TMRF e-Book
Advances in Semantic Computing
(Eds. Joshi, Boley & Akerkar), Vol. 2, pp 32 – 51, 2010

Chapter 3

OWL Ontology for Solar UV Exposure and Human Health

Cambillau Mathieu, El-Shanta Eltaher, Purushotham Sarathy, and Simeu Gilles[*]

Abstract Various knowledge systems represent information in different ways, which may be easy to read and utilize for humans even though it is time consuming. Some systems represent knowledge in their code making it difficult to port and share, though they can perform reasoning tasks effectively. Conversely the Semantic Web is geared towards making this knowledge shareable for creating a global interlinked knowledge source with reasoning capabilities, to cater for humans and agents. The work described in this chapter is one step in the direction of the Semantic Web to show how knowledge can be consistently represented and utilized.

Protégé along with description logic helps to create consistent Web Ontology Language (OWL) ontologies and perform reasoning from the represented domain knowledge. This chapter illustrates the use of Protégé to create an OWL ontology for Solar UV rays and their effects on humans. This ontology represents information about UV rays effects on humans, factors influencing such effects, and protective measures against UV. Our work takes this a step further to show how reasoning tasks can be performed by inferring the knowledge represented to determine the risk level of a person from UV rays effects. We also discuss how this ontology and others similar can be useful in the Semantic Web endeavor.

INTRODUCTION

One of the numerous functions of the human mind is reasoning. However, one needs to be a domain expert to be able to perform good reasoning. Machines can do some reasoning tasks effectively if the relevant knowledge can be represented in a machine-readable way. When a domain of knowledge is large, especially with a lot of factors influencing any reasoning task, machine reasoning support can be helpful.

Ontologies are a good way to represent this knowledge. However it is difficult to represent a vast amount of knowledge in one monolithic ontology. Such ontology can be complex and difficult to maintain. “Constructing large ontologies typically requires collaboration among multiple individuals or groups with expertise in specific areas, with each participant contributing only a part of the ontology” *(Paolo Ciccarese, et al. 2009)*. A modular approach simplifies such a task and also promotes reuse. Modularity has long been introduced in software engineering, where by complete software systems are built from separate modules, to deliver the desired functionality. Modularity allows flexibility, better reuse potential, generalization and ease of maintenance. A number of completed and ongoing researches address the issues of studying and

[*] All Authors are affiliated with Faculty of Computer Science, University of New Brunswick, Canada
Cambillau Mathieu, email: mathieu.cambillau@mines-saint-etienne.org
El-Shanta Eltaher, email: X1H4R@unb.ca
Purushotham Sarathy, email: sarathy@unb.ca
Simeu Gilles, email: gilles.simeupoaka@mines-saint-etienne.org

developing frameworks, formalisms, and tools to facilitate the sharing and modularization process like (Bernardo et al 20) and (Heiner et al 20). Semantic Web Applications in Neuromedicine (SWAN) Ontology is a good example of modular ontology approach.

With the ozone layer depleting, UV rays intensity has been increasing in the earth's atmosphere. Understanding the effects of UV rays would mean having a good knowledge of the underlying factors. Our work is a first step towards representing effects of UV rays on human health for the Semantic Web representation and reasoning.

The goal is to build an OWL (Web Ontology Language) ontology for the domain of "Solar UV Exposure and Human Health". For this purpose we are using Protégé to create our ontology and Pellet as our reasoning tool. The scope of this undertaking is to use the World Health Organization (WHO) factsheets as our domain knowledge, though the final ontology has been expanded to represent additional information from other sources. This ontology includes factors influencing UV rays, effects on humans and protective measures that can be used. The reasoner can be employed to query the ontology to infer knowledge or conclusions.

In the following sections of this chapter, we first introduce the methodology that we followed to build this ontology. Then, we go through the taxonomy and the general structure of the ontology, highlighting important concepts and relationships. After explaining the taxonomy, we present the implementation and some of the limitations we faced with the Protégé ontology editor version 3.4.1, and why we switched to version 4 of Protégé.

Even with version 4, where we needed to apply certain restrictions, reasoning was not a smooth sailing. But eventually, we managed to build the ontology which can provide the required knowledge when queried. Once we have explained the implementation part, we demonstrate some example Description Logic (DL) queries on the ontology to give a flavor of the kind of knowledge that can be retrieved from the ontology. Finally, we make some concluding remarks and suggest possible ways this ontology can be used effectively.

BACKGROUND

The philosophical origin of the term ontology means "the science of being". Others were introduced by people working in specific domains such as Artificial Intelligence. Some definitions that are more linked to Computer Science characterize it as follows: "an ontology is a formal description of objects and their inter-relationships"*(W3C UK and Ireland Office),* and "a rigorous and exhaustive organization of some knowledge domain that is usually hierarchical and contains all the relevant entities and their relations" *(thefreedictionary.com)*. Tom Gruber's definition is probably still the most popular in this field. He defines an ontology as "an explicit specification of a conceptualization". Ontology is a tool largely used to share knowledge. It defines an agreement between its users to use the same vocabulary and concepts hierarchy *(Ontologies and Semantic Web).*

Ontologies are used to represent a particular domain of knowledge and facts in that domain. Reasoners are used with the ontologies to perform operations such as inferring taxonomic relationships and answering queries.

The OWL language is focused on DL which is widely used in the Semantic Web to build ontologies since the DL reasoning algorithms are well understand and have been used successfully in several projects. The Protégé OWL tool allows us to build ontologies with respect to the DL paradigm. It includes some key features such as logical expressions, class descriptions, reasoning and scalability *(Holger Knublauch 2004).*

Having the tool to build an ontology is not enough. We also need project support and a technical methodology. Methontology *(Mariano Fernandez, et al. 1997),* has been known to be a good project-supporting methodology for building ontologies. We explain it in more detail later in this chapter. "Ontology Development: 101" (*N. F. Noy 2001*) is a good starting point for the technical organization of ontologies. It gives methods about determining the domain and assessing the scope of an ontology project. It also explains how to define the classes and the class hierarchy from important domain terms and from the reuse of existing ontologies. This article also gives clues on when to use an individual instead of creating a class and provides several other conceptual suggestions concerning the taxonomy.

We based the domain development of our ontology on the WHO factsheets (Ultraviolet radiation: solar radiation and human health) which gives the important relations between UV exposure and its various good and bad effects such as diseases, vitamin D production, etc. It also lists some influencing factors and types of risks. For further information that we needed, we carried out more research on the UV rays domain. The article "Toxic effects of ultraviolet radiation" *(Yasuhiro Matsumura, et al. 2003)* explains the difference between the three types of UV: UVA, UVB and UVC. It also gives their respective wave lengths (Figure 1), and explains that the most dangerous one (UVC) is completely blocked by the Ozone layer in the Stratosphere. However, the two others reach the earth and affect human health.

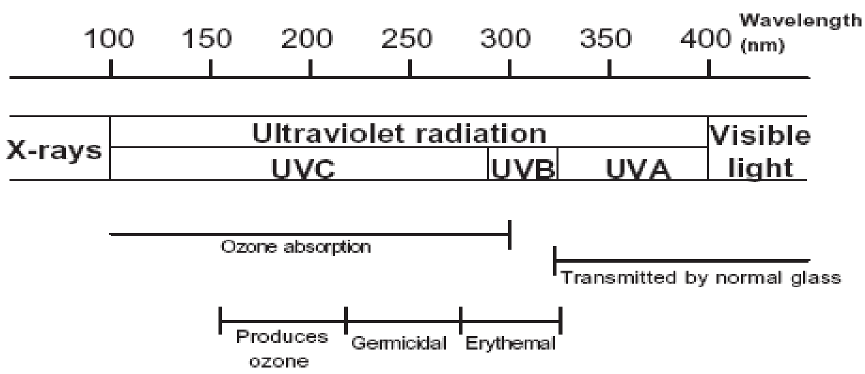

Figure 1. *(Yasuhiro Matsumura, et al. 2003).*

The WHO practical guide about Global Solar UV Index *(UNEP)* explains in detail what the Global Solar UV Index is, the different categories that exist, and the health effects of UV radiation exposure. It helped us to define the risk classes in a realistic way but also gave us more effects and influencing factors to include in our knowledge base.

