**Medical Entity vocabulary extension proposal of schema.org**

1. **The goal:**

To extend the medicalEntity vocabulary of schema.org.

1. **Rationale:**

MedicalEntity and subtypes vocabulary provided in schema.org are largely used by content publishers that wish to mark up health and medical content on the web. Like all schema.org schemas, the health and medical schema is intended to make it easier for people to find the right web pages by exposing structured information contained in web pages to search engines, and may also enable other applications that make use of the structure.

However, given the complexity of health and medical content, some classes and predicates were missing, or incompletely defined to achieve this indexing use. Today, most of health and medical web content are structured documents, based on clinical information models or HL7 standards (HL7 V3 RIM, CDA, C-CDA, FHIR etc.) and are integrating existing vocabularies and terminology (SNOMED, ICD, LOINC, ATC, RxNorm etc.).

Moreover, some case use scenarios go beyond this simple indexing: schema.org has been used in recent health APIs, as a pillar source of medical ontology/vocabulary for formalization of healthcare information, through RDF graphs which is a high contribution to the semantic web interoperability. To meet those challenges it was necessary to extend the MedicalEntity vocabulary and take this opportunity to refine some of existing concept (classes and predicates).

The intention here is neither to have a comprehensive medical ontology nor provide a clinical information model nor define or codify a new controlled medical vocabulary, but to provide a scalable, use-driven vocabulary schema, in-line with the schema.org philosophy.

1. **Approach used:**

In this design, none of existing concepts (classes and predicates) are renamed or deleted. There are only additions of new classes/predicates and extensions (definitions or domain and range) on the existing concepts. For granular classes the priority is given to use the schema.org vocabulary in clinical information models which use classes from existing vocabularies and terminology (SNOMED, ICD, LOINC, ATC, RxNorm etc.) so the new concepts are minimized. This is a scalable approach as schema.org can’t include all those granular concepts.

Additional, staying in schema.org best practice and philosophy we avoided some discussions interfering with high-level schema.org design. Finally to refine existing concepts or for a better integration of new concepts, there was hierarchy rebuild, like declaring existing schema classes/properties as subClass of or subPropertyOf.

1. **High level overview of the principle changes**

The proposal contains 3 types of changes:

1. Changes: only hierarchies (purpose: corrections of existing hierarchy or for integrating new proposed classes/predicates)
   1. Class : 18
   2. Predicates : 9
2. Deletions: 0
3. Additions/extension
   1. Class : 172 (Overview [here](http://webprotege.stanford.edu/#Edit:projectId=d14a74e0-3328-4079-8b60-6b2bf11fd313))
   2. Predicates : 173 (Overview [here](http://webprotege.stanford.edu/#Edit:projectId=d14a74e0-3328-4079-8b60-6b2bf11fd313))

A table showing all changes details is provided below, but in nutshell changes are:

***Classes:***  
-Added AnatomicalOrgan and AnatomicalRegion classes under hierarchy of AnatomicalEntity class and related properties  
-Extended subclass of AnatomicalSystem class and related properties  
-Extended subclass of MedicalSignOrSymptom class and related properties  
-Big extension in MedicalProcedure with a one change (actually additional declaration) of making the MedicalTherapy a subclass of TherapeuticProcedure class, itself a subclass of MedicalProcedure class. The PreventiveProcedure become a subclass of MedicalProcedure class as well.  
-extending Person class with Patient class as subclass.  
-making the Drug also a subclass in low level hierarchy of Substance (the new class).

- Extended subclass MedicalEnumeration (some of subclasses are the existing ones: see below table).

-New classes added.

***Properties:***  
-adverseOutcome, location, provider, medicalSpeciality and procedure were extended with granular subproperties  
-status was extended with 2 granular subproperties

-introduction of Snomed concepts use as domain/range objects (not always but where there was no clear similar schema.org existing class)

-new properties added.

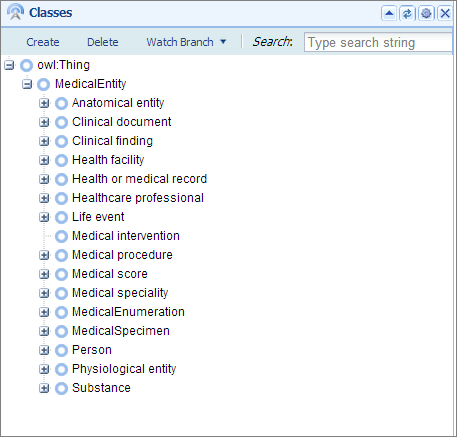
Comparing the *existing* schema.rdfa file and the *proposed* schema.rdfa patch, there are in total 2578 insertions (+), 59 deletions (-), 25 fixes (changes in *subPropertyOf*/*subClassOf*/*rdfs:comment* in existing properties/classes) , 0 class/predicate rename and 0 class/predicate deletions.

We are aware that some improvement may be required even beyond official publication, but those will be minimal.

