssertion axioms count	0
NegativeObjectPropertyAssertion axioms count	0
NegativeDataPropertyAssertion axioms count	0
SameIndividual axioms count	0
DifferentIndividuals axioms count	0
Annotation axioms	
AnnotationAssertion axioms count	128
AnnotationPropertyDomain axioms count	0
AnnotationPropertyRangeOf axioms count	0

Ontology metrics:

Metrics	
Axiom	603
Logical axiom count	234
Class count	75
Object property count	6
Data property count	17
Individual count	87
DL expressivity	ALCH(D)
Class axioms	
SubClassOf axioms count	70
EquivalentClasses axioms count	0
DisjointClasses axioms count	4
GCI count	0
Hidden GCI Count	0
Object property axioms	
SubObjectPropertyOf axioms count	0
EquivalentObjectProperties axioms count	0
InverseObjectProperties axioms count	0
DisjointObjectProperties axioms count	0
FunctionalObjectProperty axioms count	0
InverseFunctionalObjectProperty axioms count	0
TransitiveObjectProperty axioms count	0
SymmetricObjectProperty axioms count	0
AsymmetricObjectProperty axioms count	0
ReflexiveObjectProperty axioms count	0
IrrefexiveObjectProperty axioms count	0
ObjectPropertyDomain axioms count	6
ObjectPropertyRange axioms count	6
SubPropertyChainOf axioms count	0
Data property axioms	
SubDataPropertyOf axioms count	11
EquivalentDataProperties axioms count	0
DisjointDataProperties axioms count	0
FunctionalDataProperty axioms count	0
DataPropertyDomain axioms count	10
DataPropertyRange axioms count	13
Individual axioms	
ClassAssertion axioms count	92
ObjectPropertyAssertion axioms count	3
DataPropertyAssertion axioms count	15
NegativeObjectPropertyAssertion axioms count	0
NegativeDataPropertyAssertion axioms count	0
SameIndividual axioms count	0
DifferentIndividuals axioms count	0
Annotation axioms	
AnnotationAssertion axioms count	180
AnnotationPropertyDomain axioms count	0
AnnotationPropertyRangeOf axioms count	0

Figure 10: The typical Ontology metrics [20] vs. our modified ORC Ontology metrics in Protégé.

Figure 11: The validation reportfor our ORC Ontology format.

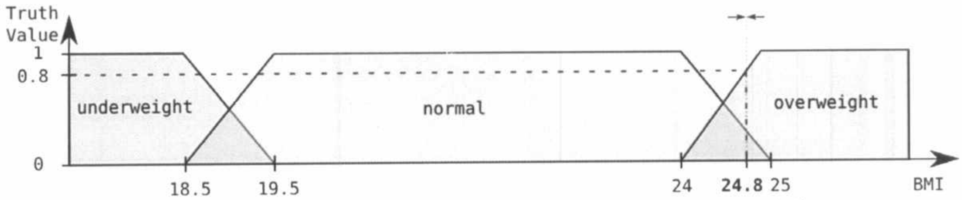


Figure 12: The fuzzy sets for underweight, normal, and overweight BMI [23].

4.2.2. Fuzzy Data Types Representation:

To represent the Fuzzy atomic data types, we need to specify the parameters k1, k2, a, b, c, d. The first four parameters are common to all of them, c and d appear in the trapezoidal function. The parameters k1 and k2 are the minimum and maximum inclusive values, respectively. These parameters are optional and, if omitted, then the minimum and maximum of the attributes (a, b, c, d) are assumed, respectively. We specified k1, k2 with 0, 300 as the heaviest human till now recorded more than 204 BMI (Kg/m^2)[23]. We represented our fuzzy datatypes using values of (k1, k2, a, b, c, d) as the following using two left and right triangle functions, and two trapezoidal functions. Figure 13 shows the underweight Fuzzy data type representation in the building environment as an example:

- Underweight_datatype = Left(0, 300, 18.5,19.5)
- Normalweight_datatype = Trapezoidal(0, 300, 18.5, 19.5, 24, 25)
- Overweight_datatype = Trapezoidal (0, 300, 24, 25, 29, 30)
- Obese_datatype = Right(0, 300, 29, 30)

The medical experts told that overlapped concepts of Colorectal Cancer were used approximately as 60% for Colorectal Cancer, 30% for Colon Cancer, and 10% for Colon Adenocarcinoma. Figure 14 shows a Sample of the used Fuzzy label annotation properties for representing both fuzzy data types and Overlapped Concepts of Colorectal Cancer with different degrees of usage (0.6, 0.3, and 0.1).

[23]http://en.wikipedia.org/wiki/List_of_the_heaviest_people

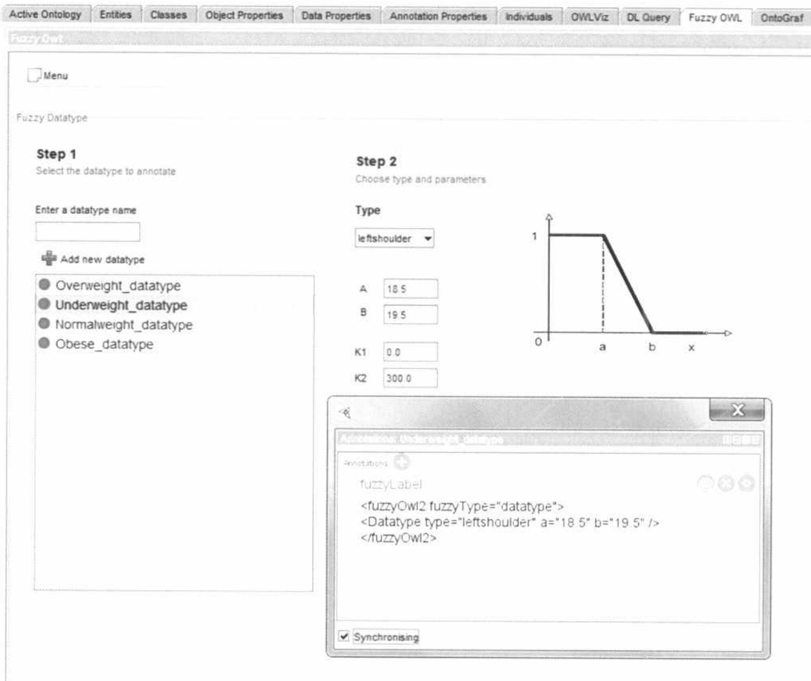


Figure 13: The underweight Fuzzy Data Type Representation (as an example).

4.2.3. Phase 3: Query the FOORC Ontology

Using the installed plugin and Gurobi software, we can send queries in specified syntax and predefined tags to our constructed FOORC and get Fuzzy answers. To check our Ontology response and consistency, we made some queries like:

(max-subs? Colorectal_cancerLarge_intestine_cancer), (min-subs?Colorectal_cancerLarge_intestine_cancer) to get the maximum and minimum values of concept implication Colorectal_cancer ->Large_intestine_cancer, as shown in Figure 15.FOORC responded with the expected answers for the given queries.

5. Results

The FOORC validation was made in two stages. First, validating the phase 1 of the regular Ontology using the regular validation tools as shown in Section 4.1.4. Then, validating the second and third phases of getting answers from the Fuzzy Dl reasoner that reflect user's queriesusing the fuzzy annotations approach.

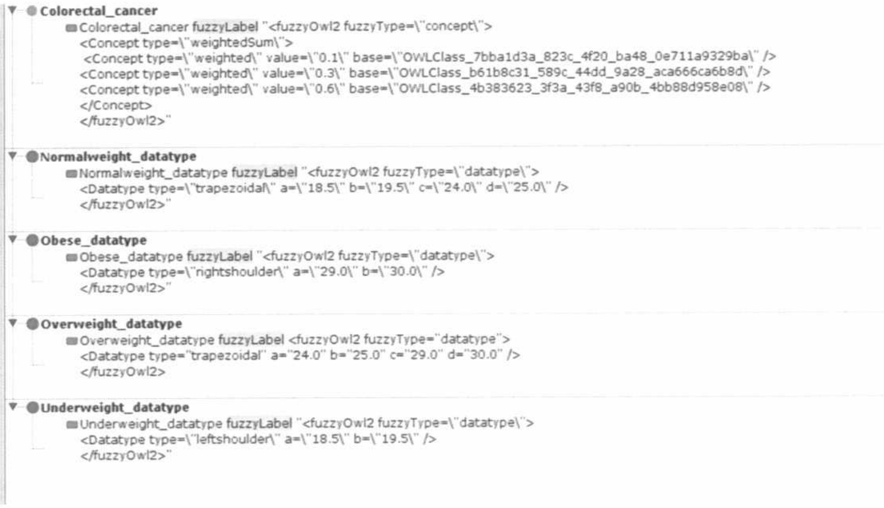


Figure 14: The used Fuzzy Label Annotation Properties for both fuzzy data types & Overlapped Concepts of Colorectal Cancer.

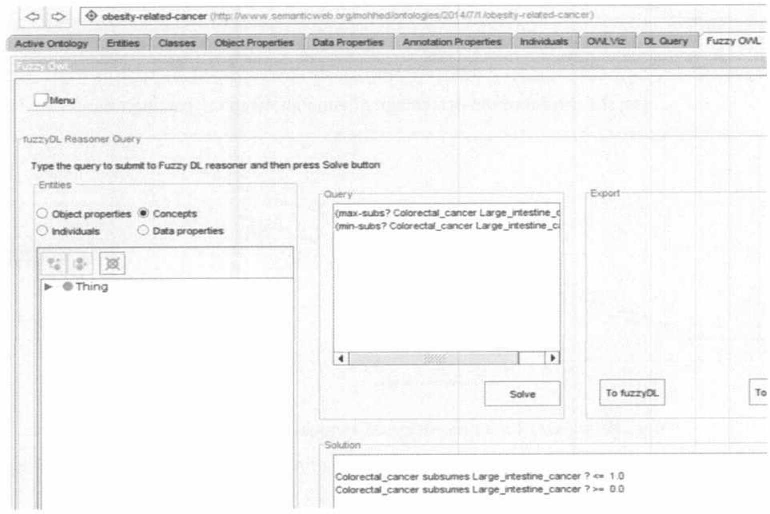
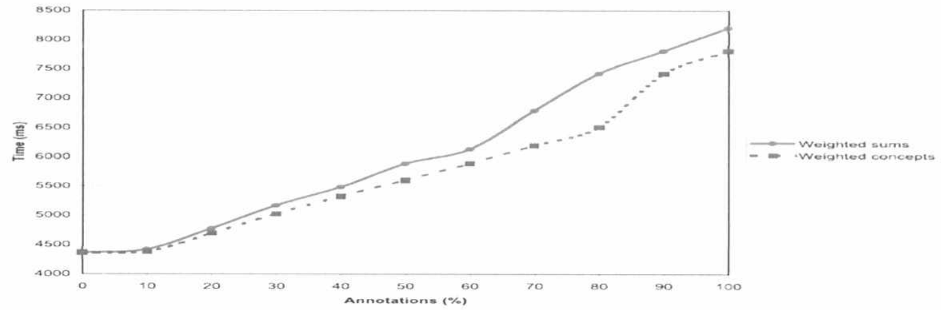


Figure 15: Results: get the minimum and maximum values of concept implication Colorectal_cancer ->Large_intestine_cancer

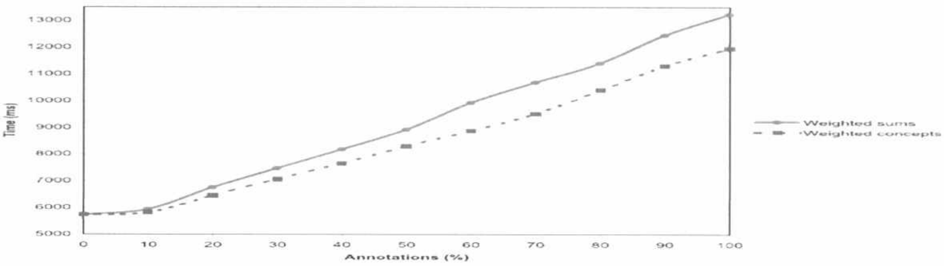
The FOORC replied with the expected answers for the given queries, if the Fuzzy Ontologyhad something wrong with Fuzzy representation, the reasoner would provide an only "ERROR" response with no result. In addition, Bobillo and Straccia [1] made an experimental evaluation ofusingFuzzy annotation properties. The final evaluationdecision was there is no additional overhead for the annotations, and good performance was acquired. Figure 16 displays their experimental evaluation results using Galen Ontology. Table c in Figure 16 shows the influence of the percentage of annotations

(%) in both PT (the parsing time) and TT (the translation time) into fuzzyDL syntax. The parsing time and the translation time are shown for both WSs (Weighted Sums) and WCs (Weighted Concepts).

The numbers of annotated elements influence in the PT is shown in Figure 16a. It is noticeable there is a semi-linear growing of the PT concerning the number of annotated elements. A Fuzzy Ontology with a forty percent of annotated elements would take one more second to be parsed than the original Galen Ontology. In addition, it is obvious that there are no considerable differences between WCs and WSs, in general, which means the types of the Fuzzy concepts are not significative. The numbers of annotated elements influence in the TT is shown in Figure 16b. Again, there is a semi-linear growing of the running time concerning the number of annotated elements, and there are no significant differences because of the type of the fuzzy concepts [1].



(a) The impact of the percentage of annotations in the parsing time.



(b) The impact of the percentage of annotations in the translation time.

(c) Influence of the percentage of annotations in the parsing time &the translation time into FuzzyDL syntax.

%	Concepts	GCIs	RIAs	PT WCs	PT WSs	TT WCs	TT WSs
0	0	0	0	4364.1	4363.9	5731.7	5726.1
10	2385	2604	88	4420.3	4382.5	5932.8	5812.6
20	4590	5151	177	4773.6	4692	6746.8	6443.9
30	6990	7675	276	5166.8	5025.2	7465.5	7059.4
40	9312	10152	383	5481.4	5320.3	8173.5	7648.1
50	11588	12760	462	5884.5	5603.4	8925.2	8295.4
60	13888	15260	569	6131.6	5889	9928.1	8875
70	16216	17764	672	6785.7	6193.9	10690.5	9521.6
80	18475	20363	785	7418.6	6509.4	11403.1	10402.8
90	20805	22906	875	7809.2	7418.8	12451.7	11303.3
100	23141	25563	958	8201.6	7813.8	13228.3	11962.6

Figure 16: The results of the experimental evaluation [1].

6. Conclusion

The Obesity Related Cancer(s) is a rich and significant medical domain. From our experiment, the proposed FOORC was better to represent this domain than the typical one for several reasons. One of them was the ability to represent overlapping concepts and linguistic variables that had not sharp edges to be represented in regular Ontologies, and this was done via the Fuzzy annotation properties (like fuzzy datatypes, weighted sum concepts, …etc.). It led us to accommodate more concepts and make a wider range of vocabularies. Second, enabling the user(s) to send queries to the fuzzyDLreasoner that in turn replies with Fuzzy Ontologies. Third, it best fits for rich domains having Fuzzy knowledge that need to be represented within Ontologies. Finally, it leads to good performance, in general. We introduced a simple three phases methodology to build the FOORC that is expected to be good practice for ontologists and knowledge engineers in medical field aiding them to solve the overlapping concepts, linguistic variables, and reasoning problems. Both physicians and intelligent systems can exploit obesity-related cancer Fuzzy Ontology in knowledge sharing, reusability, and reasoning. In future, we may extend this work to include all the cancer types with their elements in the fuzzificationprocess or may work on a particular cancer type with the study of patients group's real data. The fuzzy plug-in may need more development to facilitate the users to submit more queries than the predefined ones.

7. Acknowledgment

The authors want to express gratitude toward Dr. Ashraf Khater,Prof. of Surgical Oncology for his efforts, Dr. Moahamed Awad Ibrahem, Assoc. Prof. of Medical Oncology (Mansoura University Oncology Center), Dr. Noureddin Sadawi (Dept. of Computer Science, Brunel University, London, United Kingdom), and Mr. Wolfram Bartussek (Lecturer in University of Applied Sciences, Darmstadt, Germany) for their support during this work.

References

1. Fernando Bobillo, Umberto Straccia. Fuzzy Ontology representation using OWL 2.International Journal of Approximate Reasoning archive, Volume 52 Issue 7, pp. 1073-1094, October 2011.
2. Disease Ontology - Institute for Genome Sciences @ University of Maryland; http://disease-Ontology.org/ (Last accessed on 01 Sep, 2014)
3. Semantic Web Wiki. http://semanticweb.org/wiki/Main_Page. (Last Accessed on 20 Mar 2015)
4. Semantic Web. http://en.wikipedia.org/wiki/Semantic_Web. (Last Accessed on 20 Mar 2015)
5. Shyi-Ming Chen, Yun-Hou Huang, Rung-Ching Chen, Szu-Wei Yang and Tian-Wei Sheu. Using Fuzzy Reasoning Techniques and the Domain Ontology for Anti-Diabetic Drugs Recommendation. Springer-VerlagBerlin Heidelberg 2012, J.-S. Pan, S.-M. Chen, N.T. Nguyen (Eds.): ACIIDS 2012, Part I, LNAI 7196, pp. 125–135, 2012.
6. David Parry. A fuzzy ontology for medical document retrieval, ACSW Frontiers '04 Proceedings of the second workshop on Australasian information security, Data Mining and Web Intelligence, and Software Internationalisation - Volume 32; pp. 121-126, published on ACM, 2004.
7. World Wide Web Consortium (W3C). OWL 2 Web Ontology Language Document Overview (Second Edition). http://www.w3.org/TR/owl2-overview. (Last Accessed on 23Dec 2013).
8. AbolfazlDoostparastTorshizi, AbolfazelDoostparastTorshizi, GhazalehDoostparastTorshizi, and KamyarEghbali. A Hybrid Fuzzy-Ontology Based Intelligent System to Determine Level of Severity and Treatment Recommendation for Benign Prostatic Hyperplasia. Computer Methods and Programs in Biomedicine, Volume 113, Issue 1, pp 301-313, 2014.

9. Jun Zhai, LixinShen, Zhou Zhou and Yan Liang. Fuzzy Ontology Model for Knowledge Management. Intelligent Information Processing IV, IFIP – The International Federation for Information Processing Volume 288, pp. 58-67, 2008.
10. OWL Web Ontology Language. http://www.w3.org/TR/owl-features/. (Last accessed on 01 Mar 2014)
11. World Wide Web Consortium (W3C). OWL 2 new features. http://www.w3.org/TR/2012/REC-owl2-new-features-20121211/. (Last Accessed on 01 Dec2013).
12. Obesity and Cancer Risk - National Cancer Institute - Fact Sheet. http://www.cancer.gov/cancertopics/factsheet/Risk/obesity ((Last Accessed on 30Dec, 2015).
13. Eugenia E. Calle and Rudolf Kaaks. Overweight, Obesity and Cancer: Epidemiological Evidence and Proposed Mechanisms. Nature Reviews Cancer 4, pp. 579-591, August 2004. doi:10.1038/nrc1408.
14. Tumor Markers - National Cancer Institute - Fact Sheet. http://www.cancer.gov/cancertopics/factsheet/detection/tumor-markers. (Last Accessed on21Dec, 2015)
15. OncoLink - The Web's First Cancer Resource - Patient Guide to Tumor Markers. http://www.oncolink.org/treatment/article.cfm?id=296 (Last Accessed on 11 Dec, 2015)
16. M. Alfonse, M. M. Aref, Abdel-Badeeh M. Salem, "Ontology-Based Knowledge Representation for Liver Cancer", Electronic Proceedings of the International eHealth, Telemedicine and Health ICT Forum for Educational, Networking and Business, pp. 840-844, 2012.
17. I. F. Moawad, G. AL Marzoqi, A-B. M. Salem, "Building OBR-Based Owl Ontology for Viral Hepatitis", Electronic Proceedings of the International eHealth, Telemedicine and Health ICT Forum for Educational, Networking and Business, pp. 821-825, 2012.
18. Abdel-Badeeh M. Salem and Hisham S. Katoua. Web-Based Ontology of Knowledge Engineering. Journal of Communication and Computer 9, pp. 1336-1342, 2012.
19. Fatimatufaridah Jusoh, Roliana Ibrahim, Mohd Shahizan Othman and Norshafarina Omar. Development of Breast Cancer Ontology Based on Hybrid Approach. International Journal of Innovative Computing 01(1), pp.15-22, 2013.
20. Mohammad A. Elhefny, Mohammed Elmogy and Ahmed A. Elfetouh. Building OWL Ontology for Obesity Related Cancer, 9th IEEE International Conference on Computer Engineering & Systems (ICCES), Cairo, Egypt. ISBN: 978-1-4799-6593-9, DOI: 10.1109/ICCES.2014.7030953, pp. 177–183, 2014.
21. Manmohan Singhal, V.R. Manchireddy Vishnu, S.V. Rama Raju and Yozana Upadhyay. Interrelationship between Obesity and Cancer (A Review). Academic Journal of Cancer Research 6 (1): pp. 13-20, 2013. ISSN 1995-8943, DOI: 10.5829/idosi.ajcr.2013.6.1.739.
22. Fernando Bobillo and Umberto Straccia. fuzzyDL: An Expressive Fuzzy Description Logic Reasoner, IEEE International Conference on Fuzzy Systems,Vols 1-5; pp. 923-930, 2008.
23. Fehre K, Mandl H and Adlassnig K-P, A Fuzzy Arden Syntax Compiler.Schreier G, Hayn D, Ammenwerth E, editors Tagungsband der eHealth2010: Health Informatics meets eHealth6-7 Mai 2010; Wien OCG; 2010.