METHODOLOGY

The methodology that we adopted in building the ontology is a modified version of the Agile Software Development proposed methodology called Methontology *(Mariano Fernandez, et al. 1997)* depicted in Figure 2.

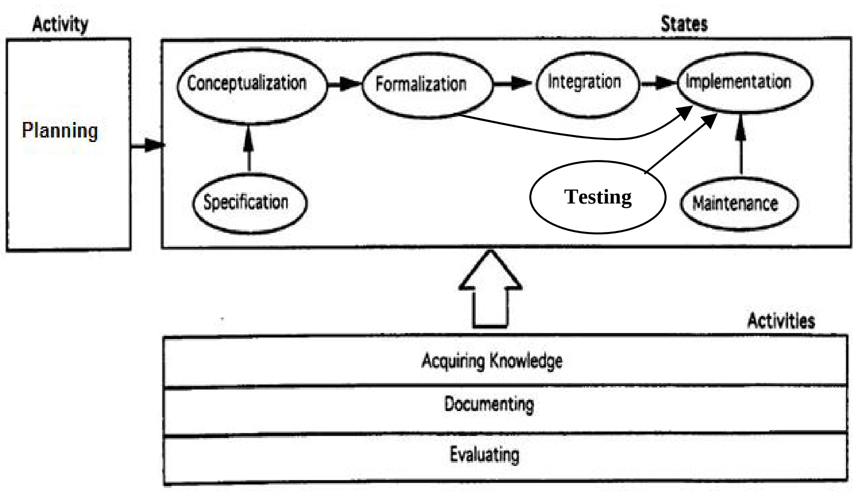

Figure 2. Methontology (*Mariano Fernandez, et al. 1997)*

We started the planning phase by discussing how to approach the project and then we divided the work load of the different steps in the methodology between members of our team. Having done that, we each started searching for more relevant information, knowledge, or descriptions that could be useful in building the ontology. The next step was to try to bring together what each of us gathered and define the concepts of the ontology based on the descriptions we had.

Once we were satisfied with our general understanding of the specification and the relationships among its components, we started formalizing the ontology and expressing the concepts we had in a formal class hierarchy (taxonomy). Then we added different relationships between those classes and any restrictions that we thought were applicable.

For the integration step, we carried out some research for any existing (formal) ontologies relevant to our topic, but we couldn't locate any. This is depicted by the arrow (Figure 2) that goes from the formalization phase straight to the implementation phase, bypassing the integration step.

After completing the final touches of the formalized ontology, we started implementing it using the Protégé ontology editor version 3.4.1. When the implementation part was done, we started the testing phase. Based on the testing results, we made some updates and amendments to the ontology and the relationships that tie its classes together until we were satisfied with the resulting ontology.

ONTOLOGY

With "*Solar UV Exposure and Human Health*" as the theme of the ontology, we created a top level layer of superclasses that represent the overall structure of the ontology. This layer of classes was aggregated with other layers of subclasses as required. Figure 3 shows the top-level layer of the ontology classes and the relationships amongst those classes.

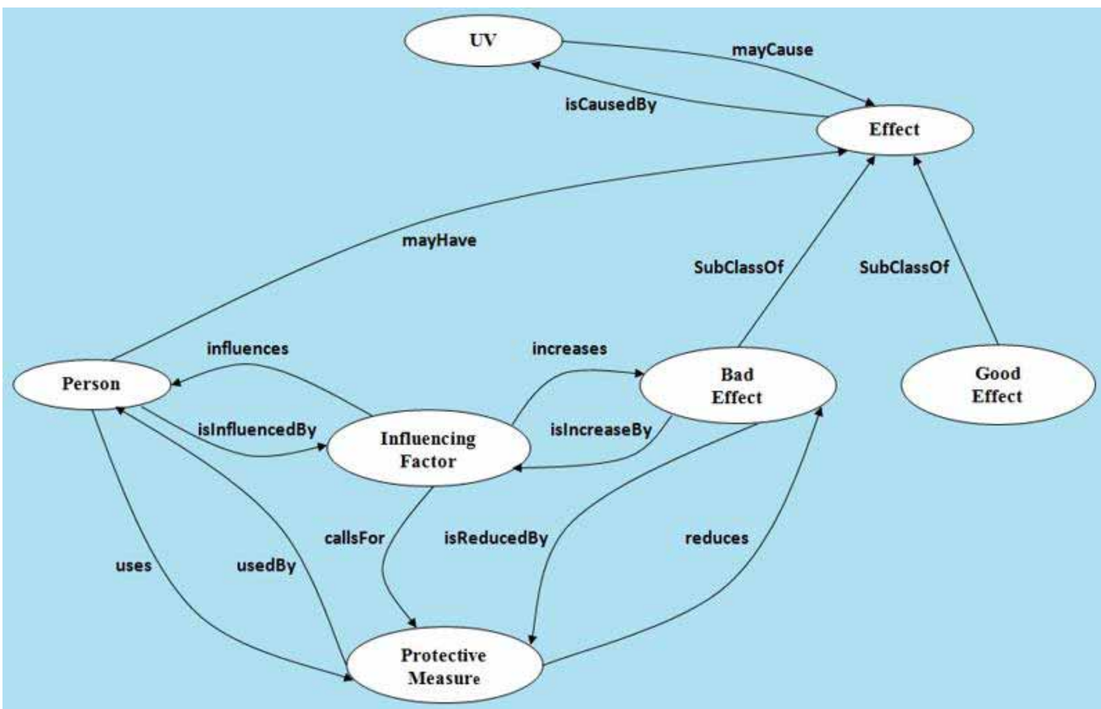

Figure 3. Top-level ontology classes and relationships.

As can be seen from Figure 3, we have the ultraviolet rays *UV* class as the top class. This class has a *mayCause* relationship with the *Effect* class. This relation is read as follows: UV rays may cause some effects on human health. Such UV rays effects can either be good or bad. This notion is represented by the two subclasses of the class *Effect*, namely, *GoodEffect* and *BadEffect*.

A human being may be the subject of UV rays effects both good and bad, which explains the *mayHave* relationship between the *Person* class and the *Effect* class. The degree of effects imposed by UV rays on human health varies based on some circumstances. We call such circumstances the influencing factors. Influencing factors may increase the bad effects of UV rays on human health. This is represented by the relationship *increases* between the *InfluencingFactor* class and the *BadEffect* class. Presence of such influencing factors requires a person to take some measures so as to reduce the risk posed by UV rays. We call such measures protective measures. The relationship between the *InfluencingFactor* class and the *ProtectiveMeasure* class is that the existence of some influencing factors *callsFor* protective measures.

Influencing factors may increase the degree of UV rays bad effects on a person. To reflect this in the ontology, we introduced the relationship that is interpreted as: influencing factor *influences* the degree of risk a person is at. *ProtectiveMeasure* class also has a direct relationship with the *BadEffect* class. This relationship can be described as: a protective measure *reduces* UV rays bad effects. It goes without saying that a person is responsible for taking those protective measures in order to reduce the UV rays bad effects on himself/herself. Such requirement was catered for in the ontology by the "person *uses* protective measure" relationship. It is worth noting that, whenever possible, we defined inverse relations for completeness of the ontology. This is clearly shown in Figure 3.

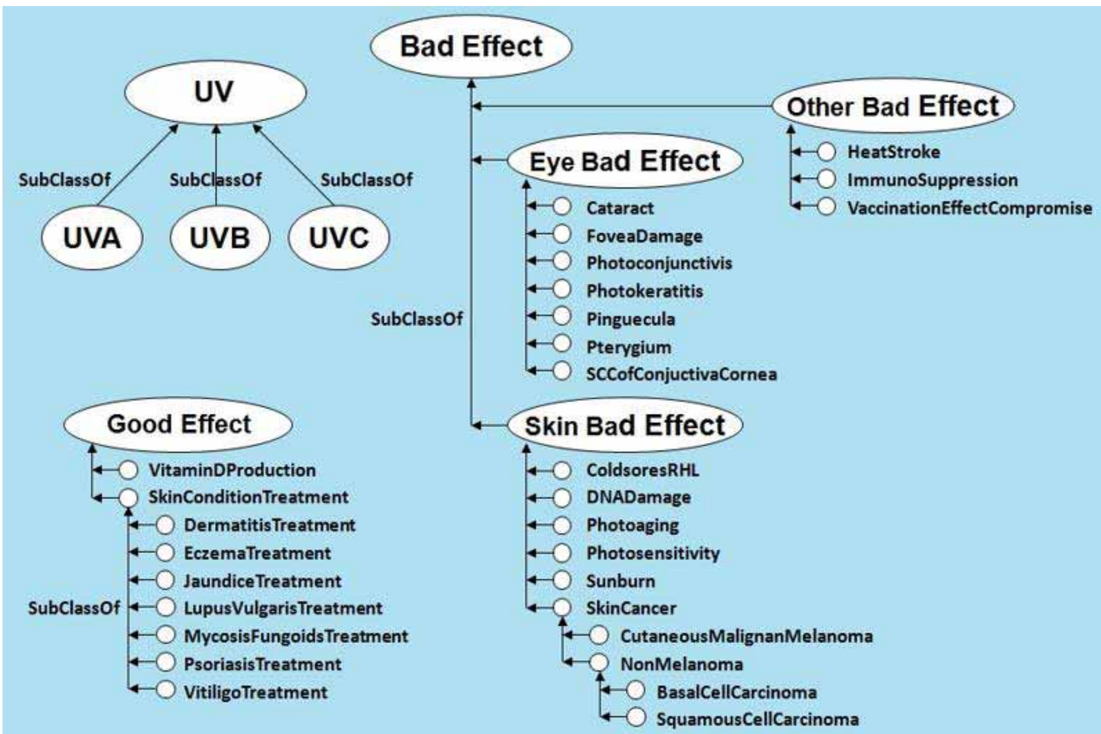

Figure 4. Sub-levels of UV and Effect classes.

As mentioned earlier, we have divided the taxonomy into several layers of classes according to the relationships among them. Figure 4 shows another layer of the subclasses that branch from some of the top layer classes explained above.

In the top left part of Figure 4, we can see that the UV class has three subclasses namely *UVA*, *UVB*, and *UVC*. These basically represent the three types of UV rays according to their wave length. To be more specific about the different types of UV rays bad effects, we assorted those types of bad effects into different subclasses of the *BadEffect* class. *EyeBadEffect* class represents the different diseases that UV rays can cause to the human eye. *SkinBadEffect* class represents the types of skin diseases – cancerous and non-cancerous – that UV rays cause to humans. Finally, *OtherBadEffect* class which represents other bad effects of UV rays a human being may be subjected to.

The major source of vitamin D for most humans is casual exposure of the skin to UVB portion of sunlight *(Tai C. Chen, et al. 2007)*. Good effects of UV rays come in the form of using such rays in the treatment of some skin conditions – under proper medical supervision of course - and the role of such rays in the production of Vitamin D in the human body.

So far, we have talked about the influencing factors that can increase human risk, but we haven't discussed what those factors are. Figure 5 illustrates in detail the different classes of the possible influencing factors and their subclasses. As can be inferred from Figure 5, influencing factors can be personal, environmental, or others that are related to the person's length of exposure to UV rays

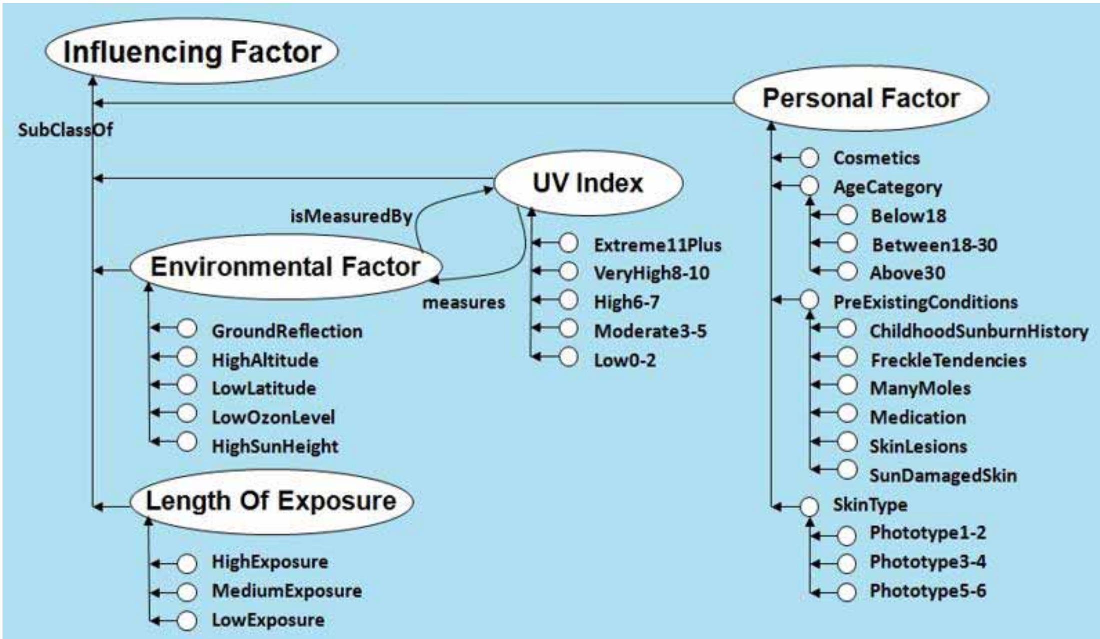

Figure 5. Sub-levels of Influencing Factor class

.

PersonalFactor class consists of a group of subclasses that represent personal related factors such as skin type, age, pre-existing conditions etc. *EnvironmentalFactor* class has a group of subclasses that are related to the environment surrounding the individual. Ground reflection, altitude, latitude and sun height are some of such subclasses. The longer a person is exposed to UV rays, the greater the risk of having bad effects. LengthOfExposure class serves that purpose in the ontology.

As we stated above, influencing factors can increase the risk that UV rays pose on human health. But we also stated that some protective measures can be taken by individuals to reduce such risk. The right hand-side part of Figure 6 shows the types of protective measures people can take to reduce the bad effects of UV rays. Subclasses of *ProtectiveMeasure* class include *ProtectiveClothing*, *WearingHat*, *UsingSunScreens*, *ReduceExposureTime, and WearingUVProtectingSunglasses*. Finally, the *Person* class is in turn divided into three subclasses namely, *LowRiskPerson*, *MediumRiskPerson*, and *HighRiskPerson*. These subclasses represent the grouping of people according to each individual's risk degree. Of course, different people will have different influencing factors based on their environments and their daily habits. Also, protective measures taken by each of them will be different – for a variety of reasons. The ontology can classify an individual as either being at low, medium, or high risk based on the given influencing factors and the protective measures for that individual.

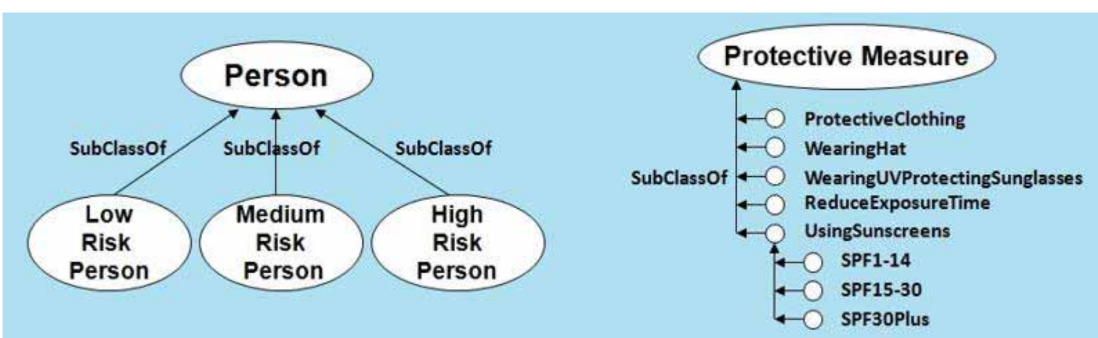

Figure 6. Sub-levels of Person and Protective Measure classes

INDIVIDUALS

Individuals play an important role in the ontology. A non-populated ontology without individuals fulfills only the function of a DL TBox (the set of the terminological components), not the DL ABox (the set of assertion components). In other words, the taxonomy is the part of the knowledge base which states how the elements can be used and describes the class hierarchy whereas the individuals would allow representing some facts that conform to the taxonomy in the knowledge base.

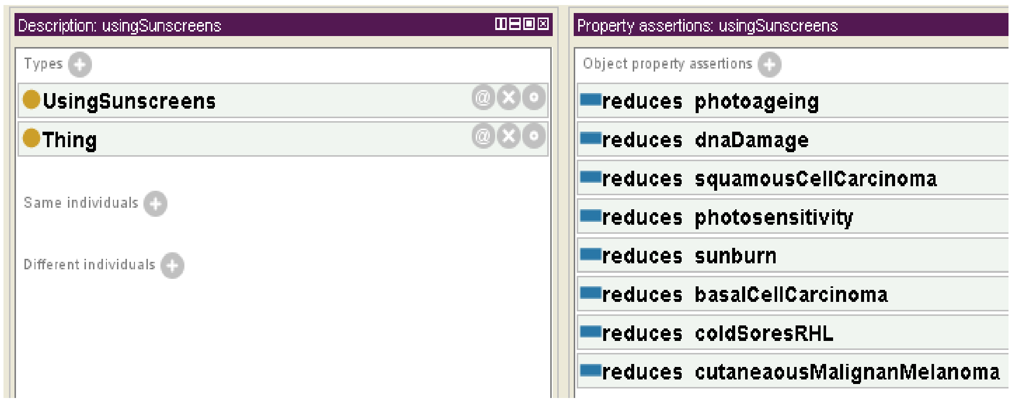

Figure 7. usingSunscreens individual example

Figure 7 shows an individual example in our ontology. For each individual, we can specify its classes and its joint or disjoint individuals (same individuals or different individuals). We can also add to each individual some object property (relations with other individuals) and data property assertions (primary type values, as integers for example, assigned as a property to the individual). In Figure 7, we can see that *UsingSunscreens* and *Thing* are classes for the individual *usingSunScreens*. This represents two facts of the knowledge base: *usingSunscreens* is a *UsingSunscreens* object but is also a *Thing* object. We also notice some object property assertions which represent other facts. For instance, the expression *reduces photoageing* implies a relation *reduces* from *usingSunscreens* to *photoageing*. It is used to represent the fact that using sunscreens reduces photo ageing. The same is true for *reduces sunburn*. It expresses that there exists a relation *reduces* from *usingSunscreens* to *sunburn*, which represents the fact that using sunscreens reduces sunburn.

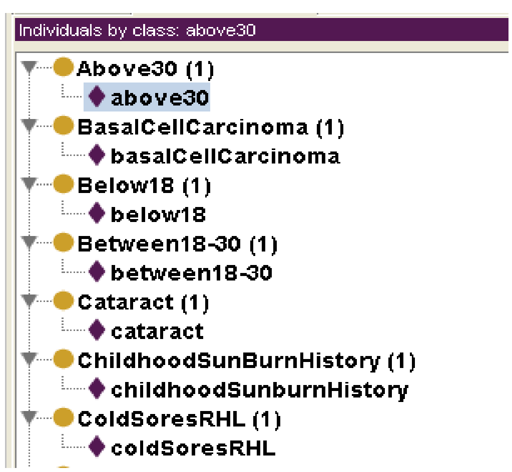

Figure 8. The singleton pattern

As shown in Figure 8, we chose to apply the singleton pattern to our individuals (except for the *Person* class for which we have many individuals used for testing our ontology classification). This means that each class has only one individual. This is more logical for objects such as diseases or protective measures for which it does not really make sense to have many individuals. It also avoids confusion when it comes to expressing facts.

REASONING WITH PROTÉGÉ 3

Restrictions

We use properties to introduce restrictions in OWL. Individuals belonging to a class can be refined by using restrictions e.g. Figure 9 below shows the restriction *UsingSunscreens only reduces SkinBadEffect*. This restriction implies that sunscreen will not reduce the UV effect for eye and other bad effects. Thus, such restrictions help to specify direct relations enabling us to restrict individuals belonging to a class. With all possible restrictions introduced we have a complete ontology.

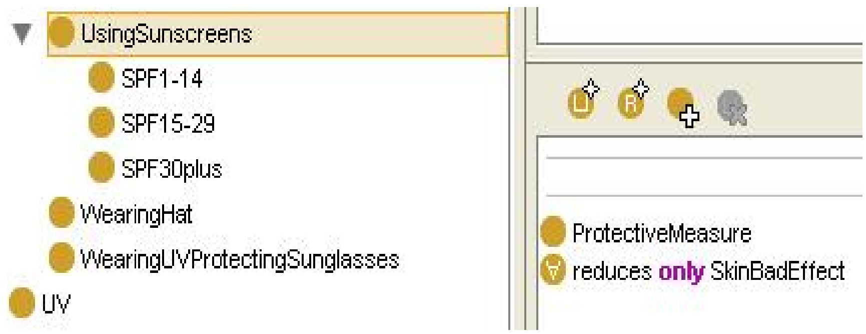

Figure 9. Sunscreen restriction

Reasoning

Upon building the ontology, the next task is to introduce reasoning in Protégé. The reasoner can check the class consistency, compute inferred super classes of classes and classify the ontology. Pellet reasoner available as a plug-in in Protégé 3was used for performing the reasoning task. From our OWL ontology we are able to perform basic queries to provide information about the knowledge represented which we will discuss in detail later in the querying section. However, to get more information like the risk level of a person from UV rays, we would have to introduce more logic expressions that can help the reasoner to categorize a person based on the influencing factors.

As discussed earlier in the taxonomy, there are several factors that influence the Solar UV rays. The degree of influence of these factors is not clear, however, each of these factors plays a role and so we cannot ignore any of them. Neither the WHO factsheets nor other sources available provide these levels of influence by these various factors. Knowing or estimating the level of influence is the key element in categorizing the risk level of individuals. Without categorizing the risk for an individual it would be like saying that all these factors influence UV rays effects on humans and all humans will have equal effects from UV rays, which is not true.

The approach used by our team was to use weights to classify the levels of influence of the various factors identified. The weighted approach is a simple way to deal with this problem and the risk categorization can be easily achieved by using the total weight of all factors a person is influenced by. Using a common sense approach, all of the factors were given weights with a scale of 1 to 5. e.g. The *UVIndex* factor *VeryHigh8-10* gets a weight of 5 compared to the *Low0-2 UVIndex factor* which is given a weight of 2. This weighted approach helps to classify the individuals to their appropriate levels of risk as high, medium or low. This classification is based on the total weights of “above 11”, “6 to 11” and “1 to 5” of the influencing factors that correspond to an individual. Table 1 shows the weight table for all Influencing factors represented in this ontology.

Table 1. Weight Table for Influencing Factors

Influencing Factors			**Weights**
PersonalFactor			
	PreExistingConditions		
		Medication	2
		ChildhoodSunBurnHistory	4
		ManyMoles	4
		SunDamagedSkin	4
		FreckleTendencies	4
		SkinLesions	4
	AgeCategory		
		Above30	1
		Between18-30	3
		Below18	5
	Cosmetics		2
	SkinType		
		Phototype5-6	1
		Phototype3-4	3
		Phototype1-2	5
UVIndex			
	Low0-2		2
	Moderate3-5		3
	High6-7		4
	VeryHigh8-10		5
	Extreme11plus		5
LengthOfExposure			
	LowExposure		1
	MediumExposure		3
	HighExposure		5

In Protégé 3 the total weight cannot be calculated directly by using any computations. So logic expressions that would categorize based on weights were necessary. Each expression would have several combinations of factors based on weights like *isInfluencedBy some Phototype1-2 and isInfluencedBy some Below18 and isInfluencedBy some VeryHigh8-10.* The more influencing factors we introduced, the larger the number of possible combinations, making the expressions even more lengthy. e.g. If we say that below a total weight of 6 is Low risk for 3 influencing factors, we would need to include all combinations for a total of > 6 like 1, 1+1, 1+1+1, 1+1+1+1, 1+1+1+1+1, 2+1+1+1 etc. A sample part of an expression used in risk categorization can be seen below in Figure 10.

Figure 10. Sample part of expression for Medium Risk Person

Due to time restrictions, we had to limit the number of factors used in the expressions to 3. This means that only 3 factors influencing an individual are considered. To build an expression for medium risk with 3 factors, more than 200 possible combinations were needed. Introducing more factors would increase the number of possible combinations making the expressions even lengthier. This would be a really difficult task to perform without using some tool to automate creating these expressions. Once these expressions were introduced into Protégé, the reasoner was able to categorize individuals into their appropriate risk categories.

With the introduction of these expressions the reasoner is also able to compute the inferred class type for the 3 categories of risk (High, Medium and Low). From Figure 11 below we can see that the *HighRiskPerson* is an inferred subclass of *MediumRiskPerson* and both these classes are inferred subclasses of *LowRiskPerson.* This is because of the fact that an individual at high risk is always at medium risk and low risk as well. An individual at medium risk is always at low risk.

Figure 11. Inferred Class types

Limitations with Protégé 3 use

The same weighted approach was used to introduce protective measures into the risk categorization. Table 2 below shows the weights used for protective measures. In theory a person at High risk using some protective measures can reduce his risk to medium or even to low. In Protégé 3, however, this theory does not work well when the protective measure expressions are included. This could be due to the fact that *MediumRiskPerson* is already an inferred super class of *HighRiskPerson* with respect to the open world reasoning. The assumption that a person under high risk would not be under medium or low risk does not fit according to this implementation of the ontology.

Table 2. Weight table for Protective Measures

Protective Measures		**Weights**
ProtectiveClothing		4
UsingSunscreens		
	SPF1-14	1
	SPF15-29	3
	SPF30plus	4
WearingHat		2
WearingUVProtectingSunglasses		3

Due to the limitations posed by Protégé 3 with introducing more influencing factors and including protective measures, we had to take a slightly different approach when we used protégé 4 as described in the next section.

Reasoning with Protégé 4

The other approach adopted by using protégé 4 is meant to simplify the classification of the ontology by introducing the weighted influencing factors directly in the OWL file which is impossible with Protégé 3. Protégé 4 allows us to do this since it uses OWL 2 which takes into account the data property assertions and is able to reason with those data properties.

The ontology is modified following three steps. The first one is to introduce the weighted values for the influencing factors directly in the individuals. The second is to construct value classes to group the different influencing factors that have the same value in a common class. The last step is to define the risk classes for persons by counting the number of relations a person has with the different value classes. These steps will be discussed in more detail in the following paragraphs.

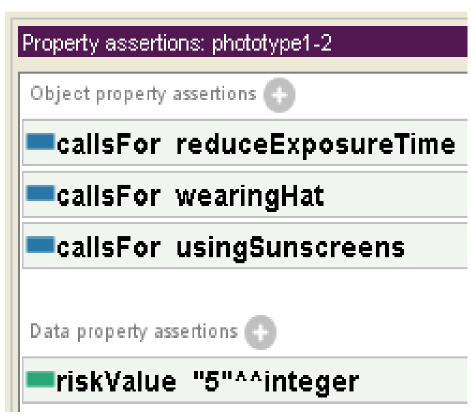

Figure 12. Using weighted factors

Figure 12 shows how we directly use the data property *riskValue* in the individual *phototype1-2*. The statement *riskValue "5"^^integer* expresses that *phototype1-2* has a *riskValue* relation with an integer of value 5. It represents the fact that a weight of 5 was chosen to express the risk level of the *phototype1-2* influencing factor. Similarly, a weight has been given to every influencing factor to represent its risk value.

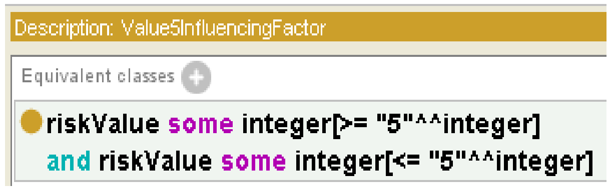

Figure 13. Classifying factors into value classes

Figure 13 shows how we classified the influencing factors into value classes. In this example, we describe the equivalent class of Value5InfluencingFactor which is meant to regroup all the individuals that have a risk value 5. This is done by the statement *riskValue some integer[>=”5”^^integer]* and *riskValue some integer[<=”5”^^integer]*. This means that *Value5InfluencingFactor* is equivalent to the intersection of the classes that have a *riskValue* relation with an integer greater than or equal to five with the classes that have a *riskValue* relation with an integer less than or equal to five. It can be interpreted as the class whose individuals have a risk value of five. The same was done for the other values by creating the classes *Value4InfluencingFactor*, *Value3InfluencingFactor*, *Value2InfluencingFactor* and *Value1InfluencingFactor*.

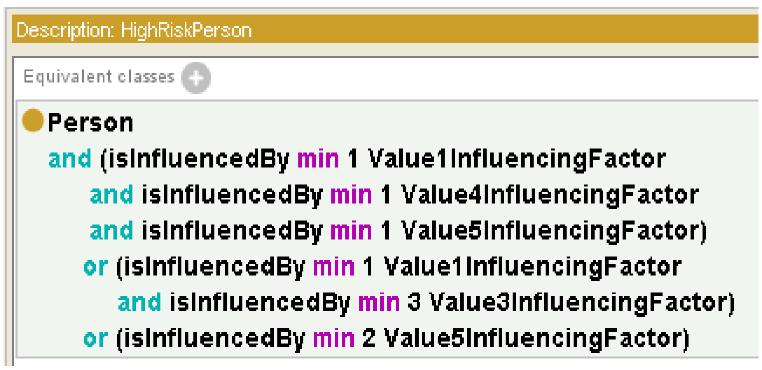

Figure 14. Define risk classes for the persons

An example of how we performed the last step (defining risk classes for a person) is shown in Figure 14. Here we express the equivalent class of *HighRiskPerson* with a statement meaning that a high risk person is influenced by at least one *Value1InfluencingFactor* and one *Value4InfluencingFactor* and 1 *Value5InfluencingFactor* or is influenced by at least one *Value1InfluencingFactor* and 3 *Value3InfluencingFactor* or is influenced by at least two *Value5InfluencingFactor*. By using similar expressions, we define the classes *MediumRiskPerson* and *LowRiskPerson*.

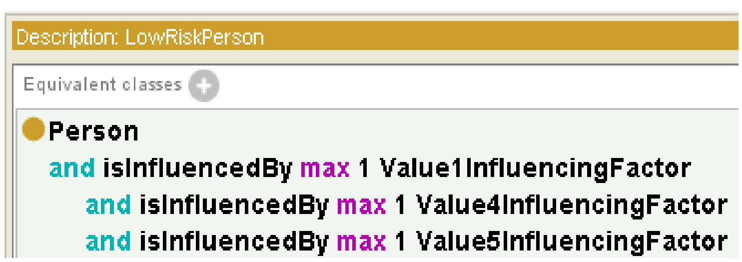

Figure 15. Problem example

But problems arise when it comes to using the maximum number of relations for *LowRiskPerson* and *MediumRiskPerson* instead of the minimum number as in Figure.15. We face the open world reasoning assumption which implies that we cannot assume something does not exist until it is explicitly stated that it doesn’t. In our example, we have defined many influencing factors, but we have not stated that these influencing factors are the only ones that exist in our ontology. The reasoner assumes that influencing factors other than *Value1InfluencingFactor*, *Value4InfluencingFactor* or *Value5InfluencingFactor* can exist in Figure.15 example. This explains why we did not have the expected results for the classification.

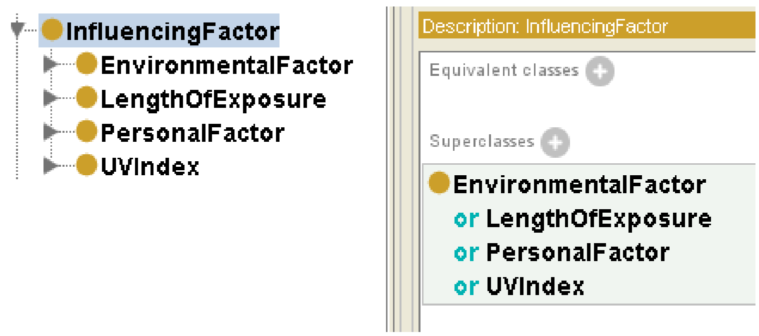

Figure 16. Class closure

Figure 16 shows how we perform a class closure over *InfluencingFactor*. This type of closure is similar to a closure axiom except that it is done over a class and not over a property. For example, we define "*EnvironmentalFactor or LengthOfExposure or PersonalFactor or UVIndex*" as the superclass of *InfluencingFactor*. This seems really strange at first sight because those classes are subclasses of *InfluencingFactor*. In fact, the "or" operator used can rather be seen as a union of the above mentioned classes and the superclass expression can be considered as the inclusion of the *InfluencingFactor* class. Figure 17 shows the InfluencingFactor (IF) set before (left hand-side) and after (right hand-side) applying the class closure. Before, we had stated that IF contained the classes A, B and C, but it could also contain other classes because of the open world reasoning assumption. By saying that IF is included in the union of A, B and C, we restrict the IF set and the reasoner takes into account the fact that no classes other than A, B or C are influencing factors in the ontology.

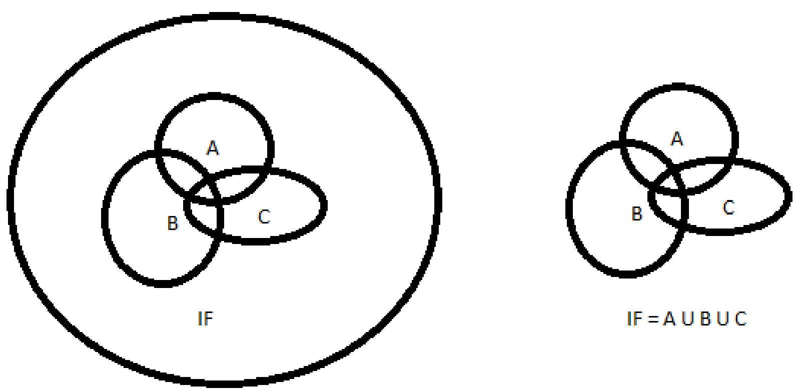

Figure 17. Inclusion statement

Even applying such closure was not enough to make the reasoning task work properly. Figure 18 shows that open world reasoning also forces us to explain the facts that do not exist. We do this by specifying for each individual of the class *Person*, both the relations it has and those it doesn't have with the different influencing factors. This is done by the statements of the form *isInfluencedBy max 0 AgeCategory* which states that *person4* is not influenced by any age category. By doing this, the maximum number of relations of *person4* with influencing factors is computed correctly during the reasoning task.

Figure 18. Individuals description

DESCRIPTION LOGIC QUERIES

Up to this point we had explained the design and the implementation of our ontology on Solar UV exposure and Human health. We also discussed the strategic decisions we made, like using weights for risk factors and using different approaches to implement this in Protégé 3 and 4.

This part mainly focuses on how knowledge can be retrieved from the ontology. The goal of this section is to show how we can get answers to questions such as: what diseases are caused by UV or what the influencing factors are or who is at high risk. This is done using the Pellet reasoner tool and the Description Logic (DL) query tab in Protégé. At this point, all of the reasoning tasks are completed by the reasoner and we are using the queries to lookup the facts. In the following section we use screenshots of DL queries in Protégé 4 to illustrate the reasoning results.

Using DL Queries in Protégé 4 to retrieve knowledge from the Ontology

The Protégé DL query tab combined with the Pellet reasoning tool are able to retrieve all instances or related classes, given a class expression query in DL. Therefore, any question which can be correctly translated into such class expression query will be answered with the facts that exist in the knowledge base. Of course, translating normal questions into DL queries involves knowing the taxonomy (classes and properties) of the ontology.

We present some examples of queries answered by the ontology with facts mainly inspired by the WHO factsheets which served as the starting point for this project. Figure 19 illustrates how to use DL queries on the ontology with an example.

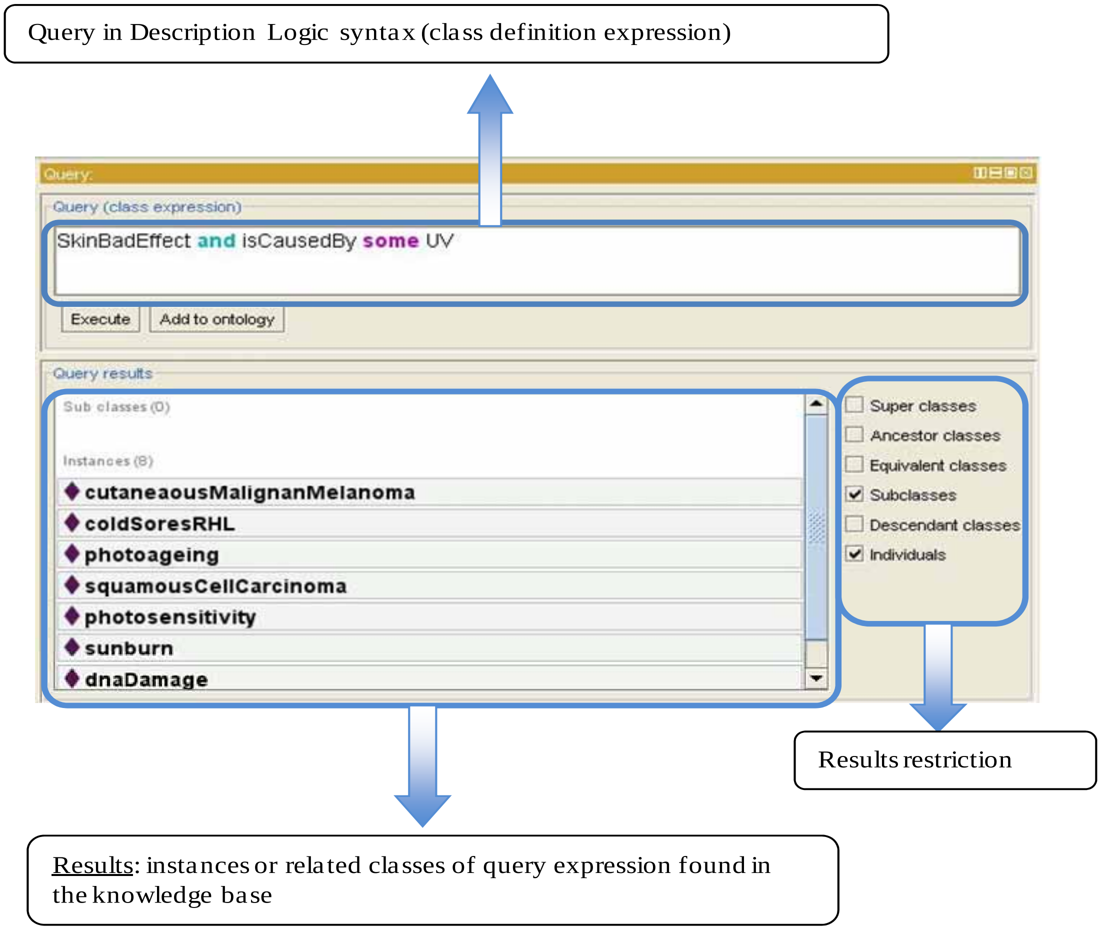

Figure 19. Overview of the Protégé 4 query tab

The query has to be in a standard DL form. The results are the individuals from the resulting fact classes. Any subclasses, superclasses or equivalent classes can also be retrieved easily by simply asking to display them in the results. Figure.20 illustrates a query for bad effects of UV rays answered with the facts from the ontology.

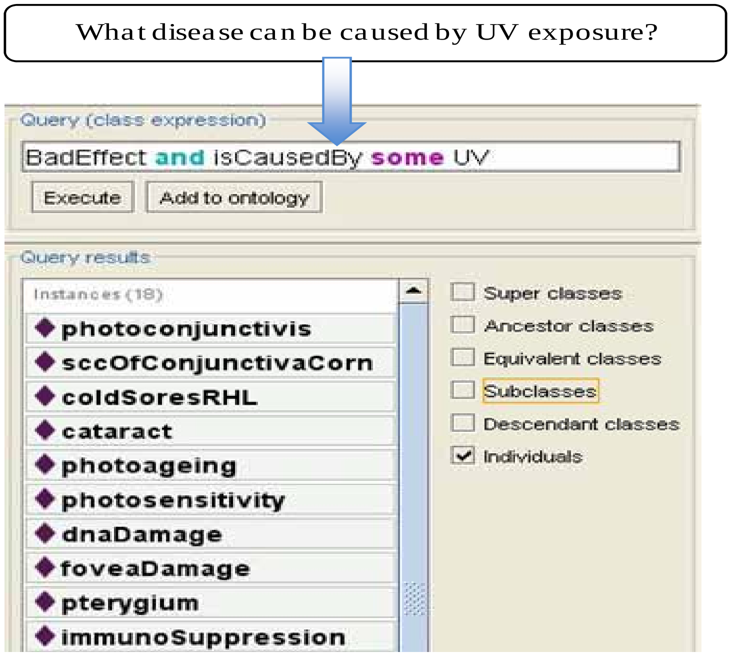

Figure 20. DL queries example 1

The Query for good effects in Figure.21 is answered in the same way we saw before.

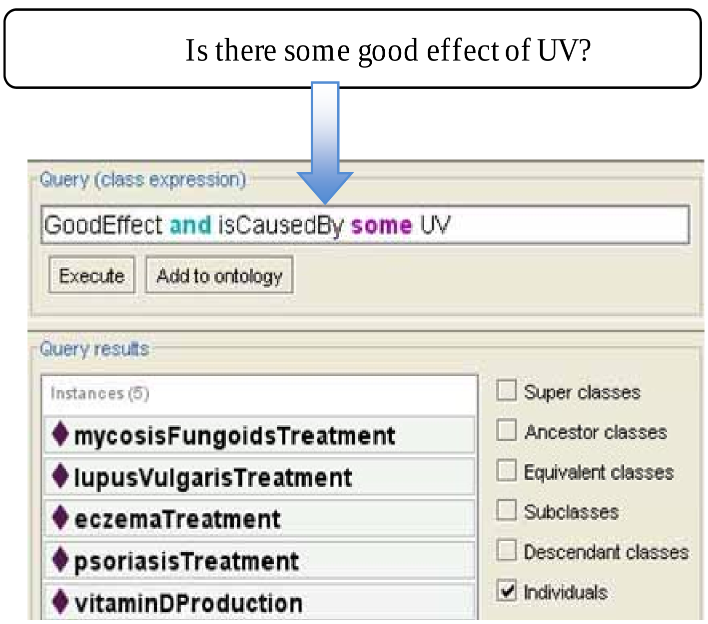

Figure 21. DL queries example 2

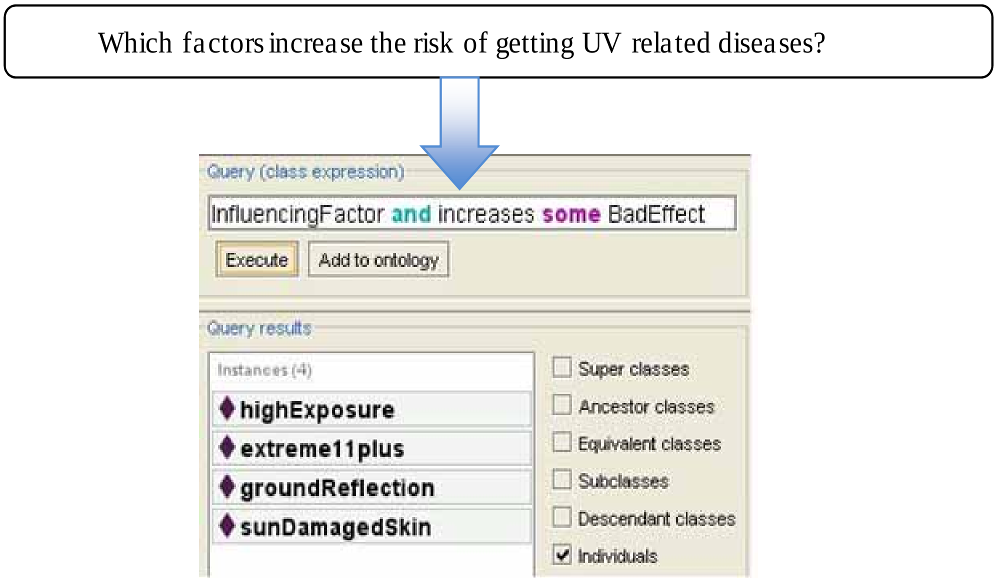

Figure 22. DL queries example 3

Figure.22 illustrates another query to find out the factors that increase the bad effects of UV rays. Figure.23 shows the protective measures that can be used to reduce the bad effects of UV rays. Thus, the ontology can be queried for any domain knowledge facts represented in it.

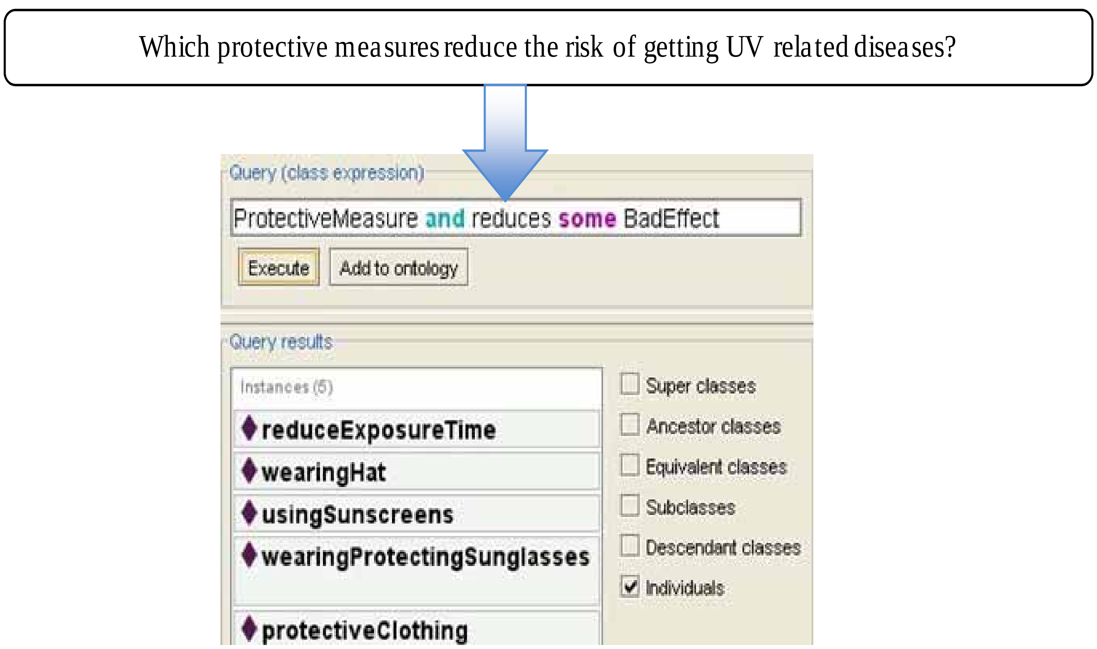

Figure 23. DL queries example 4

Querying Risk Categories using Protégé 4

As stated before, another capability of this ontology is to classify persons in the knowledge base into three risk categories given their influencing factors. To retrieve this information using DL query, we can simply ask who the high, medium or low risk persons in our knowledge base are.

If we want to know the risk category of a person say *person1*, we have to add this individual to our knowledge base with the corresponding properties regarding risk factors; say *person1* is less than 18 years old, has a skin photo type 1 and has many moles. Representing information about the person in the ontology is enough to be automatically classify their appropriate risk category based on the total risk weight. Here, the risk weight for *person1* is 5 + 5 + 4 > 12, which means that this person will appear under the high risk category. Figure.24 shows an example of all the persons classified under high risk.

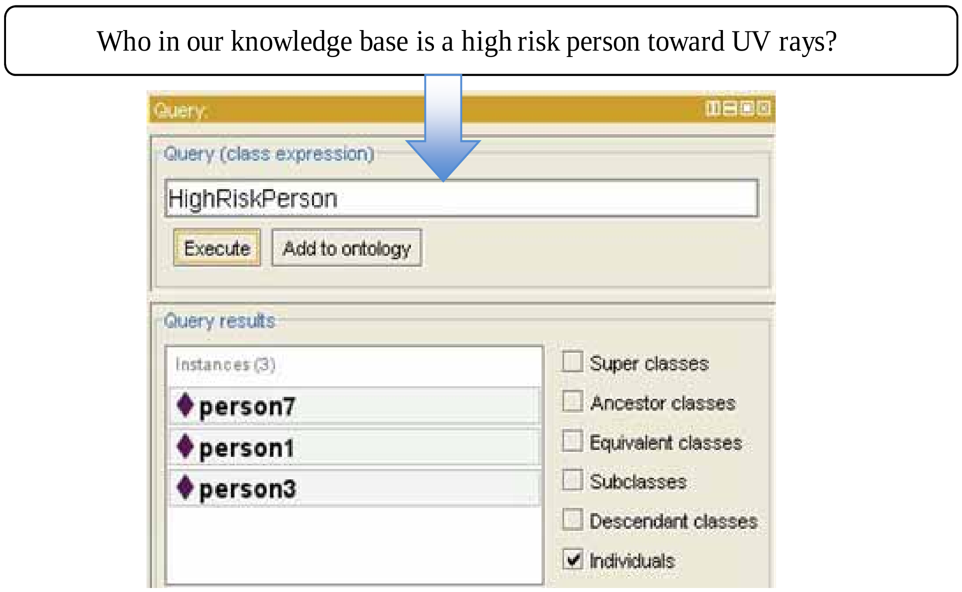

Figure 24. DL queries example 6

Querying Risk Categories using Protégé 3

The approach for querying the risk category in Protégé 3 is almost the same as in Protégé 4: representing the information about the person and determining which category he or she is classified into. Apart from the Protégé interface, which is quite different from version 3 to 4, the main differences between Protégé 3 and Protégé 4 implementations of the ontology are the following:

- We can represent only up to 3 risk factors of a person in Protégé 3, for the reasons we explained earlier
- A person is classified to the appropriate risk category, and also to all of the ones below his/her risk category. Therefore the exact risk category of a person is the highest risk category in which that person appears.

To explain the risk category of a person in Protégé 3, if we consider our earlier example of *person1*, this individual will be classified under high, medium and low risk. From this, we can state that the exact category for *person1* is high risk.

IMMINENT TRENDS

With the standardization of OWL2 about to be concluded by the W3C OWL working group, The OWL community already has OWL3 in mind and some suggestions are being discussed to improve OWL2. Efforts are being made to provide OWL with its own rule language which would fit better with the OWL paradigm than the standard rules paradigms do. Other challenges regarding the use of Boolean role constructors and the incorporation of DL-safe SWRL are being investigated *(P Hitzler 2009)*.

As for the Pellet reasoning services, some extensions are planned in the near future to provide better reasoning with large number of individuals and full Semantic Web Rule Language (SWRL) support. *(Sirin, E, et al. 2007)*

CONCLUSION

In this chapter we described how our team built the "Solar UV and Human Health" OWL ontology using Protégé. This process mainly involved adopting a suitable methodology and following its guidelines through the modeling and the implementation of the ontology. However, implementing such ontology was subject to the specificities of the tool used, and some strategic decisions had to be made to achieve our goal of building the OWL ontology.

The last section of this chapter about querying the ontology gives an idea of its possible uses. Indeed, these queries can be executed from a JAVA program using the ontology and lead to very useful applications such as a personal UV risk evaluator, a diagnosis assistant tool for medical professionals and so on. Besides, this ontology can also be extended to include more facts and reasoning (for example about UV related diseases and their treatment). It can inspire the design of other ontologies such as X-rays and Human Health etc. where the framework of this ontology can be reused. Several such modular ontologies can be combined to work as a knowledge base to be used in health care, tourism or other related fields. This ontology can also be combined with other Semantic Web tools like SWRL to leverage its knowledge.

References

Journals

Tai C. Chen, Farhad Chimeh, Zhiren Lu, et al. 2007. Factors that influence the cutaneous synthesis and dietary sources of vitamin D. Archives of Biochemistry and Biophysics, Volume 460, Issue 2, 15 April 2007, Pages 213-217.

Yasuhiro Matsumura and Honnavara N. Ananthaswamy. 2003. Toxic effects of ultraviolet radiation on the skin. Toxicology and Applied Pharmacology, Volume 195, Issue 3, 15 March 2004, Pages 298-308.

M Fernandez-Lopez, A Gomez-Perez, N Juristo.1997.Methontology: from ontological art towards ontological engineering, Proceedings of the AAAI97 Spring, 1997.

P. Hitzler .2009. Suggestions for OWL 3. In Proceedings of OWL Experiences and Directions, Sixth International Workshop, Washington D.C., USA, October 2009.

Sirin, E., Parsia, B., Cuenca, B. Grau, Kalyanpur, A., Katz, Y. 2007. Pellet: a practical OWL-DL reasoner. J. Web Semantics 5(2), 51–53 (2007).

Holger Knublauch, Mark A. Musen, Alan L. Rector. 2004. Editing Description Logic Ontologies with the Protégé OWL Plugin. International Workshop on Description Logics - DL2004, Whistler, BC, Canada - A technical discussion for logicians.

N. F. Noy and D. L. McGuinness. 2001. Ontology development 101: A guide to creating your first ontology. Technical Report KSL-01-05, Stanford Knowledge Systems Laboratory.

Bernardo Cuenca Grau, Ian Horrocks, Yevgeny Kazakov, et al. 2007. A Logical Framework for Modularity of Ontologies. International Joint Conferences on Artificial Intelligence, IJCAI-07, Technical Papers and Posters, Pages 298-303.

Heiner Stuckenschmidt, Michel Klein, 2007. Reasoning and change management in modular ontologies. Data & Knowledge Engineering, Volume 63, Issue 2, November 2007, Pages 200-223.

Online Documents

W3C UK and Ireland Office http://www.w3c.rl.ac.uk/pasttalks/slidemaker/EPS_DTI/slide11-0.html (accessed November 5, 2009).

thefreedictionary.com http://www.thefreedictionary.com/ontology (accessed November 5, 2009).

Ontologies and Semantic Web http://www.obitko.com/tutorials/ontologies-semantic-web/ (accessed November5, 2009).

UNEP (United Nations Environmental Programme), Global Solar UV. Index, http://www.unep.org/PDF/Solar_Index_Guide.pdf. (accessed October 26, 2009).

Matthew Horridge, Simon Jupp, Georgina Moulton, Alan Rector, Robert Stevens, Chris Wroe. 2007. A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools, Edition 1.1. Protégé User Documentation http://protege.stanford.edu/doc/users.html#tutorials (accessed October 10, 2009).

WHO factsheets (http://www.who.int/mediacentre/factsheets/fs305/en/index.html (accessed October 6, 2009).

Paolo Ciccarese, Tim Clark, Marco Ocana. 2009. Semantic Web Application in Neuromedicine (SWAN) Ontology. http://www.w3.org/TR/hcls-swan/ (accessed March 4, 2010).