***Table1. The changeLog details:***

|  |  |  |
| --- | --- | --- |
| **Current version** | **Proposal version** | **Rationale** |
| :AnatomicalStructure  *subClassOf*  :MedicalEntity | :AnatomicalStructure  *subClassOf*  : AnatomicalEntity | *AnatomicalEntity* was suggested as parent of all anatomical related concepts what is more scalable if one day we need to extend anatomy vocabulary |
| : AnatomicalSystem  *subClassOf*  :MedicalEntity | : AnatomicalSystem  *subClassOf*  : AnatomicalEntity | *cfr supra* |
| : MedicalProcedure  *rdfs:comment* | *rdfs:comment* Definition extended | The definition change to cover all medical related procedure |
| : MedicalTherapy  *subClassOf*  : MedicalEntity | : MedicalTherapy  *subClassOf*  : TherapeuticProcedure | *TherapeuticProcedure* was suggested as parent of all therapeutic concepts in scope of rebuilding *MedicalProcedure* entity (mainly to comply with medical terminologies: Snomed and HL7 standards) |
| : Drug  *subClassOf*  :MedicalEntity | *rdfs:comment* Definition extended  : Drug  *subClassOf*  : PharmaceuticalSubstance | *PharmaceuticalSubstance* was suggested as parent of all Pharmacological substance which extends the vocabulary |
| : DrugClass  *subClassOf*  : MedicalTherapy | : DrugClass  *subClassOf*  : MedicalEnumeration | Since the *MedicalTherapy* is defined as a medical procedure it cannot have a *DrugClass (defined here as* list item/concept*)*as children. |
| : MedicalEnumeration  *rdfs:comment* | *rdfs:comment* Definition extended | This concept class seems to be the same as the medical observation defined with Snomed. The definition was updated to reflect its hierarchy content. |
| : MedicalCondition  *subClassOf*  :MedicalEntity | : MedicalCondition  *subClassOf*  : ClinicalFinding | *ClinicalFinding* was suggested as parent of all medical condoition (disorders) concepts what is in line with Snomed terminology. |
| :AnatomicalStructure  *subClassOf*  :MedicalEntity | :AnatomicalStructure  *subClassOf*  : AnatomicalEntity | *AnatomicalEntity* was suggested as parent of all anatomical related concepts what is more scalable if one day we need to extend anatomy vocabulary |
| : Patient  *subClassOf*  : MedicalAudience | : Patient  *subClassOf*  : MedicalAudience  : Patient  *subClassOf*  : Person |  |
| : MedicalClinic  *subClassOf*  : MedicalOrganization | : MedicalClinic  *subClassOf*  : MedicalOrganization  : MedicalClinic  *subClassOf*  : HealthFacility | *HealthFacility* was suggested as parent of all health related premises what is more covering other places like medical room, wards, etc. |
| : Ultrasound  *subClassOf*  : MedicalImagingTechnique  : MedicalImagingTechnique  *subClassOf*  : MedicalEnumeration | : Ultrasound  *subClassOf*  : MedicalImaging | *MedicalImaging* is the most correct parent as it’s defined here as a medical procedure. |
| : SurgicalProcedure  *rdfs:comment* | *rdfs:comment* Definition updated |  |
| : MedicalSignOrSymptom  *subClassOf*  :MedicalEntity  : MedicalSignOrSymptom  *rdfs:comment* | : MedicalSignOrSymptom  *subClassOf*  : MedicalCondition  *rdfs:comment* Definition updated | Following the Snomed hierarchy. |
| : MedicalSymptom  *rdfs:comment* | *rdfs:comment* Definition updated | Disambiguation (vs MedicalSign) |
| : PhysicalExam  *subClassOf*  : MedicalEnumeration  : MedicalSymptom  *rdfs:comment* | : PhysicalExam  *subClassOf*  : MedicalProcedure  *rdfs:comment* Definition updated | *PhysicalExam* is here defined as a procedure. So is fitting well in medical procedure hierarchy and not in enumeration. |
| : VitalSign  *subClassOf*  : PhysicalExam | : VitalSign  *subClassOf*  : MedicalSign | *VitalSign* is here defined as a sign. So is fitting well in medical sign hierarchy and not in physical exam (which is a medical procedure). |
| : PsychologicalTreatment  *subClassOf*  : MedicalTherapy | : PsychologicalTreatment  *subClassOf*  : TherapeuticProcedure | *PsychologicalTreatment* is is fitting well in *TherapeuticProcedure* hierarchy and not in *MedicalTherapy*. |
| : cause  *rdfs:comment* | *rdfs:comment* Definition updated |  |
| : comprisedOf  *rdfs:comment* | *rdfs:comment* Definition updated |  |
| : diagnosis  *:domainIncludes* | *domainIncludes*  Domain/range extended  *rdfs:comment* Definition updated |  |
| : clincalPharmacology | *:clinicalPharmacology* | Typo error fixing (in published version) |
| *adverseOutcome*, *location*, *provider*, *medicalSpeciality*, *procedure*, |  | extended with granular subproperties |

1. **Current overview of the hierarchy**
   1. **Class hierarchy overview** (Can be seen in full [here](http://webprotege.stanford.edu/#Edit:projectId=d14a74e0-3328-4079-8b60-6b2bf11fd313))



* 1. **Class hierarchy details** (the displayed text reflects labels not item name)

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |