

XML i el web semàntic

Projecte OntoTrauma

Definició d'una ontologia d'aprenentatge
en l'àrea de cirurgia traumatològica

Treball fi de carrera

Enginyeria tècnica en informàtica de sistemes



Universitat Oberta
de Catalunya

www.uoc.edu

Nom: Jordi Varela Agrelo

Consultor: Joan Anton Perez Braña

10 de gener de 2012

Resum

Facilitar l'accés a la informació va ser un dels principis que va fonamentar el naixement del web, i el fil conductor durant la seva evolució fins els nostres dies. Son indubtables les fites aconseguides durant els seus escassos 30 anys d'existència tant en el volum d'informació generada com consumida.

Les tecnologies web semàntiques han anat evolucionant durant aquest temps, fins a perfilar-se en evolució natural a la web actual per a permetre a les màquines entendre l'estructura de la informació que s'emmagatzema al web, obrint la porta a noves possibilitats.

A part del seu ús en la web pública, les tecnologies semàntiques aporten un model d'explotació de dades que pot ser de gran interès també per les empreses.

Aquest treball s'emmarca en aquest context, amb l'objectiu de l'estudi de les tecnologies del web semàntic i l'elaboració d'una aplicació pràctica que reculli la informació necessària dintre d'un procés quirúrgic traumatològic.

Generarem una ontologia per modelar l'estructura de les dades i farem ús d'un wiki semàntic per la generació i l'emmagatzematge de la seva informació.

Índex

| | |
|-----------------------------------------------------------------------------|-----------|
| Resum | 3 |
| 1 Presentació del projecte | 2 |
| 1.1 Justificació i aportació del TFC | 3 |
| 1.2 Objectius del TFC | 3 |
| 1.3 Enfocament i planificació | 4 |
| 1.4 Productes obtinguts | 7 |
| 1.5 Descripció dels següents capítols | 7 |
| 2 El web semàntic | 8 |
| 2.1 Evolució del web | 8 |
| 2.1.1 Què és una ontologia? | 10 |
| 2.1.2 Aplicacions sobre el web semàntica | 12 |
| 2.2 Tecnologies del web semàntic | 13 |
| 2.2.1 Llenguatges de descripció | 15 |
| 3 Elaboració de la ontologia ontoTrauma | 19 |
| 3.1 Anàlisi d'ontologies quirúrgiques traumatològiques | 19 |
| 3.2 Metodologia per a la creació d'una ontologia | 20 |
| 3.3 Creació de l'ontologia ontoTrauma | 23 |
| 3.3.1 Determinació de requeriments | 23 |
| 3.3.2 Elaboració del model conceptual | 25 |
| 3.3.3 Implementació del model conceptual amb Protégé | 32 |
| 4 Elaboració de la base de dades de coneixement: els wikis semàntics | 37 |
| 4.1 Sematic Media Wiki | 38 |
| 4.2 Semantic Media Wiki Plus | 41 |
| 4.3 OntoWiki | 42 |
| 4.4 Diferències i selecció de programari | 44 |
| 5 Conclusió i línies de futur | 47 |
| Bibliografia | 48 |

Índex de figures

| | | |
|------|------------------------------------------------------------------------------------------|----|
| 1.1 | Versions del web | 2 |
| 2.1 | Evolució de les tecnologies del web semàntic | 8 |
| 2.2 | Linked Data | 13 |
| 2.3 | Pila semàntica presentada per Tim Berners-Lee | 13 |
| 2.4 | Visualització d'una pila de tecnologies semàntiques | 14 |
| 2.5 | Descripció gràfica d'un document XML | 15 |
| 3.1 | Relació d'ontologies proposades a [Bur09] segons el seu nivell d'abstracció | 20 |
| 3.2 | Descripció del mètode de desenvolupament plantejat per [Alv10] | 22 |
| 3.3 | Diagrama d'activitat d'una operació quirúrgica traumatològica | 23 |
| 3.4 | Diagrama de casos d'ús involucrats en l'etapa de diagnosi | 24 |
| 3.5 | Diagrama dels casos d'ús involucrats dintre d'un tractament quirúrgic | 24 |
| 3.6 | Diagrama de la classe Persona | 25 |
| 3.7 | Diagrama de classes d'ontoTrauma | 27 |
| 3.8 | Diagrama de relacions de ontoTrauma | 28 |
| 3.9 | Classes i propietats de Persona | 29 |
| 3.10 | Classes i propietats de Material | 29 |
| 3.11 | Classes i propietats de Diagnosi | 30 |
| 3.12 | Classes i propietats de Procediment | 30 |
| 3.13 | Classes i propietats de Tractament | 31 |
| 3.14 | Definició de metadada Protégé | 32 |
| 3.15 | Definició de classes Protégé | 33 |
| 3.16 | Definició de relacions Protégé | 34 |
| 3.17 | Definició d'atributs Protégé | 35 |
| 3.18 | Definició d'instàncies Protégé | 36 |
| 4.1 | Edició millorada en SMW+ | 41 |
| 4.2 | Incorporació ontologia a SMW+ | 42 |
| 4.3 | Incorporació ontologia a Ontowiki | 44 |
| 4.4 | Rendiment de diversos triplestores en una consulta sobre la DBPe- dia [Bec] | 46 |

1 Presentació del projecte

Context

El web, el sistema de distribució d'informació basat en l'hipertext ideat per Tim Berners-Lee al 1989 [BL09], està sotmès a una revisió constant de la seva pròpia concepció i forma en què es gestiona la seva informació.

Alguns factors d'aquest dinamisme són, probablement: el gran ritme de creixement de la informació emmagatzemada, l'evolució de les tecnologies que la sustenten i exploten i sobretot, l'interès i ús propiciat per part dels consumidors de la seva informació.

Un indicador d'aquest ritme de canvi és clarament visible a través d'aquestes tecnologies que giren en el seu entorn i que, progressivament, milloren o simplement, es substitueixen per obsolescència.

Sembla apropiat parlar de l'existència de diverses “versions” del web intentant agrupar les diverses tecnologies que són madures en una determinada “època” web, tot i que existeix molta controvèrsia al voltant d'aquesta visió del web [Bra09].

Segons aquesta explicació del web, ens trobaríem situats sobre l'anomenat web 2.0, - terme assignat per Tim O'Reilly durant una conferència al 2004 - referint-nos a la web dinàmica i social, i estaríem a les portes del web 3.0, generalment conegut com el web semàntic.

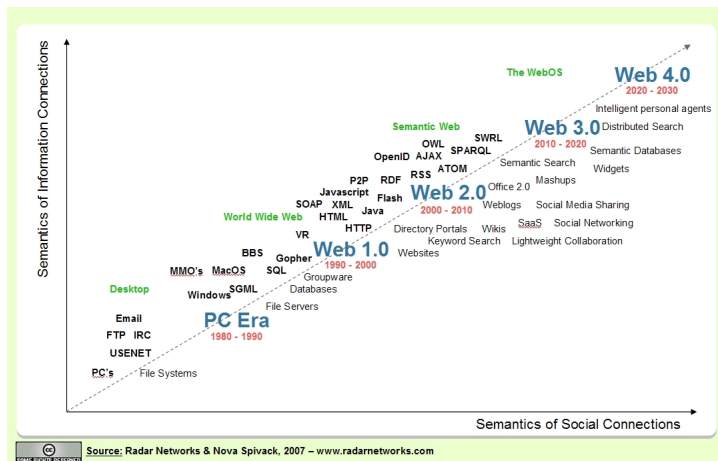


Figura 1.1: Versions del web

El gran volum d'informació existent actualment al web, fa que cada cop sigui més costós el seu aprofitament, perdent cada cop més valor i interès.

Precisament, un dels principals objectius del web semàntic és dotar de significat a les dades en si mateixes, aportant una estructura que permeti el seu tractament automatitzat i possibilitat d'obtenció del seu màxim profit; fent més fàcil la seva cerca, compartició i combinació.

La consultora Gartner, al seu anàlisi anual de tendències tecnològiques (hypecycle), va presentar l'any 2006 la xarxa semàntica corporativa com una tecnologia emergent en un termini de 5 a 10 anys (és a dir, entre l'any 2011 i 2016) [Gar06].

La forta penetració de la telefonia mòbil del tipus smartphone que s'ha produït entre el 2010 i 2011, juntament amb la maduresa de les tecnologies basades en XML, fa intuir la necessitat d'un canvi de paradigma que permeti millorar l'explotació de les dades independentment del dispositiu, amb restriccions de computació i ample de banda. La xarxa semàntica pot aportar una solució a aquestes necessitats.

1.1 Justificació i aportació del TFC

En aquest context, la necessitat de generar fonts d'informació de prou qualitat per ser explotades de forma semàntica és clau, tal com mostren les diverses propostes de Linked-Data.

Aquest TFC s'emmarca en el context de les ciències de la salut, concretament en l'àrea de cirurgia traumatològica, amb la intenció d'aprofundir en les eines necessàries per poder emmagatzemar una font d'informació especialitzada en les operacions quirúrgiques, la seva explotació i la seva possible interconnexió amb altres fonts de dades existents.

Un cop realitzat, ens trobarem en posició d'emetre unes conclusions fonamentades sobre l'interès, maduresa, i la conveniència de l'us de xarxes semàntiques aplicades en el camp de la medicina.

1.2 Objectius del TFC

El principal objectiu d'aquest treball és la introducció als conceptes relacionats amb el web semàntic i la consecució d'un producte funcional, basat en ontologies. Més concretament:

- Estudiar els conceptes relacionats amb la web semàntica

- Elaboració d'una ontologia semàntica dintre de l'àmbit mèdic
- Conèixer l'organització i estructura dels wikis semàntics i el seu ús
- Ús de la ontologia generada dintre d'un wiki semàntic

1.3 Enfocament i planificació

Es tracta d'un projecte amb les següents objectius:

- Estudi de les tecnologies semàntiques
- Desenvolupament d'una ontologia amb l'eina protégè
- Estudi i ús d'un wiki semàntic (mediawiki/ontowiki)
- Incorporació de la ontologia al wiki semàntic

L'abordarem en dues fases, coincidint amb l'entrega de les pac:

- Primera fase (PAC2): marcada pels conceptes de la xarxa semàntica i l'elaboració de la ontologia
- Segona fase (PAC3): relacionada amb les eines d'emmagatzematge i la incorporació de dades a la seva estructura

En cada fase s'incorpora una entrega prèvia la setmana abans de l'entrega final, amb l'objectiu d'avaluar amb el consultor el correcte avançament del projecte i disposar de temps per corregir les possibles desviacions que es puguin produir.

Planificació

Farem servir les següent franges horàries:

- dilluns, dimecres i dijous: 18:00-21:00
- dimarts i divendres: no dispo d'un horari fix, es destinarà principalment per la lectura de documentació
- Caps de setmana i festius: inicialment no estan contemplats, es faran servir puntualment per recolzar els dimarts/divendres que no s'hagin pogut dedicar i en general per evitar desviacions de projecte, en cas que sigui necessari

Donades les circumstàncies familiars del projectista s'han descartat també dies específics de vacances.

1.3 Enfocament i planificació

Detall

Fase 1: Pla de treball (30/09/11 – 03/10/11)

- Recollida de requisits
- Realització del document del pla de treball.

| | Nombre de tarea | Duración | Comienzo | Fin | 11 | 03 oct '11 | | | | | | | |
|---|----------------------------------------------|----------|--------------|--------------|----|------------|---|---|---|---|---|---|---|
| | | | | | J | V | S | D | L | M | X | J | V |
| 1 | [-] Fase 1:Pla de treball | 2 días | vie 30/09/11 | lun 03/10/11 | | | | | | | | | |
| 2 | Recollida de requisits | 1 día | vie 30/09/11 | vie 30/09/11 | | | | | | | | | |
| 3 | Realització del document del pla de treball. | 1 día | lun 03/10/11 | lun 03/10/11 | | | | | | | | | |

Fase 2: Desenvolupament ontologia (4/10/11 – 7/11/11)

- Estudi XML: Conceptes generals DTD, XML-Schema, RDF, OWL i SPARQL
- Estudi sobre ontologies, i ontologies existents relacionades amb l'àmbit d'aplicació
- Estudi sobre els conceptes amb l'àmbit d'aplicació i fonts d'informació
- Programari PROTÉGÉ: Instal·lació i configuració
- Elaboració d'una ontologia i una instanciació d'exemple
- Redacció i revisió de la documentació
- Pre-entrega PAC2
- Correcció PAC2 incorporant les indicacions del consultor
- Entrega PAC2

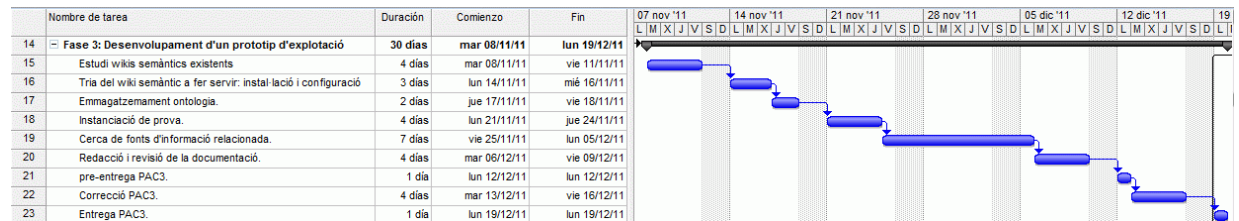
| | Nombre de tarea | Duración | Comienzo | Fin | 03 oct '11 | 10 oct '11 | 17 oct '11 | 24 oct '11 | 31 oct '11 | 07 nov '11 | | | | | | | | | | | | | | | | |
|----|----------------------------------------------------------|----------|--------------|--------------|------------|------------|------------|------------|------------|------------|---|---|---|---|---|---|---|---|---|---|---|---|---|--|--|--|
| | | | | | L | M | X | J | V | S | D | L | M | X | J | V | S | D | L | M | X | J | V | | | |
| 4 | [-] Fase 2: Desenvolupament ontologia | 25 días | mar 04/10/11 | lun 07/11/11 | | | | | | | | | | | | | | | | | | | | | | |
| 5 | Estudi XML | 4 días | mar 04/10/11 | vie 07/10/11 | | | | | | | | | | | | | | | | | | | | | | |
| 6 | Estudi sobre ontologies. | 5 días | lun 10/10/11 | vie 14/10/11 | | | | | | | | | | | | | | | | | | | | | | |
| 7 | Estudi sobre els conceptes relacionats. | 3 días | lun 17/10/11 | mié 19/10/11 | | | | | | | | | | | | | | | | | | | | | | |
| 8 | Programari PROTEGE. | 1 día | jue 20/10/11 | jue 20/10/11 | | | | | | | | | | | | | | | | | | | | | | |
| 9 | Elaboració d'una ontologia i una instanciació d'exemple. | 4 días | vie 21/10/11 | mié 26/10/11 | | | | | | | | | | | | | | | | | | | | | | |
| 10 | Redacció i revisió de la documentació. | 2 días | jue 27/10/11 | vie 28/10/11 | | | | | | | | | | | | | | | | | | | | | | |
| 11 | pre-entrega PAC2. | 1 día | lun 31/10/11 | lun 31/10/11 | | | | | | | | | | | | | | | | | | | | | | |
| 12 | Correcció PAC2. | 4 días | mar 01/11/11 | vie 04/11/11 | | | | | | | | | | | | | | | | | | | | | | |
| 13 | Entrega PAC2. | 1 día | lun 07/11/11 | lun 07/11/11 | | | | | | | | | | | | | | | | | | | | | | |

Fase 3: Desenvolupament d'un prototip d'explotació (8/11/11 – 19/12/11)

- Estudi wikis semàntics existents
- Tria del wiki semàntic a fer servir: Instal·lació i configuració
- Emmagatzemament ontologia

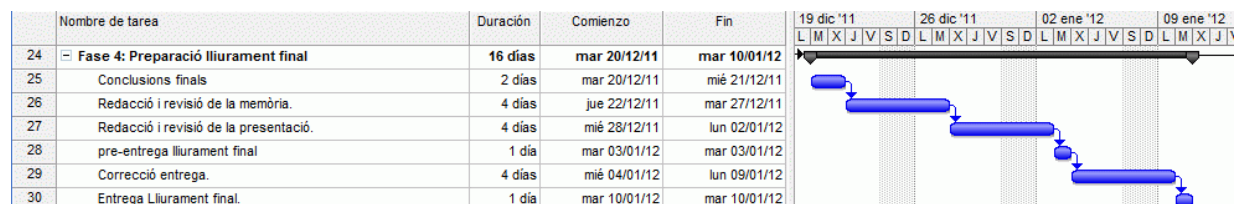
1.3 Enfocament i planificació

- Instanciació de prova
- Cerca de fonts d'informació relacionada per millorar l'instanciació
- Redacció i revisió de la documentació
- Pre-entrega PAC3
- Correcció PAC3 incorporant les indicacions del consultor
- Entrega PAC3



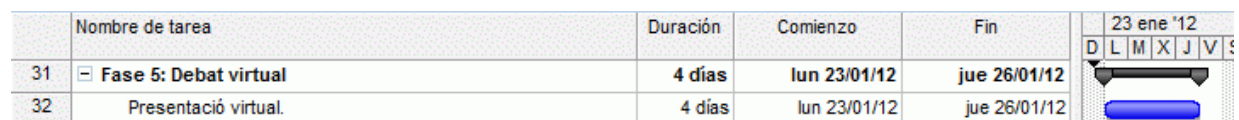
Fase 4: Preparació lliurament final (20/12/11 – 10/01/12)

- Conclusions finals
- Redacció i revisió de la memòria
- Redacció i revisió de la presentació
- Pre-entrega lliurament final
- Entrega Lliurament final



Fase 5: Debat virtual (23/01/12 – 26/01/12)

- Presentació virtual



Anàlisi de riscos

Cal tenir en compte els següents riscos associats a compaginar estudis, feina i família:

- A nivell laboral: es poden donar puntes de treball no previstes que afectin al projecte.
 - Per minimitzar aquest risc compto com contingència els caps de setmana, i en cas de detectar una possible desviació substancial de la planificació, em posaré en contacte amb el consultor per avaluar una possible retallada de funcionalitat i replanificació de projecte.
- A nivell personal i familiar.
 - Per minimitzar aquest risc, s'han tingut en compte al calendari les dates crítiques relatives a la seva escolarització, i algunes d'opcionals. És molt probable que hagi de reorganitzar tasques per compatibilitzar-les amb l'elaboració del projecte.

1.4 Productes obtinguts

Com a resultat d'aquest TFC s'obtidran els següents productes:

- Una ontologia en format OWL amb el model de domini que modeli les operacions traumatològiques proposades.
- Sel·lecció, configuració i integració de la ontologia a un wiki semàntic que permeti la seva explotació.

1.5 Descripció dels següents capítols

Començarem introduint-nos al capítol 2 en els conceptes de la web semàntica que seran d'aplicació en aquest projecte.

Al capítol 3 aprofundirem al concepte d'ontologia, i aplicarem un mètode per aconseguir modelar l'àmbit de coneixement de les operacions traumatològiques.

Un cop tinguem l'ontologia, per poder-la posar en pràctica, farem ús d'un wiki semàntic. En el capítol 4 plantejarem els wikis semàntics mes oportuns per aquest àmbit, sel·leccionarem un i l'adaptarem per incorporar la nostra ontologia.

2 El web semàntic

2.1 Evolució del web

El web (World Wide Web o WWW) va néixer com un espai per millorar l'intercanvi d'informació de la comunitat científica. Amb aquest objectiu, Tim Berners-lee va proposar l'ús del sistema d'hipertext per la publicació de textos científics, com a resultat d'un encàrrec mentre treballava al CERN a l'any 1980, donant vida al que mes tard seria HTML.

Al 1993 el CERN va decidir obrir la xarxa acadèmica per usos empresarials, provocant un creixement gairebé exponencial de múltiples facetes de la xarxa: tant en número de persones que hi accedien, com en número de planes publicades, formats, etc.

L'impacte d'ús es va veure reflectit al llenguatge HTML, que va anar-se adaptant a les noves demandes amb constants revisions.

Aquesta primera etapa del web, coneguda posteriorment com el “web 1.0”, es pot caracteritzar des del punt de vista tecnològic per les pàgines web amb escàs dinamisme que es feien a partir dels llenguatges mes comuns de l'època (principalment: HTML4.01, Javascript, Flash i Java). Des del punt de vista de la generació de

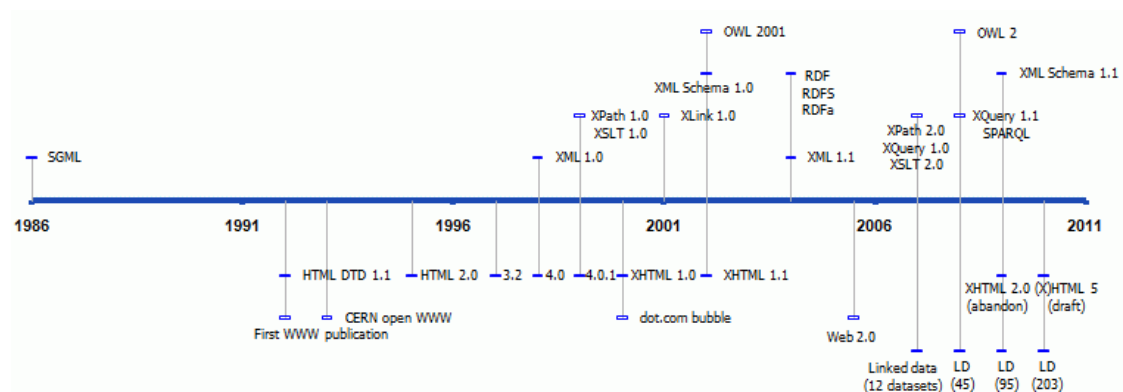


Figura 2.1: Evolució de les tecnologies del web semàntic

continguts, era remarcable el nivell de coneixements tècnics necessaris per fer servir les eines i conèixer els llenguatges de programació per poder crear contingut web (figura del programador - webmaster), i des de la vesant de la compartició d'informació, aquesta només es produïa a través de l'ús de l'hiperenllaç.

Amb el temps la tecnologia va anar madurant i simplificant la forma de crear noves pàgines web, baixant progressivament la barrera dels coneixements tècnics necessaris per aportar contingut a la xarxa fins al punt de fer-la mínima. Moment al qual es va produir un segon boom en accés i producció de continguts a la xarxa, al que se li va donar el nom de “web 2.0” [O'R05]. Grans projectes col·laboratius com la Viquipèdia, els blocs, les xarxes socials, en són els representants més clars d'aquesta nova forma de publicar.

Uns primers afectats per aquest creixement exponencial d'informació poc estructurada, van ser els motors de cerca al web, que es trobaven cada cop amb més dificultats per realitzar una indexació efectiva que permetés la seva correcta explotació.

Mes actualment, s'han afegit uns segons afectats: els smartphones. Dispositius, que per les seves capacitats reduïdes en relació amb un ordinador tradicional (tant en processament, com memòria o ample de banda), es troben força dificultats en el tractament eficient i eficaç de la informació.

Aquest fets, suggereix la creixent necessitat d'un re-disseny de la informació existent a la WWW per fer-la més aprofitable, transformant-la en una mena de base de dades de les quals es pugui extreure la informació necessària en cada moment.

Resulta curiosa la silenciosa arrencada al 1997 del format d'intercanvi RSS, que a partir del concepte de “sindicar contingut” permet compartir aquest contingut entre diversos llocs web o dispositius, i que no serà fins mes actualment que guanyi rellevància.

En aquest escenari, l'aposta del web semàntic o la “web de dades” [BLL01], és la d'aconseguir que les màquines puguin comprendre el contingut de les pàgines web perquè puguin explotar eficientment la seva informació.

Reorganització de la informació

Aquesta necessitat de reorganització pot tenir un petit símil amb el naixement d'Internet, on es va partir d'ordinadors aïllats (elements), que es van interconnectar formant xarxes, i aquestes a la seva vegada un cop connectades van donar lloc a la xarxa d'Internet.

El web semàntic proposa partir d'elements d'informació dintre de les pàgines web, per connectar-los posteriorment a estructures més grans fins donar lloc a la xarxa semàntica.

Llavors el primer pas passaria per identificar aquest “objectes” elementals d'informació que ens serveixin per poder-los “reensamblar” posteriorment d'altres formes. Aquest pas d'abstracció ens permet separar-los del domini més operacional i portar-los a un domini més ampli de coneixement.

Un cop identificats, per poder ser expressats, cal triar un format de representació. Existeixen múltiples maneres d'organització i representació d'aquest coneixement, amb diversos graus de formalitat i nivells d'expressivitat.

Des del punt de vista semàntic alguns dels exemples més coneguts poden ser:

- Vocabularis:
 - Son un conjunt de mots recopilats amb algun criteri.
 - Habitualment els trobem representats per:
 - * diccionaris: llista de definicions mots.
 - * thesaurus: llista de sinònims, antònims, etc.
 - Un exemple pot ser: Wordnet
- Classificacions:
 - Es una distribució d'un conjunt d'objectes en un cert nombre de classes, coordinades o subordinades segons un criteri.
 - Habitualment els trobem representats amb:
 - * Taxonomies: estructures jeràrquiques, seguint relacions pare-fill (arbres genealògics) o tipus-subtipus.
- Xarxes de conceptes.
 - Ontologies: permeten representar dominis de coneixement.

Per la seva capacitat expressiva profunditzarem en aquest últim model.

2.1.1 Què és una ontologia?

El terme “ontologia” prové dels camps de la filosofia i la epistemologia: filosòficament, es considera una explicació sistemàtica de l'existència [Wik]; mentre que científicament, s'ocupa de la part de la metafísica que tracta de descriure la realitat del ser i les seves propietats transcendents [Her02].

Va ser introduït al camp de la informàtica per Tom Gruber a l'any 1993 per referir-se a l'especialització d'una conceptualització:

“In the context of knowledge sharing, the term ontology means a specification of a conceptualisation. That is, an ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents. This definition is consistent with the usage of ontology as set-of-concept-definitions, but more general. And it is certainly a different sense of the word than its use in philosophy.” [Gru93] [Gru09]

Des d'aquest punt de vista, una ontologia és una descripció formal i explícita d'un domini, definint:

- conceptes.
- relacions, propietats i atributs.
- restriccions.

Així doncs, les ontologies permeten establir un vocabulari i una representació de coneixement comuns, que ens facilitarà la seva reutilització tant per persones com màquines (o agents de programari).

D'aquesta forma s'aconsegueix una separació entre el domini del coneixement i el domini operatiu del coneixement, fet que permet analitzar-lo i fer assumpcions explícites sobre el model.

Per representar el significat dels conceptes, les ontologies contenen categories jeràrquiques amb relacions “es-un”.

Donat que en un determinat domini del coneixement hi han categories més genèriques que d'altres, porta a pensar que és possible l'existència d'algun tipus de classificació entre ontologies respecte la seva “generalitat”

Tot i que no existeix un consens sobre aquesta classificació atenent el seu àmbit de domini podem destacar-ne 3 tipus de més general a més específic:

Ontologies superiors (Upper ontology [Hoe10]) modelen conceptes genèrics que son aplicables dintre d'un ample rang de dominis. Alguns exemples poden ser: Cyc, UMBEL, Wordnet, entre d'altres.

Ontologies de domini modelen els conceptes d'un domini específic. Alguns exemples: Dublin core (documents i publicació), Friend of a Friend (per descriure persones), rosetta-net (e-business), UMLS (medical), EngMath (Enginyeria)

Ontologies d'aplicació modelen conceptes molt específics d'una aplicació concreta. Alguns exemples: EFO (representen dades experimentals en genètica), NIFSTD (per aplicacions de neuriciència). [JM10]

2.1.2 Aplicacions sobre el web semàntica

Aquesta nova forma d'estructurar i definir la informació, permet donar vida a un ecosistema d'aplicacions format per aplicacions noves o bé redefinició de models anteriors, que permet una explotació completa de les dades aportades.

En destaquem la iniciativa de dades vinculades: Linked Data.

Dades vinculades (Linked Data)

Partint de l'enfocament de BBDD distribuïda que aporta el web semàntic, “linked data” és un grup de treball de W3C amb l'objectiu d'aconseguir aquest ús, estenent el web amb dades comunes de Open Data sets.

Tim Berners-Lee va donar quatre principis que haurien de fer-la possible:

1. Fer servir URIs per identificar objectes.
2. Fer servir URIs HTTP per que aquests objectes puguin ser seguits o desreferenciats, tant per persones com agents de programari.
3. Donar informació útil sobre l'objecte al desreferenciar la URI, fent servir formats estàndards com RDF/XML.
4. Incloure enllaços relacionats en les dades exposades per millorar el descobriment d'altra informació relacionada amb el web.

Al 2009 va fer servir servir tres regles mes simples:

1. Tots els tipus d'objectes conceptuals, han de tenir noms que comencin per HTTP
2. S'ha de retornar informació important en un format estàndard que sigui útil sobre l'objecte.

La unió europea ha impulsat a traves de diversos projectes:

- Linking Open Data Around the clock (LATC), 2007-2013, amb l'objectiu d'impulsar la publicació i el consum de dades obertes, a través de l'entitat LATC.
- PlanetData.
- Linking Open Data 2, 2010-2014, amb l'objectiu de “crear coneixement fora de Inteerlinked Data”

A data de setembre de 2010 es van créixer fins a 25 bilions de tripletes RDF i 395 milions d'enllaços RDF. L'èxit de la iniciativa es pot veure amb la gràfica de fonts obertes aconseguida:

2.2 Tecnologies del web semàntic

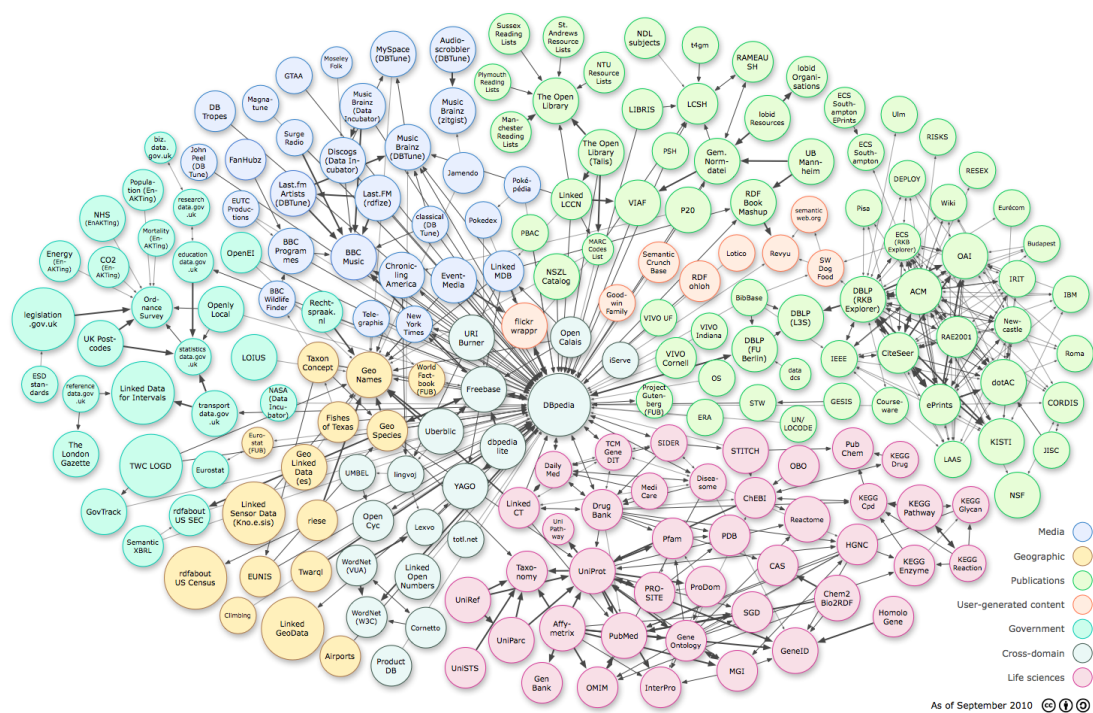


Figura 2.2: Linked Data

2.2 Tecnologies del web semàntic

Per fer possible aquesta nova forma d'aportar contingut, cal una estructura tecnològica que la pugui sustentar. A l'any 2000, Tim Berners-lee va donar la seva visió sobre com podria ser [BL00].

La seva proposta inicial és una pila formada de diverses capes, de les quals ens centrem a les següents :

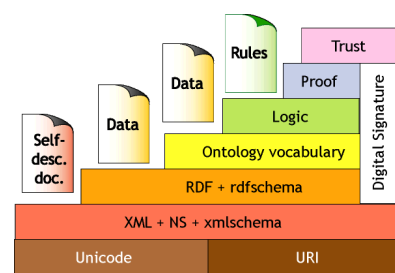


Figura 2.3: Pila semàntica presentada per Tim Berners-Lee

XML+NS+xmlschema amb Unicode+URI Que constitueixen una capa destinada a garantir la interoperabilitat sintàctica del document. Per poder definir una estructura de document flexible i restringida que permeti extreure el contingut.

RDF+RDF Schema Que formen una capa per mantenir l'interoperabilitat semàntica. Una primera capa minimalista per aportar significat sobre les propietats i definir taxonomies de classes.

Ontology Vocabulary (OWL) Afegeix una capa mes rica semànticament per permetre aplicacions mes complexes (raonadors) a capes superiors.

Amb el temps no s'ha aconseguit establir una única visió de la pila semàntica, i actualment existeix una visió mes complexa de la pila com posa de manifest la següent figura.

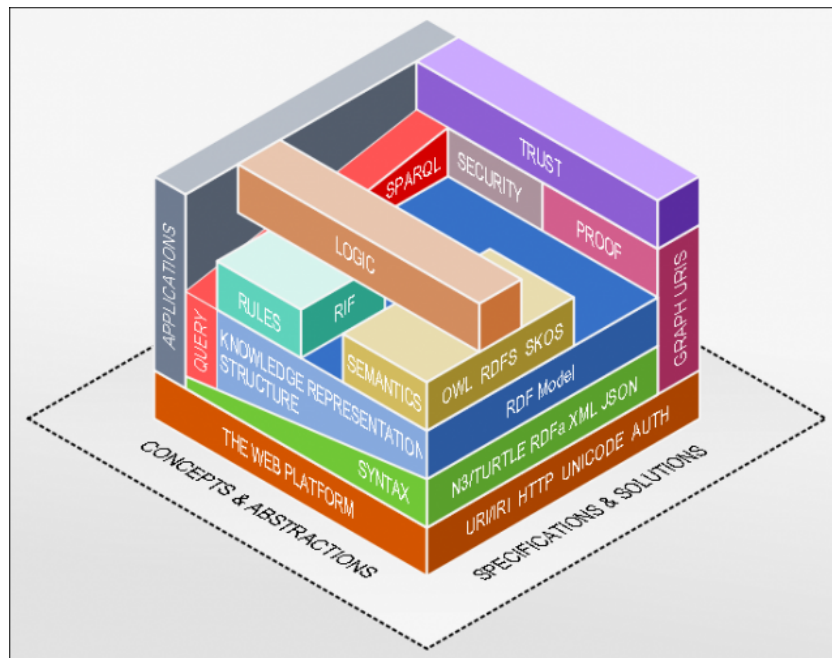


Figura 2.4: Visualització d'una pila de tecnologies semàntiques

2.2.1 Llenguatges de descripció

2.2.1.1 La descripció sintàctica: XML

XML (Llenguatge extensible de marcat) és un llenguatge de marcat desenvolupat pel W3C per superar les limitacions a les que es va arribar amb HTML.

Mentre HTML es focalitza en l'aparença i presentació del document, XML es centra en descriure la seva estructura. Per tant XML no pretén substituir HTML, el complementa: un es centra en la capa de presentació, mentre XML forma part de la capa de dades.

XML s'elabora amb dos canvis de filosofia essencials respecte HTML: no predefineix etiquetes, i és estricte.

En no predefinir etiquetes, ens permet fer servir les etiquetes més apropiades per descriure l'estructura del document. Això transforma XML en un metallenguatge: un llenguatge per definir llenguatges.

En ser estricte, vol dir que la seva estructura ha de ser correcta. Per aquest efecte es defineixen les DTD.

La DTD permet efectuar un anàlisi sintàctic del document i determinar:

Documents invàlids aquells que incompleixen les regles de sintaxi especificades al DTD

Documents ben formats són els que segueixen les regles de sintaxi XML però no corresponen a cap DTD o esquema determinat.

Documents vàlids compleixen la sintaxi de XML i les regles definides al DTD.

Aquest nou enfocament, fa que ens trobem amb els següents avantatges respecte al HTML:

- Permet un codi més lleuger i senzill de processar, identificant fàcilment les parts de la informació que necessitem en cada cas. Aquesta estructuració ens permet explotar les dades en certa manera com si es tractés d'una taula d'una BBDD, facilitant les cerques, consultes.
- Un segon efecte d'aquesta simplificació, és que facilita l'intercanvi de dades. Permetent ajustar diferents representacions gairebé de forma automàtica (veure transformacions XSLT)

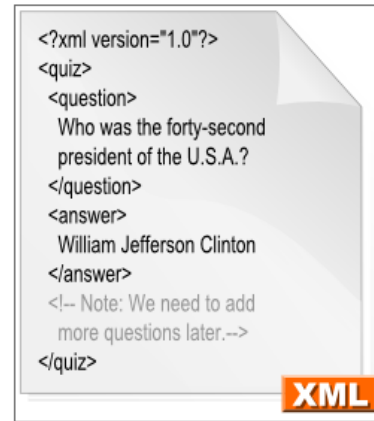


Figura 2.5: Descripció gràfica d'un document XML

Un cop publicat XML 1.0 es van desplegar tota una serie d'especificacions al seu voltant entre les que podem destacar:

- Presentació
 - CSS, Cascade Style Sheets, permet detallar l'aparença del document.
 - XSL, eXtensible Style Language, permet efectuar transformacions sobre el document orientades a la visualització del document.
- Estructura
 - XML Namespaces, que permet evitar les col·lisions que es poden produir quan es fan servir diversos vocabularis al mateix document.
 - XML Schemas 1 i 2, que aporten restriccions sobre el document XML. Prové del món de les BBDD relacionals, on un esquema defineix un model al que les dades s'han d'ajustar.
- Consulta i actualització
 - XPath, que permet identificar diversos components del document (elements, atributs).
 - XLink, que descriu una manera estàndard d'afegir hiper-enllaços dintre d'un document XML.
 - Xpointer, s'utilitza per identificar l'estructura interna d'un document XML seleccionant parts específiques del document en funció de la informació de l'arbre del document.
 - XQuery, dota de mètodes d'accés, manipulació i recuperació per XML.
 - XUpdate, és una millora sobre XPath i XQuery que permet fer actualitzacions.
 - DOM, exposa l'estructura de dades en forma d'arbre de XML a llenguatges de programació.

XML estableix és un format per representar la informació en forma d'arbre.

XML permet mantenir l'interoperabilitat sintàctica.

Format d'un document XML

Els documents XML es formen amb tres components: etiquetes, elements i atributs:

Una etiqueta és un text delimitat entre els símbols “<” i “>”, distingim dos tipus: d’inici (<nom>) i de final (</nom>).

Un element és el contingut delimitat entre etiquetes d’inici i final. .

Un atribut és una tupla nom-valor dins d’una etiqueta d’inici d’un element. .

Regles bàsiques:

- només pot existir un element arrel
- els elements no es poden encavalcar
- les etiquetes de final son obligatòries
- els elements son sensibles a majúscules i minúscules
- els atributs han de tenir valor i estar entre cometes
- comentaris els de html <!-- -->
- instruccions de processament <?parser ?>
- entitats html pels caràcters reservats.
- UTF8

2.2.1.2 La descripció semàntica: RDF

Un cop tenim una sintaxi estricta sobre la que podem extreure i relacionar informació, podem afegir contingut semàntic als seus elements, aquest és l’objectiu de RDF.

RDF no és un llenguatge pròpiament, és un framework que defineix unes regles a complir pels llenguatges que vulguin afegir aquest contingut semàntic. Algunes implementacions son RDFa...

RDF representa grafs i permet mantenir l’interoperabilitat semàntica.

Limitacions de RDFS

- És dèbil per descriure recursos amb un suficient nivell de detall.
- El rang i el domini no pot especificar-se. Per exemple, el rang de teFill es una persona quan s’aplica a una persona i un elefant quan s’aplica a un elefant.
- No té restriccions d’existència o cardinalitat. Per exemple, totes les instàncies de persones tenen una i només una mare que es una persona, i tenen exactament dos pares.



2.2.1.3 La descripció ontològica: OWL

El Web Ontology Language (OWL) va ser desenvolupat per donar més expressivitat i poder representar ontologies amb una lògica formal decidible.

Dona construccions addicionals que no es troben presents en RDFS per definir classes i propietats, com a resultat, owl està millor preparat per comprovacions de consistència i tasques de classificació.

Ens trobem tres variants:

- OWL Full: dona una expressivitat completa, però sense garanties de decidibilitat.
- OWL DL: es restringeix l'expressivitat de OWL Full a canvi d'obtenir decidibilitat, formant un llenguatge de lògica descriptiva (DL) capaç de suportar aplicacions de raonament.
- OWL Lite: és un subconjunt de OWL DL orientat en un inici a la creació de tesaurus i vocabularis sense cardinalitat múltiple.

OWL afegeix construccions addicionals no presents en RDFS per definir classes i propietats. Com a resultat, owl està més preparat per comprovacions de consistència i tasques de classificació.



3 Elaboració de la ontologia ontoTrauma

L'ontologia serà la que determinarà l'estructura de les dades que es guardaran a la base de coneixement que posteriorment ha de servir per ser explotada per eTrauma.

Abordarem la seva creació en tres etapes: primer estudiarem l'estat actual de les ontologies relacionades amb les àrees de la salut, després presentarem la metodologia que farem servir i finalment en farem ús per aconseguir una ontologia pròpia.

3.1 Anàlisi d'ontologies quirúrgiques traumatològiques

S'ha fet una cerca d'ontologies prèvies amb la intenció d'estudiar la seva reutilització, o si mes no, aproximar-les per poder facilitar l'aprofitament de les dades que recullen.

Ha estat difícil trobar material actualitzat i mantingut, tot i així, cal destacar el catàleg d'ontologies obertes biològiques i biomèdiques de Berkeley [Bio11] i el catàleg de conjunt de dades vinculades [linked data sets].

Ambdós catàlegs disposen d'una gran varietat d'ontologies tant d'alt nivell com de dominis més concrets de la biologia, però si bé moltes poden servir com extensió a la base de dades de coneixement que estem tractant, no detectem cap que cobreixi el domini que estem treballant.

Una altra font d'informació que s'ha estudiat és la generada per conferències. En aquest context, la documentació més completa ha estat la de les ponències del 2009 del Innovation Center Computer Assisted Surgery [Bur09].

Aquesta documentació mostra un recull d'ontologies orientades a la realització d'un CAS (Computer Aided Surgery): SOCAS, SWOnt, PatientOnt, FESSOnt,

3.2 Metodologia per a la creació d'una ontologia

MLSOnt; malauradament no es disposa de les fonts d'aquestes ontologies per fer-ne ús, i probablement seran comercials.

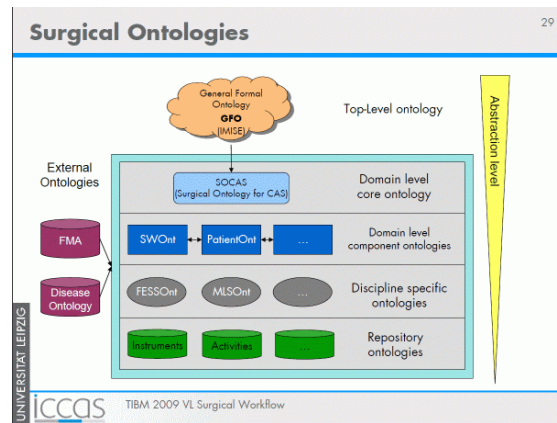


Figura 3.1: Relació d'ontologies proposades a [Bur09] segons el seu nivell d'abstracció

També he cercat entre diversa documentació acadèmica, com a "Ontological modeling knowledge" [RM11].

Un cas d'ús d'ontologies que he trobat interessant per tenir en compte és el de la National Highway Traffic Safety Administration, que fent ús de la tecnologia de dades vinculades, disposa d'un sistema de The Crash Outcome Data Evaluation System (CODES) And Applications to Improve Traffic Safety Decision-Making - [NHT09]

Donades aquestes circumstàncies i la natura d'aquest projecte s'ha optat pel desenvolupament d'una ontologia pròpia.

Tot i això, es tindran en compte fer servir les ontologies necessàries per poder facilitar la federació futura de la nostra ontologia amb la DBPedia, o les de salut americanes de data.gov.

3.2 Metodologia per a la creació d'una ontologia

A data d'aquest projecte, no existeix una metodologia única per elaboració d'ontologies i cada una presenta avantatges i inconvenients depenent del domini al que es vulgui aplicar l'ontologia. Algunes destacades podrien ser:

- Competency Questions, o la metodologia de "les preguntes rellevants". Consisteix en efectuar una sèrie de preguntes per determinar el domini i l'abast

de la ontologia. Es a dir, per una banda, les respostes ens permetran determinar les classes i instàncies que tindrà la ontologia, a partir de les quals podem anar generalitzant. I per l'altra banda, les preguntes ens permetran avaluar si hem representat suficient informació per ser respostes.

- Methontology. Va ser desenvolupada per la Universitat Politècnica de Madrid i proposa un model basat en un cicle de vida iteratiu per prototipatge. En cada iteració, s'elabora un refinament de la ontologia basant-se en les següents etapes: Especificació, conceptualització, formalització, implementació i manteniment.
- NeOn. Sorgeix d'un projecte de la Unió Europea. Presenta nou escenaris que intenten cobrir totes les possibles necessitats on cal fer servir una ontologia, fent molt d'èmfasi a la reutilització d'ontologies existents per sobre de la creació d'una nova. Com methontology presenta també una serie d'etapes per la confecció de l'ontologia dintre d'aquests escenaris.

Aquestes dues últimes (Methontology i NeOn) son possiblement les mes completes i millor documentades, però també son massa extenses per l'abast d'aquest projecte, per aquest motiu plantegem una simplificació que doni mes èmfasi en la comprensió de la ontologia que en el seu grau de formalització.

Aquesta simplificació és una adaptació de la que planteja Rubén Darío al seu article titulat "Metodología para el desarrollo de ontologías" [Alv10], basant-se en Methontology i Noy & McGuinness.

En línies generals la seva proposta es basa en un procés iteratiu amb les següents etapes:

1. Determinar els requeriments de l'ontologia: definint el domini, l'abast, la granularitat i el manteniment de l'ontologia.
2. Elaboració del model conceptual:
 - a) Definició de termes de la ontologia.
En base a la documentació i les preguntes rellevants del punt 1, efectuarem un glossari de termes del domini d'aplicació de la ontologia.
 - b) Definició de les classes i de la jerarquia.
A partir del glossari anterior, seleccionem els conceptes que descriuen objectes independents i efectuem una jerarquia de classes. Alguns dels criteris que tindrem en compte a l'hora de afegir una classe nova son:
 - Ha d'incorporar noves propietats, nous valors d'aquestes o participar en relacions diferents.
 - Els sinònims d'un mateix concepte no generaran classes diferents.

3.2 Metodologia per a la creació d'una ontologia

- En cas que els conceptes formin una jerarquia natural, es representaran com classes encara que siguin abstractes.
- Les classes germanes han de tenir el mateix nivell de generalitat.

c) Definició de les propietats de les classes.

En aquesta etapa, definirem els atributs i les relacions entre les classes a partir dels conceptes que hagin quedat pendents d'analitzar. També afegirem noves classes si son requerides per determinar millor aquests atributs.

Descriurem també algunes restriccions sobre les propietats com son: els dominis, rang, cardinalitat.

d) Creació de consultes sobre el model.

3. Implementació del model conceptual OWL amb protégé.

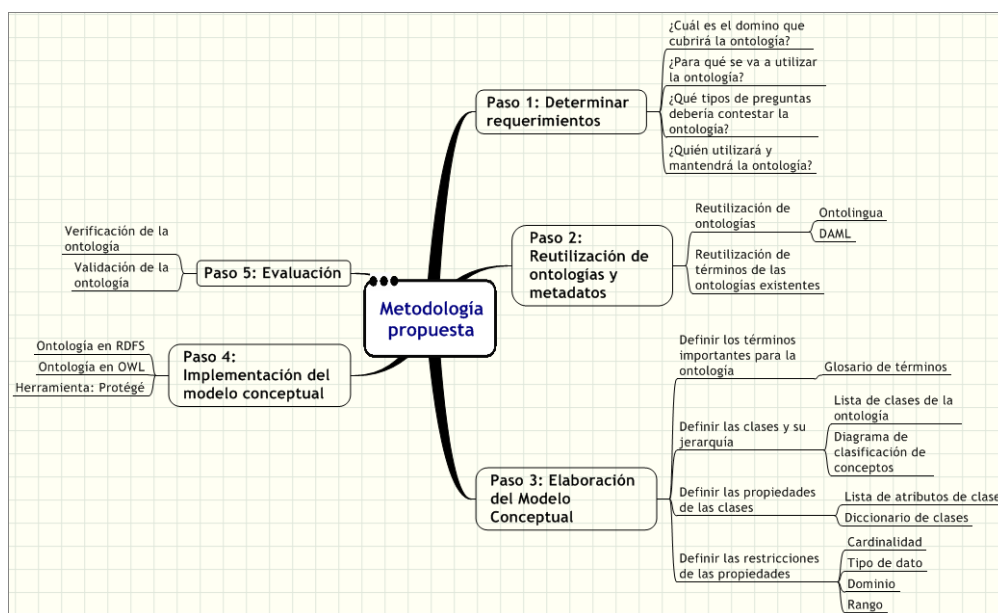


Figura 3.2: Descripció del mètode de desenvolupament plantejat per [Alv10]

3.3 Creació de l'ontologia ontoTrauma

3.3.1 Determinació de requeriments

Quin domini cobrirà l'ontologia? ontoTrauma ha de descriure una intervenció quirúrgica de l'especialitat de traumatologia.

Perquè es farà servir l'ontologia? ha de permetre efectuar el seguiment de la informació rellevant d'un pacient (història clínica) necessària pel preoperatori, així com donar assistència durant el procediment intraoperatori, ja sigui amb els protocols dels diversos de l'operatori amb la informació de material i procediments preoperatori. Mes concretament:

- es mantindrà en tot moment l'historial clínic a disposició
- en cas d'intervenció quirúrgica mostrarà la seqüència de procediment que ha de seguir cada membre de l'equip i el material relacionat en cada pas.

Per descriure millor el seu ús mostrarem el procediment sencer que haurà de cobrir l'ontologia a través del següent diagrama d'activitat:

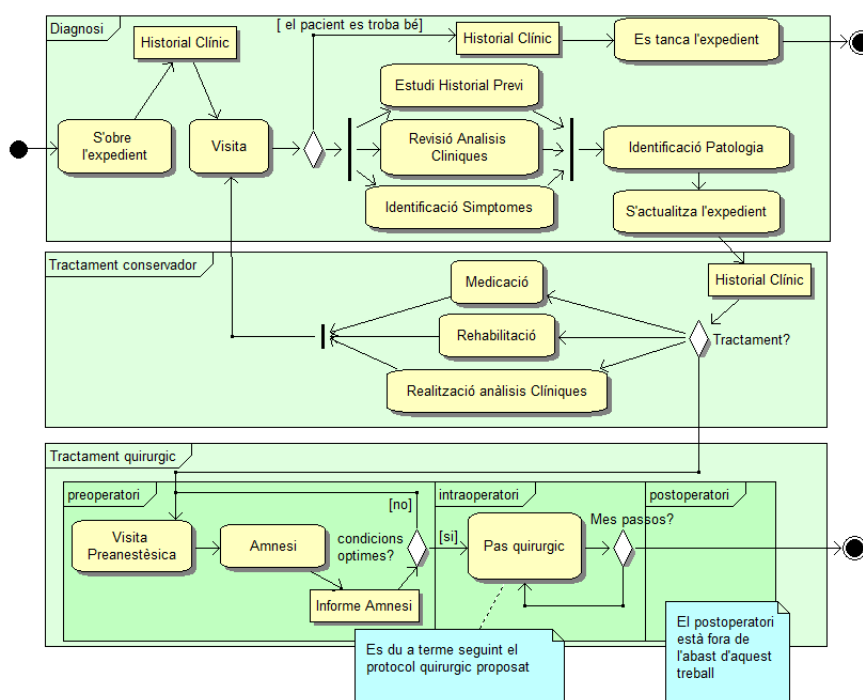


Figura 3.3: Diagrama d'activitat d'una operació quirúrgica traumatològica

Quines preguntes hauria de contestar l'ontologia? Recollim les següents preguntes:

- En quin estat es troba el pacient? Està en tractament conservador? Cal operar?
- L'anestesiòleg ha autoritzat el procediment quirúrgic?
- Quin és el protocol previst per la intervenció quirúrgica? Quines passes ha de seguir el cirurgià? i l'anestesiòleg? i el traumatòleg? i els infermers?
- Quin material cal en cada pas del protocol?
- En finalitzar s'ha seguit el protocol previst? quins son els passos que s'han modificat?

Mes concretament haurem de poder respondre les preguntes involucrades dintre dels casos d'ús en l'etapa de diagnosi i tractament conservador:

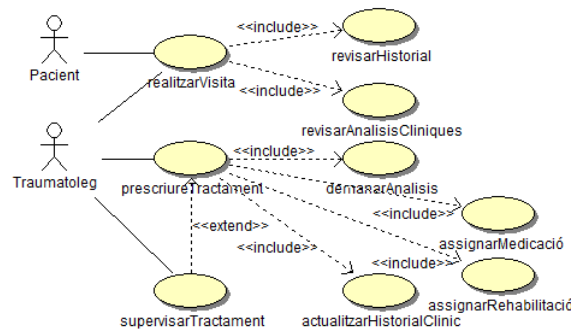


Figura 3.4: Diagrama de casos d'ús involucrats en l'etapa de diagnosi

I farem el mateix dintre pel tractament quirúrgic:

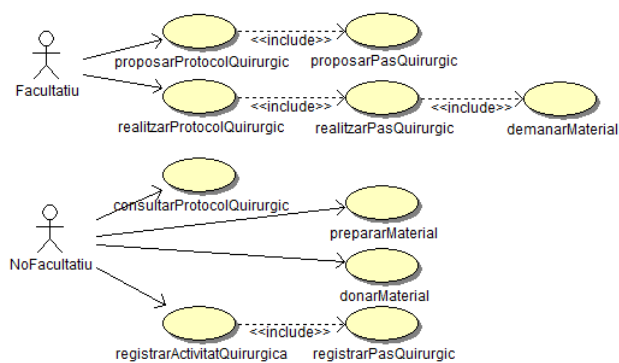


Figura 3.5: Diagrama dels casos d'ús involucrats dintre d'un tractament quirúrgic

Qui farà servir l'ontologia? l'usuari és el personal facultatiu i no facultatiu dintre del quiròfan. Tot i que, en les fases de recopilació de l'historial del pacient pot col·laborar altre tipus de personal administratiu que efectui el seguiment del pacient.

Mes concretament ens trobem amb el següent personal involucrat:

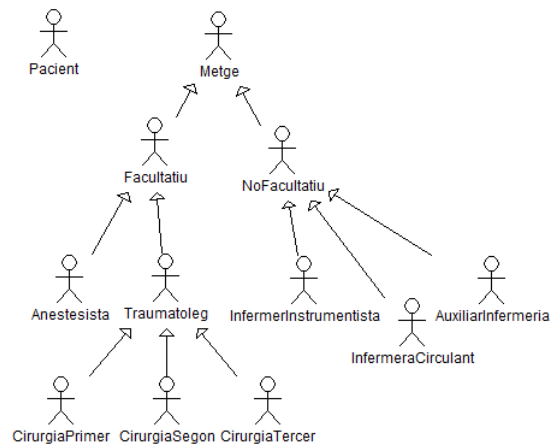


Figura 3.6: Diagrama de la classe Persona

3.3.2 Elaboració del model conceptual

3.3.2.1 Definició de termes de la ontologia

Amb la documentació recollida, i l'estudi de les dades existents, efectuem un glosari dels termes que farem servir. Tot i que no seria necessari, per facilitat, els recullo en una jerarquia atenent a la seva definició:

Diagnosi És la fase a la qual es recopila informació del pacient, per la seva patologia, i el tractament per guarir-lo.

Historial Documentació històrica del pacient, que es complementa a través de les entrevistes

Anàlisi clínica Cadascuna de les proves que s'efectuen al pacient

Síntoma Cadascun dels indicis que mostra el pacient i que poden constituir una patologia

Patologia Malaltia que pateix el pacient

Tractament És la fase a la qual s'intenta guarir el pacient del seu mal

Conservador Conjunt de tractaments alternatius a la cirurgia

Medicació Productes químics destinats a guarir al pacient

Oral/Endovenosa/Infiltració Vies de subministrament de la Medicació

Rehabilitació Conjunt de procediments físics que permeti restablir al pacient.

Quirúrgic Tractament per guarir el pacient a través de cirurgia

Procediment

FaseQuirúrgica Cadascuna de les fases de les que es compon la cirurgia requerida

Preoperatori Fase de recopilació d'informació i aprovació del protocol per l'elaboració de la cirurgia

Intraoperatori Fase d'execució del protocol quirúrgic.

postoperatori Fase posterior a la cirurgia.

ProtocolQuirúrgic Cadascun dels passos necessaris per desenvolupar una operació quirúrgica.

PasQuirúrgic Pas dintre d'un protocol quirúrgic

Material Material requerit dintre d'un pas quirúrgic.

DobleBosa/DobleCapa/CaixaTancada presentació del material.

Fungible/NoFungible recull si el material es desgasta o no.

Esteril/NoEsteril estat del material respecte l'absència de contaminants.

MaterialEspecífic material especialitzat per efectuar una Cirurgia

Persona

Pacient Persona afectada per una patologia

Metge Professional de la salut encarregat de vetllar pel pacient

Facultatiu/NoFacultatiu Defineix tipus de personal facultat per prescriure o realitzar tractaments al pacient

3.3.2.2 Definició de les classes i de la jerarquia

Agafant el glossari de termes anterior, efectuarem una primera jerarquia de classes modelant els conceptes principals.

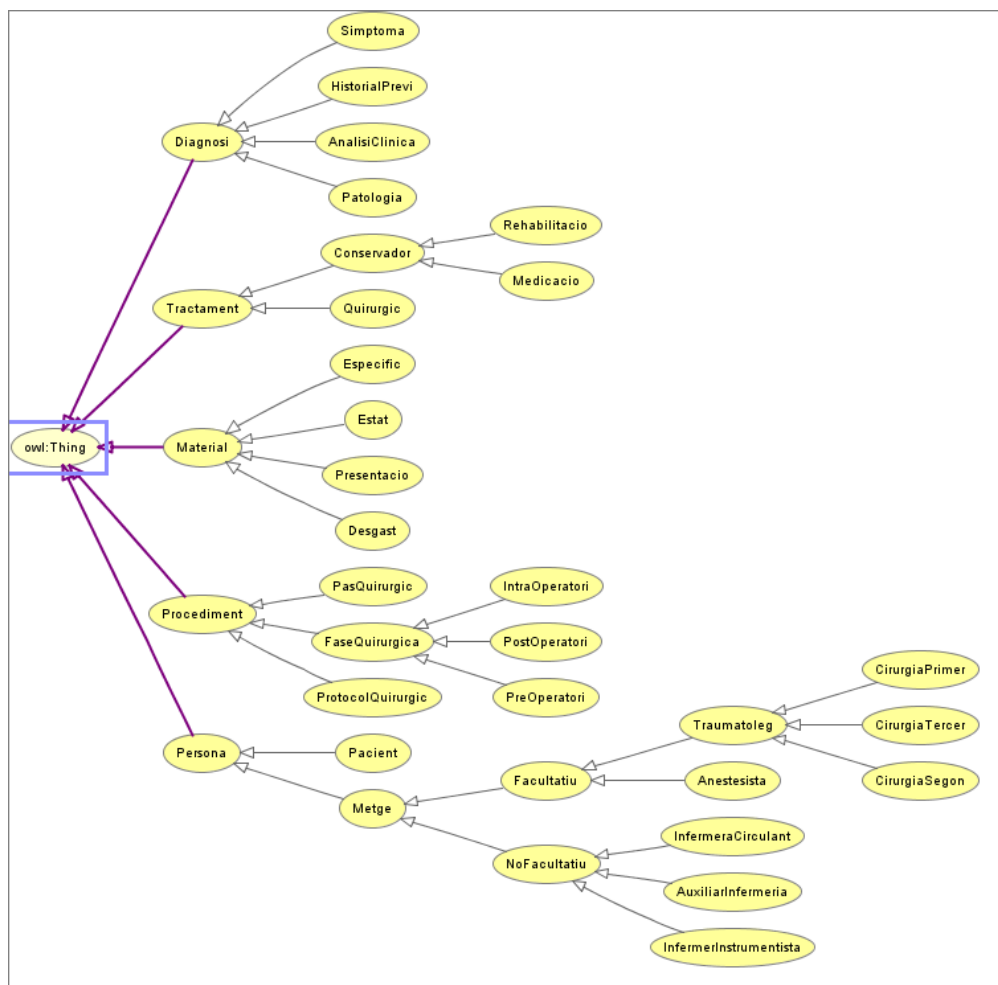


Figura 3.7: Diagrama de classes d'ontoTrauma

3.3.2.3 Definició de les propietats de les classes

Primerament determinarem les propietats d'objecte, es a dir, les relacions entre objectes. Les cercarem de forma sistemàtica centrant-nos a una classe i tractant de respondre les preguntes clau a les que està involucrada:

- teSituacioDescritaPer: Representa la documentació i els anàlisis que clíniques que descriuen al Pacient
- encarregatDe: Representa l'infermer encarregat d'un material en un determinat estat (Esteril/NoEsteril)Inversa: responsabilitatDe
- faServir: Representa el material que es fa servir a un determinat pas quirúrgic. Inversa:esFetServirPer

3.3 Creació de l'ontologia ontoTrauma

- **propostaProtocol**: Representa la proposta de protocol que es seguirà a la intervenció quirúrgica. Inversa: **esProposat**
- **realitza**: Representa el metge encarregat de realitzar un pasQuirúrgic determinat. Inversa: **esRealitzatPer**
- **supervisa**: Representa el tractament que supervisa un determinat Traumatòleg. Inversa: **esSupervisatPer**
- **teSimptomes**: Representa el conjunt de símptomes d'una Patologia. Inversa: **formenPatologia**
- **tePasosQuirurgics**: Representa els passos quirúrgics que s'han d'efectuar a una intervenció quirúrgica. Inversa: **formenProtocolQuirurgic**
- **teFasesQuirurgiques**: Representa les fases de les que consta un determinat tractament quirúrgic. Inversa: **formenTractament**

Les relacions quedaran doncs de la següent manera:

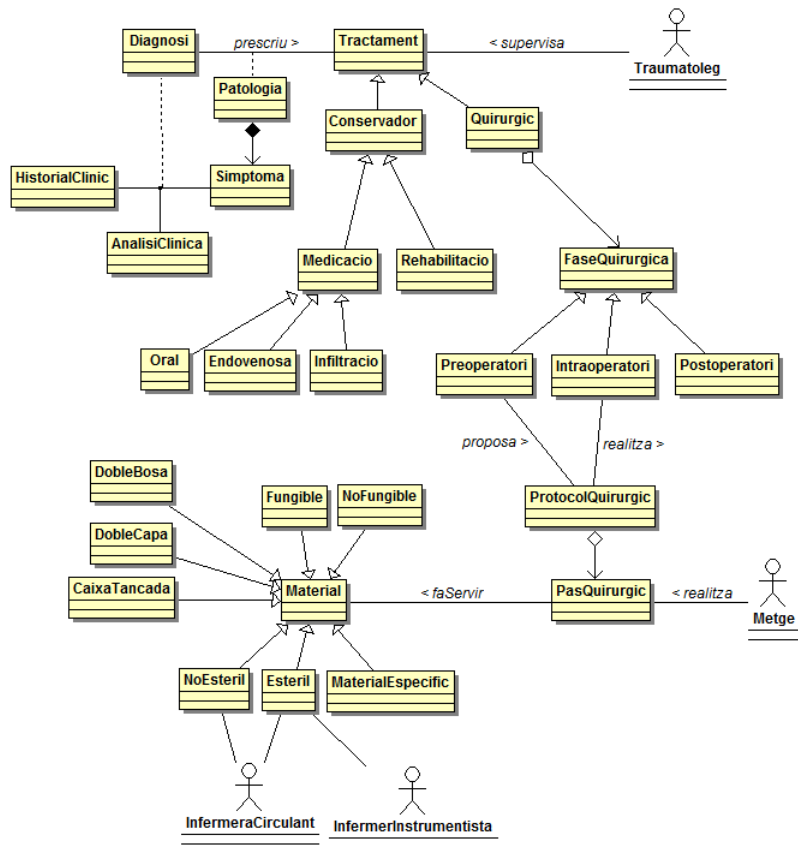


Figura 3.8: Diagrama de relacions de ontoTrauma

3.3 Creació de l'ontologia ontoTrauma

Per finalitzar definirem les propietats de les dades, o atributs de les classes pròpiament dites:

- nom, descripció, son transversals a tota la ontologia, permetent afegir aquests atributs a tots els elements.
- idProtocol, és l'identificador d'un protocol quirúrgic.
- idPas, és l'identificador d'un pas quirúrgic.
- dc:identificer i dc:title, ens permet referenciar un document i el seu títol. Trobarem documents tant als resultats de Diagnosi, Materials i Medicacions

3.3.2.4 Esquemes d'objectes i propietats

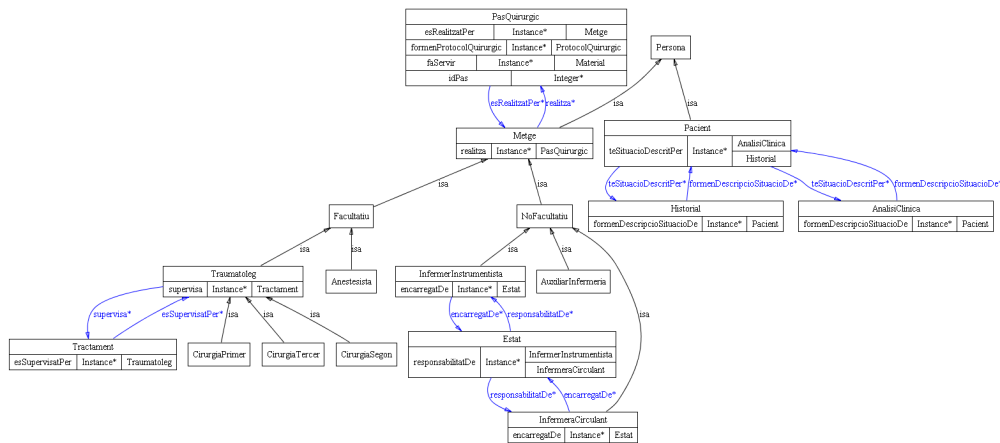


Figura 3.9: Classes i propietats de Persona

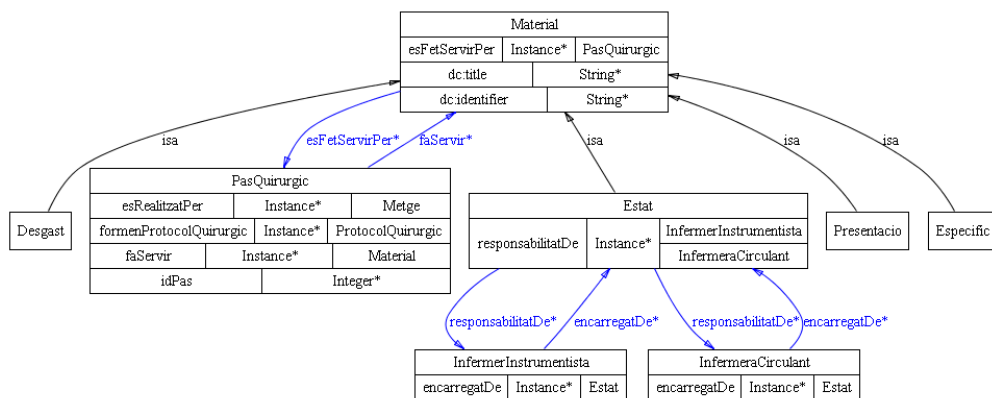


Figura 3.10: Classes i propietats de Material

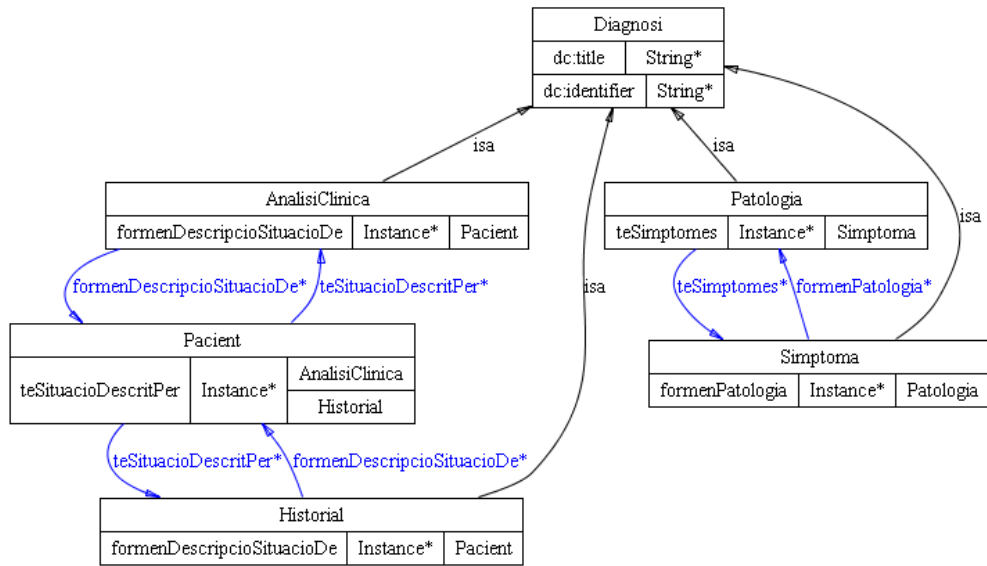


Figura 3.11: Classes i propietats de Diagnosi

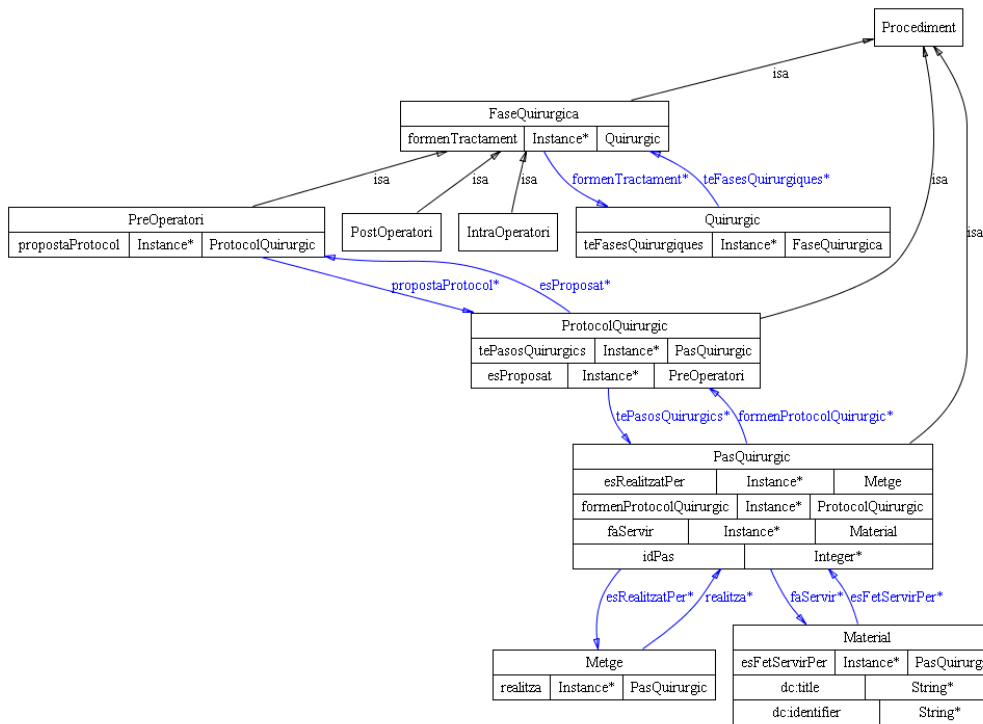


Figura 3.12: Classes i propietats de Procediment

3.3 Creació de l'ontologia ontoTrauma

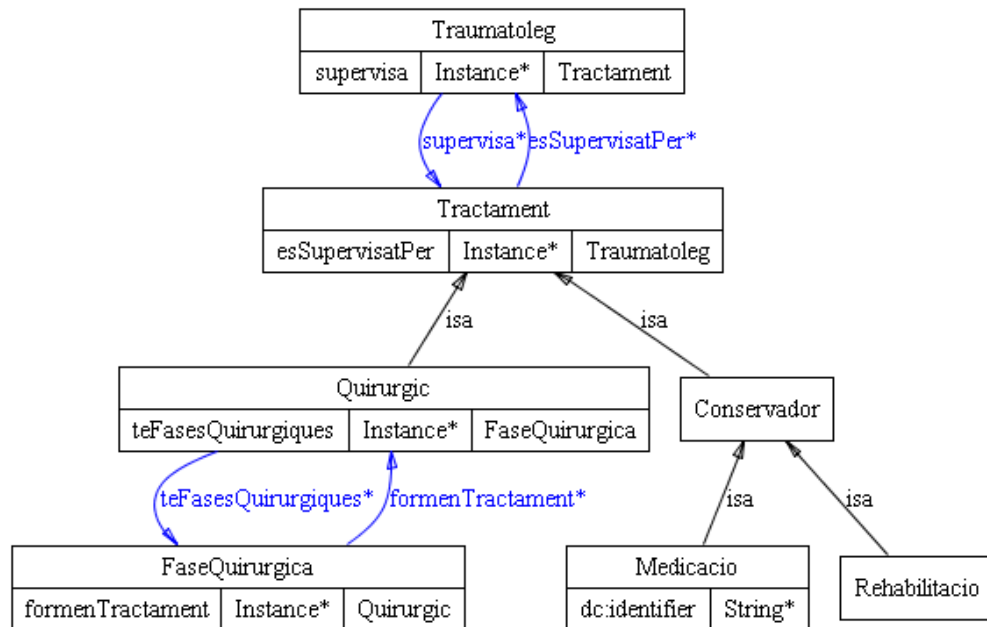


Figura 3.13: Classes i propietats de Tractament

3.3.3 Implementació del model conceptual amb Protégé

Farem servir el tutorial de protégé per crear la ontologia i especificar la metadada que farem servir, en el nostre cas farem servir atributs de dublin-core:

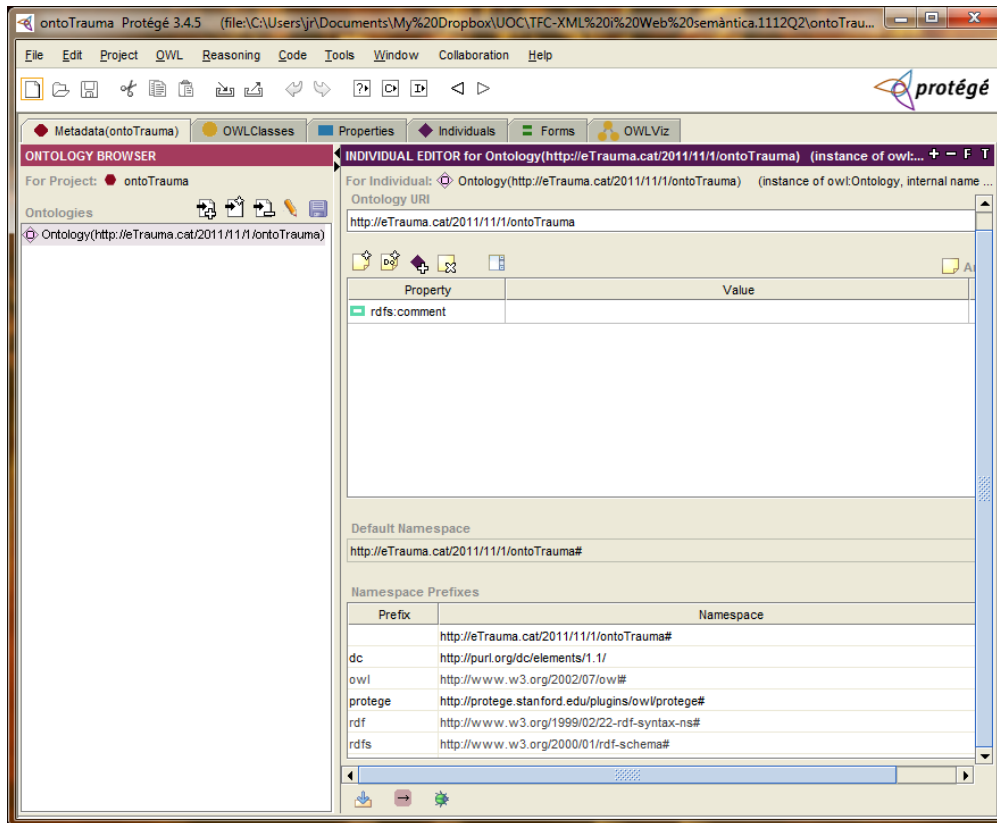


Figura 3.14: Definició de metadada Protégé

3.3 Creació de l'ontologia ontoTrauma

Seguidament, crearem el diagrama de classes a través de la pestanya OWLClasses, seguint la jerarquia proposada i tenint especial cura de determinar les classes que son disjunctes (no permeten elements de mes d'un subtipus a la vegada):

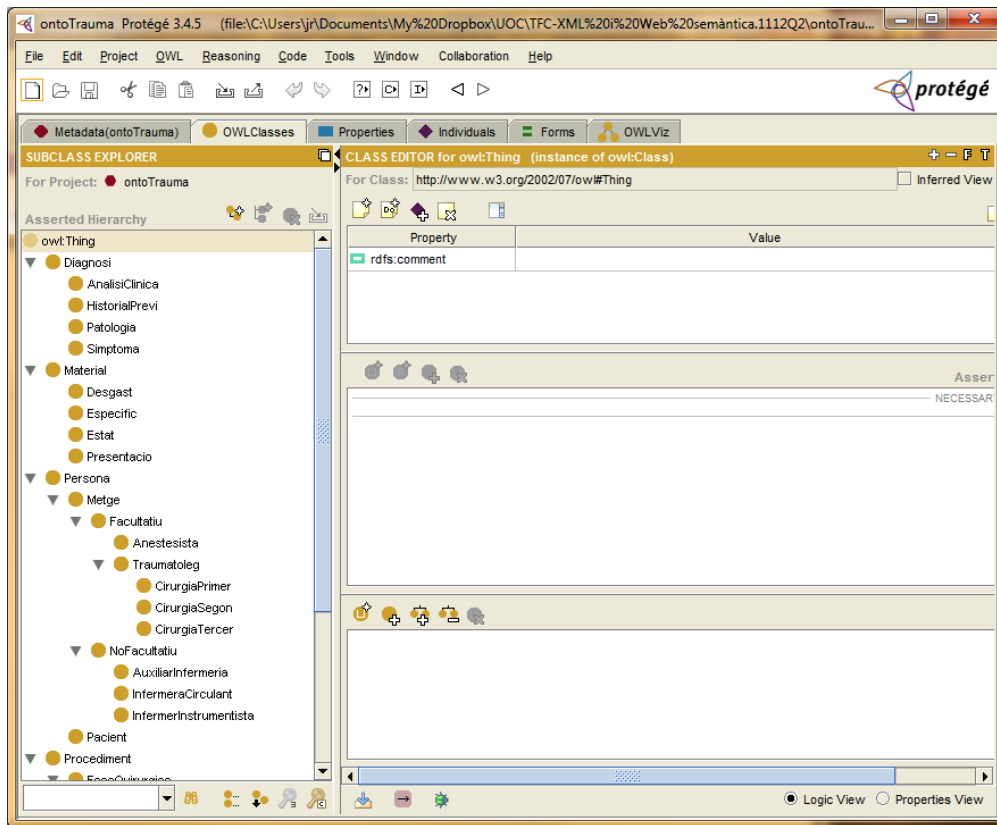


Figura 3.15: Definició de classes Protégé

3.3 Creació de l'ontologia ontoTrauma

Continuarem per la definició de les relacions entre les classes. En aquesta etapa codificarem també, la jerarquia de funcions, els seus rangs i dominis i les propietats (transitives, funcionals o simètriques).

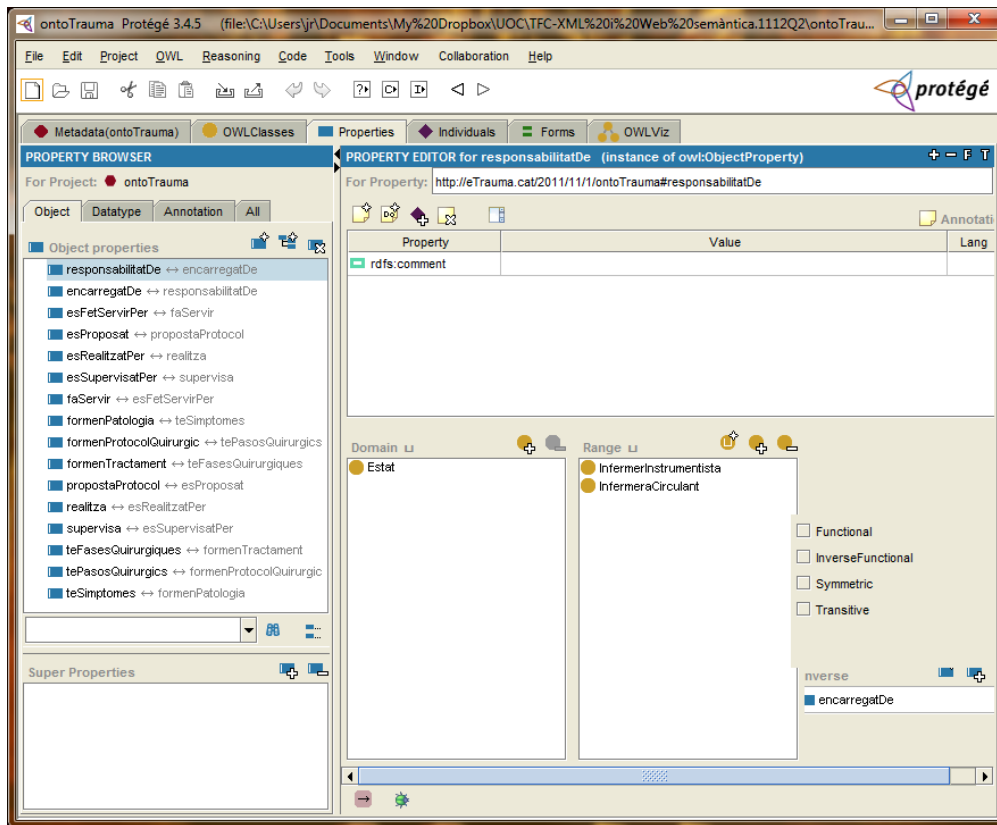


Figura 3.16: Definició de relacions Protégé

3.3 Creació de l'ontologia ontoTrauma

Ara caldrà determinar els atributs de les classes tenint cura dels Dominis als que se'ls hi aplica i els seus rangs.

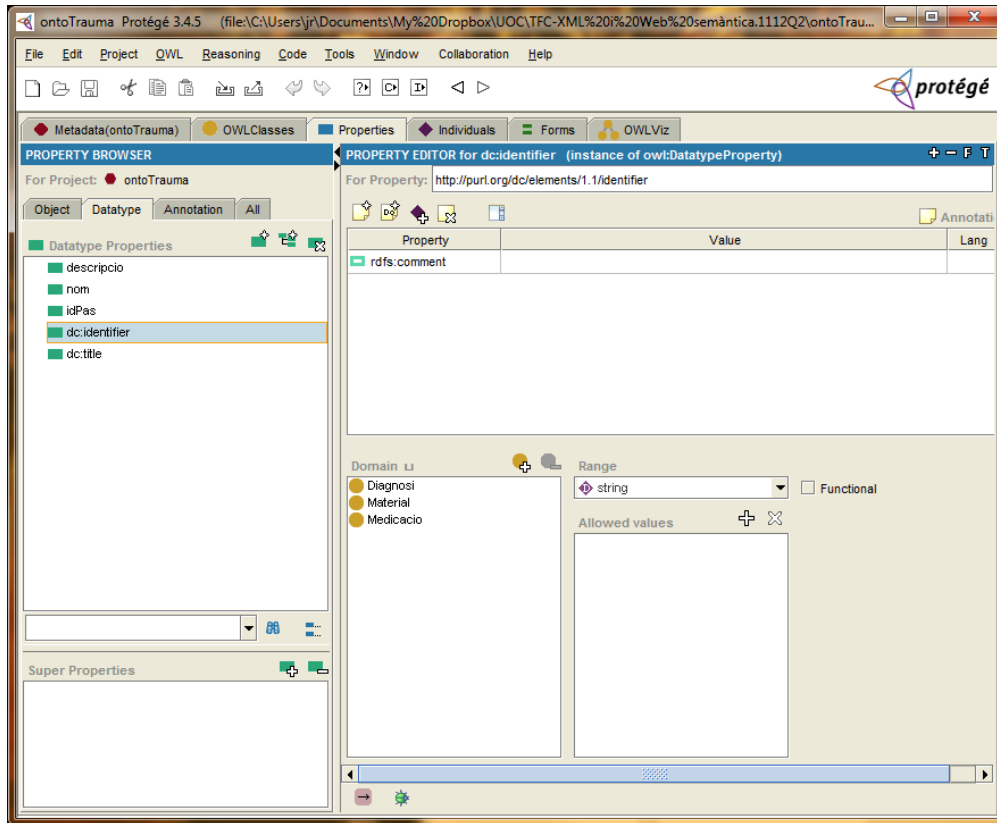


Figura 3.17: Definició d'atributs Protégé

3.3 Creació de l'ontologia ontoTrauma

Finalment crearem les instàncies a partir de les plantilles que ens aporten totes les definicions anteriors.

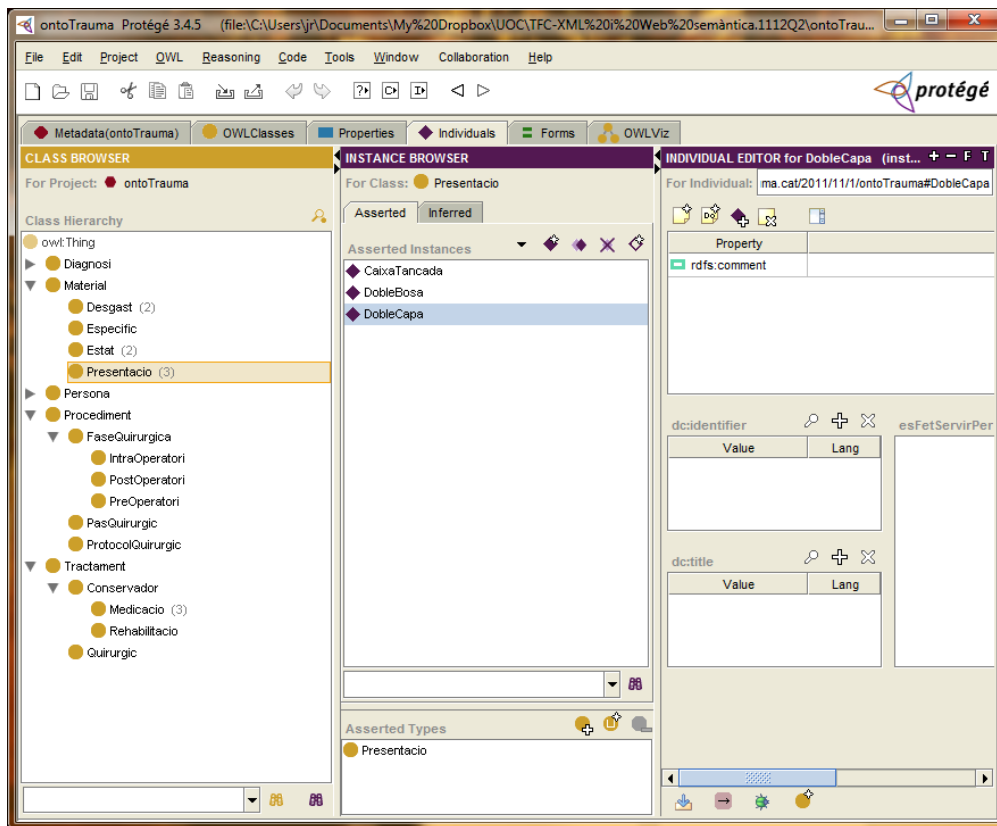


Figura 3.18: Definició d'instàncies Protégé

4 Elaboració de la base de dades de coneixement: els wikis semàntics

Les ontologies ens permeten modelar i donar estructura al domini de coneixement, per poder-ne fer ús emmagatzemant i recuperant informació necessitem generar una base de dades de coneixement.

Els últims anys, els wikis han guanyat rellevància com eines flexibles de generació de coneixement de forma col·laborativa, sobretot a partir del cas d'èxit que representa el projecte Wikipedia.

Els wikis son eines orientades a generar contingut ràpidament, estructurades en articles i una sintaxi lleugera anomenada wikitext, una simplificació del HTML. L'estratègia dels wikis facilitant la introducció i modificació de contingut, fa que els col·laboradors siguin a la vegada generadors, consumidors i revisors de la correcció de les dades que s'hi emmagatzemen.

Això relaxa l'estratègia més tradicionals d'altres programaris basats en formularis estrictes que imposen una estructura i correcció a priori sobre les dades, i és clau per aconseguir el dinamisme de les dades que proposen amb un bon compromís entre volum i qualitat.

L'espectacular creixement de la wikipedia, va propiciar que el sorgiment de projectes com el de DBPedia, que cerca explotar les seves dades de forma estructurada. I aquesta estructuració en el marc de la xarxa semàntica, possibilita el seu enllaç en estructures més grans com Linked Data.

En aquest context, han sorgit diverses propostes per aplicar tecnologies semàntiques als wiki. Estudiarem tres:

- Semantic Media Wiki. Una millora sobre MediaWiki, base de la wikipedia, per afegir contingut semàntic.
- Semantic Media Wiki Plus. Una solució comercial sobre Semantic Media Wiki.
- Ontowiki. Una recent iniciativa en desenvolupament, impulsada per la Unió Europea.

4.1 Sematic Media Wiki

Sematic Media Wiki sorgeix com una extensió de MediaWiki, que permet als usuaris afegir anotacions semàntiques al contingut que generen.

MediaWiki és la plataforma de programari lliure sobre la que es basa la wikipedia, està elaborada en PHP i disposa d'una comunitat molt extensa i madura que col·labora en el seu desenvolupament.

Ofereix una estructura de programació modular, que facilita als desenvolupadors diversos punts d'extensió (Pàgines especials, parsers o altres elements).

Això fa que existeixi una àmplia oferta de mòduls per cobrir moltes funcionalitats per una banda, però per l'altra, el seu manteniment i actualització pot ser desigual, dificultant el manteniment sobretot quan es produeixen grans dependències entre mòduls.

Representació de la metadada

MediaWiki basa la seva estructura en pàgines:

- Cada pàgina té un nom únic que la identifica.
- Les pàgines s'agrupen en col·leccions anomenades “espais de noms” o “namespaces”, que permeten una primera ordenació d'alt nivell.
- Les plantilles són pàgines estàndards de contingut que s'afegeixen dintre d'altres pàgines, oferint la possibilitat de reaprofitar contingut.
- Les categories són un “espai de noms” especial, que permet agrupar pàgines relacionades de forma transversal a qualsevol altre espai de noms, això ens ofereix poder estructurar les pàgines en taxonomies diferents independents.

Sematic Mediawiki, aprofita aquesta estructura i l'amplia amb els components necessaris per representar el contingut semàntic, afegint:

- Propietats i tipus (`[[nom propietat::valor]]`), permeten modelar les propietats d'objecte i de Datatypes, tipus especials (url, string...).
- Millora les categories, generant inferències a través de l'herència de subcategories. De forma que una cerca per una categoria, ens donaria com a resultat les pàgines que pertanyen també a les seves subcategories. Aquest tipus d'inferència també s'aplica a les propietats i subpropietats.
- Millora les plantilles, amb l'objectiu de simplificar l'anotació semàntica facilitant la feina d'edició. Cal destacar l'ús d'infobox, que permeten representar el contingut semàntic clarament en una caixa separada, facilitant la seva visualització a part de poder-lo fer servir com element de navegació.

Per tal de representar una ontologia OWL DL amb els elements anteriors haurem d'efectuar el següent mapeig:

| Construcció OWL | Construcció a Semantic Media Wiki |
|-----------------------------------|-------------------------------------|
| Classe | Categoria |
| Propietat del Datatype | Propietat |
| Propietat de l'objecte | Propietat |
| Instanciació de classe | Pàgina categoritzada |
| Subclasse | Subcategoria |
| Element | Pàgina dintre del seu espai de noms |
| Propietat de Datatype instanciada | Anotació de l'atribut |
| Propietat d'objecte instanciada | Enllaç amb tipus |

Semantic Media Wiki també afegeix també la possibilitat d'efectuar cerques semàntiques introduint:

- El llenguatge de consulta ASK, que és una simplificació de SPARQL, a través del qual podem fer diversos tipus de cerca:
 - Dinàmiques, a través del cercador o la pàgina especial (Special:ASK)
 - Emmagatzemades o en línia, representant el resultat dintre de la mateixa pàgina.
 - * Un ús particular d'aquestes cerques es dona a través de la generació de Conceptes a través de cerques, i que disposen d'un espai de noms individual (Concept:namespace)

Entre les coses que no permet:

- No permet l'expressivitat completa de OWL DL (arriba fins EL++), entre d'altres coses:
 - no suporta transitivitat.
 - no permet restriccions de domini ni rang.
 - no permet propietats funcionals ni restriccions numèriques.

Incorporació de la ontologia

La incorporació de la ontologia és un treball complicat, existeixen múltiples mòduls que efectuen aquesta tasca però es troben desactualitzats.

4.1 Sematic Media Wiki

La mateixa documentació aporta com a solució l'elaboració d'un script en python que recorre l'ontologia i la transforma generant les pàgines segons el criteri establert al mapeig.

Una altra alternativa és la generació gradual i manual a través de formularis semàntics

4.2 Semantic Media Wiki Plus

SMW+ és una millora comercial sobre Semantic Media Wiki Plus, per tant disposa de les mateixes característiques base, aportant millores.

Entre aquestes millores destaquem la pila de mòduls creats i mantinguts per l'empresa ontoprise que aporten:

- Consistència i continuïtat dels mòduls. Donant estabilitat a la plataforma de dades, facilitant la seva gestió i actualització.
- Afegeix nous mòduls d'edició (wysiwyg), navegació i consulta sobre la ontologia, que faciliten la feina de generació i manteniment de les dades.
- Sistema d'importació i exportació de dades, inclòs amb interconnexió amb fonts de Linked Data, que faciliten l'explotació de les dades contingudes a la plataforma.

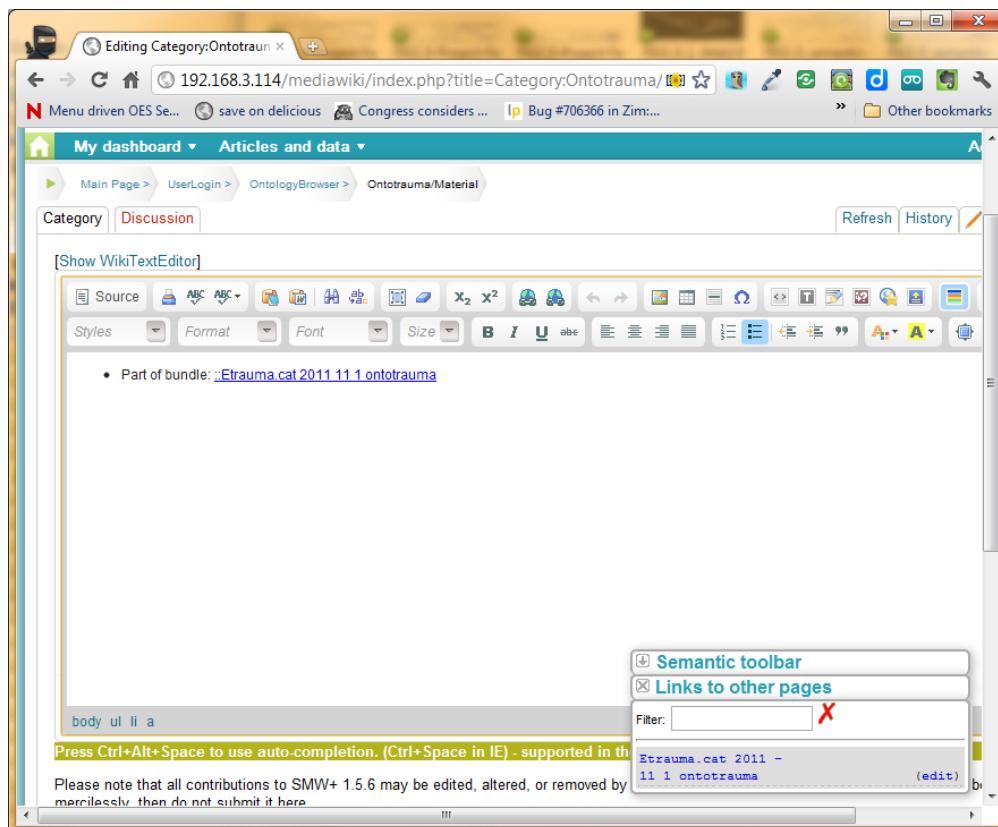


Figura 4.1: Edició millorada en SMW+

Incorporació de la ontologia

Es pot realitzar una importació directa a través de l'eina incorporada.

Un cop disposem de les dades, disposem de diversos mecanismes d'instanciació de noves dades: el tradicional aportat per MediaWiki, o bé, alternatius inclòs drag&drop sobre el seu gestor de dades.

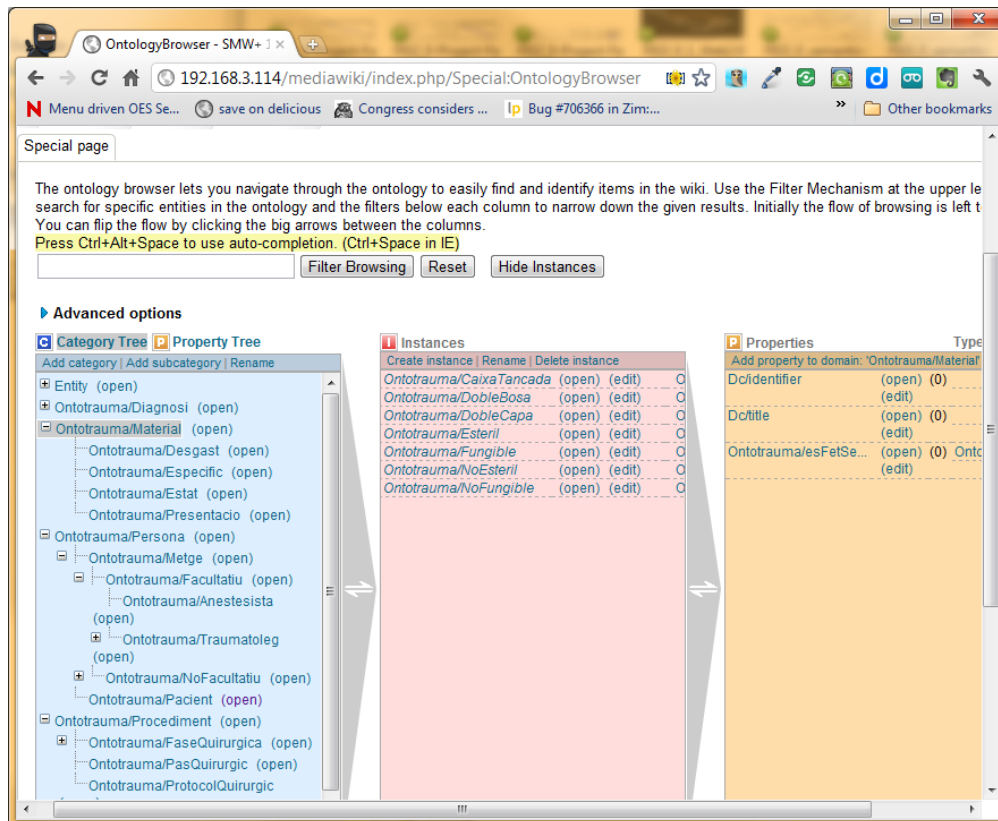


Figura 4.2: Incorporació ontologia a SMW+

4.3 OntoWiki

Sorgeix del mateix grup que va impulsar la DBpedia, per aconseguir una plataforma que permeti la generació de contingut semàntic directament.

Cerca seguir el mateix principi dels wikis (fer fàcil corregir errades, en comptes de fer difícil que es cometin), sense fer ús d'una sintaxi especial per anotacions RDF sobre les pàgines de wiki, com efectua SMW.

Per aconseguir-ho, introdueix el concepte de “mapes d’informació”, on cada node es representa visualment en una pàgina independent, que pot ser enllaçada amb els recursos digitals relacionats. I segueix les següents directrius:

- una interfície totalment basada en formularis, eliminant la sintaxi.
- vistes semàntiques, que permetin agregar contingut dintre de la base de dades.
- cerques semàntiques, full-text de totes les dades literals que poden ser filtrades i cercades, fent servir les seves relacions semàntiques. Així com l’ús del llenguatge SPARQL per les cerques, i un end-point (/sparql) pel seu enllaç i publicació de dades.
- Suport de la comunitat, versionatge i història, que permeti la discussió i revisió de cada element d’informació.
- Sindicació semàntica, que permeti distribuir el contingut de la manera més adaptada.

Emmagatzematge de les dades semàntiques

Fa servir el concepte dels “mapes d’informació” permeten modelar tots els elements d’una ontologia OWL DL completament de forma directa i visual.

Malauradament no efectua comprovació de la correctesa de tipus en efectuar instanciament dels elements.

Incorporació de la ontologia

La incorporació és directa i trivial a través de l’eina d’importació incorporada.

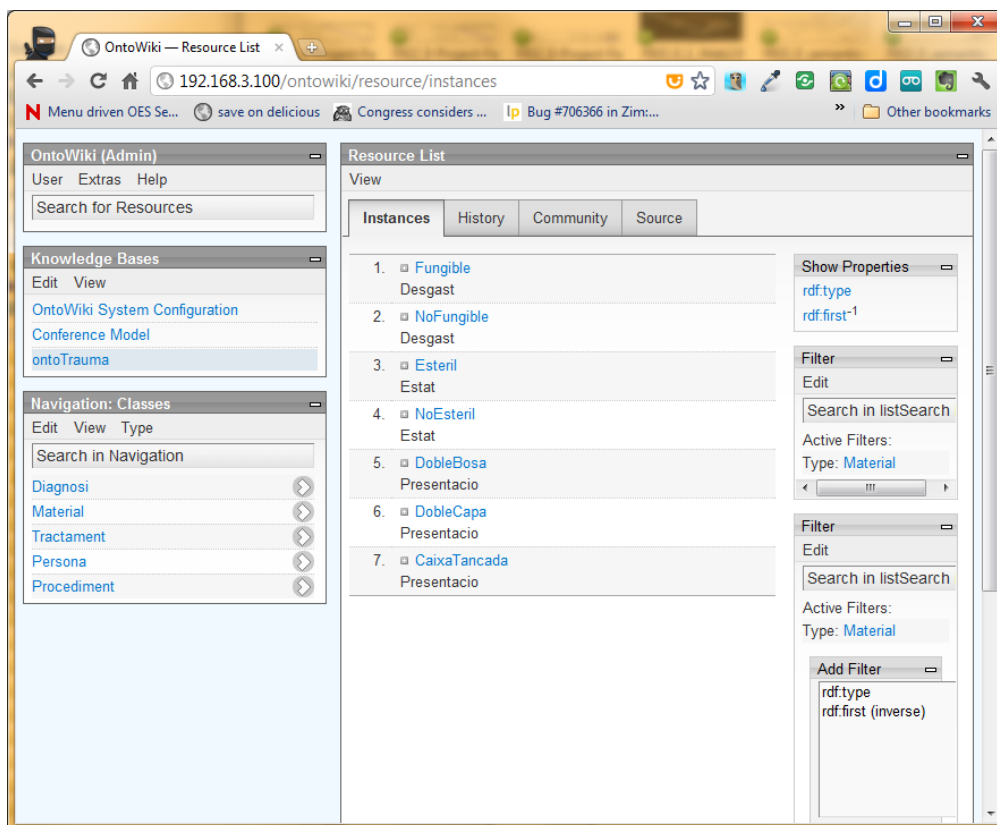


Figura 4.3: Incorporació ontologia a Ontowiki

4.4 Diferències i selecció de programari

Entre MediaWiki i SMW+ la diferència es troba bàsicament en la qualitat de la solució integrada, amb tots els mòduls que aporta. Tot i que la filosofia de base és la mateixa: aprofitar un wiki i estendre'l amb anotacions semàntiques.

En el cas d'Ontowiki, es troba un gran canvi de filosofia en la representació de la ontologia que permet una expressivitat completa de OWL DL. Això pot ser important a llarg plaç dependent de l'evolució de la plataforma.

Donat que l'àmbit d'aplicació pot arribar a considerar-se “de missió crítica”, ja que es tracta de l'àmbit de la salut, i l'estat de maduresa incipient de Ontowiki, crec oportú fer ús de SMW+ en la seva versió Community Edition.

En aquest entorn és important disposar d'un manteniment, i un suport més intensiu que pot ser aportat per l'empresa ontowiki. Partim de la Community Edition

lliure per temes de cost, però es pot escalar fàcilment a altres versions amb més funcionalitats.

Un mòdul aportat per SMW+ molt interessant per la definició que hem fet és la de visualització i gestió de projectes, s'intentarà aprofitar el màxim possible d'aquesta funcionalitat per tal de visualitzar el seguiment dels pacients, així com la planificació i seguiment de la intervenció.

Emmagatzematge físic de les dades semàntiques

Els tres wikis plantejats fan servir com a base de dades principal un sistema relacional (RDBMS, Relational DataBase Management System) i ofereixen addicionalment l'interconnexió amb altres sistemes d'emmagatzematge especialitzats en dades semàntiques.

L'emmagatzematge de dades semàntiques en una base de dades relacional requereix un mapeig costós que pot limitar l'ús d'aquestes dades depenent de la seva implementació.

Un sistema d'emmagatzematge de tripletes (triplestore) és un sistema optimitzat per l'emmagatzematge i recuperació d'aquestes. En emmagatzemar els grafs RDF ens permet millorar la qualitat i integració de les dades, a la vegada que ens ofereix la possibilitat d'explorar les seves possibilitats d'inferència.

Una implementació eficient per consultar els triplestores ens ofereix normalment un punt de consulta SPARQL.

SPARQL (Simple Protocol and RDF Query Language) és un llenguatge que defineix dos estàndards: un protocol d'accés a les dades, i un llenguatge de consulta sobre els grafs RDF.

Virtuoso és un exemple de sistema natiu d'emmagatzematge de triples que es pot integrar amb els tres wikis. Ens ofereix una API, suport per SPARQL i un webservice per fer les consultes a través de HTTP.

4.4 Diferències i selecció de programari

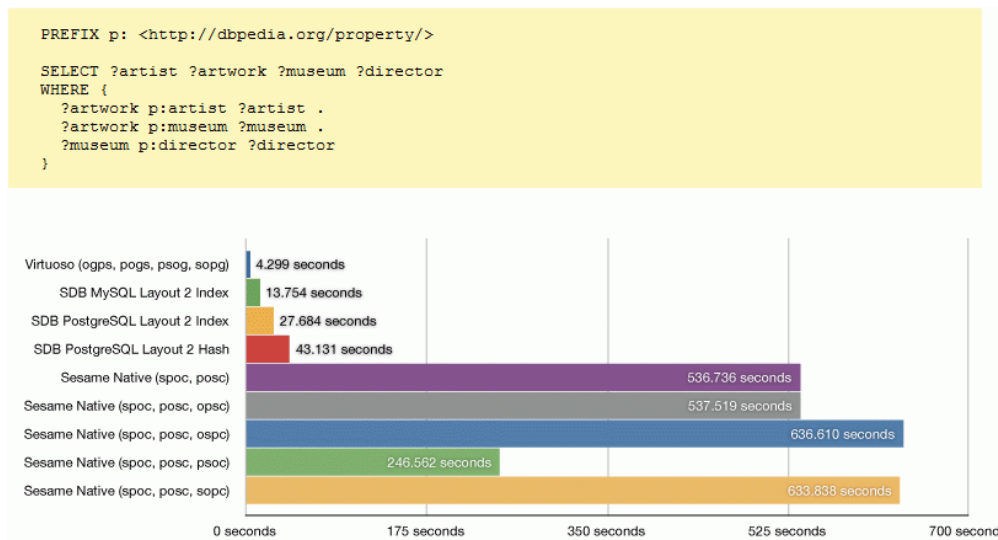


Figura 4.4: Rendiment de diversos triplestores en una consulta sobre la DBPedia [Bec]

Tots aquests sistemes ens ofereixen diverses eines per efectuar-ne el manteniment, còpia de seguretat de les dades.

5 Conclusió i línies de futur

Després de fer un repàs de la situació del Web fins a l'actualitat, queda palesa la necessitat de trobar un nou mecanisme que permeti el creixement del Web mantenint la seva qualitat de dades.

Precisament, mentre es finalitza el present document, tres empreses de cerca competidores, com son Bing, Yahoo i Google, han presentat una iniciativa conjunta anomenada `schema.org`, i que s'orienta a facilitar als propietaris de webs, etiquetes semàntiques per etiquetar contingut. Alguns comparen ja aquest esdeveniment com l'empenta mes important a la xarxa semàntica des d'ençà el 2001.

Les tecnologies del web semàntic es perfilen doncs com una solució a aquesta necessitat de creixement, construint un nou model web, anomenat de vegades web 3.0 o “el web de les coses”.

El grau de maduresa d'aquestes solucions fan que siguin d'especial interès dintre del mon empresarial per poder gestionar les seves pròpies dades. Aquest interès es posa clarament de manifest amb el sorgiment d'empreses especialitzades com `ontoprise (SMW+)` o `OpenLink Software (Virtuoso)`, o les fortes inversions públiques que ha efectuat la unió europea com el `LOD2` (segon projecte `Linked Open Data`).

En conclusió, la web semàntica és una bona proposta per afrontar els reptes actuals en la cerca i catalogació d'informació tant de la web pública, com dintre de les organitzacions.

Tot i que és molt probable que l'adopció de tecnologies semàntiques es vagi produint de forma gradual a mesura es vagin trobant els límits a les eines tradicionals, solucionant problemes puntuals, i que els veritables avantatges no puguin ser gaire visibles fins que no s'hagi implementat almenys la base de la pila semàntica proposada per Tim Berners Lee.

Línies de futur

S'apunten per futurs treballs l'estudi i implementació d'un triplestore integrat al wiki semàntic.

També l'aprofitament de dades existents a `Linked Open Data`, així com la possibilitat d'integrar la plataforma per oferir-ne les seves dades.

Bibliografía

- [Alv10] Rubén Darío Alvarado. Metodología para el desarrollo de ontologías. 2010.
- [Bec] Christian Becker. Rdf store benchmarks with dbpedia. Visited january 2012.
- [Bio11] Berkeley Bioinformatics. The open biological and biomedical ontologies, 2011.
- [BL00] Tim Berners-Lee. Semantic web - xml2000, slide 10, 05 2000. Semantic web stack.
- [BL09] Tim Berners-Lee. On the next web, 2009.
- [BLL01] Tim; James Hendler Berners-Lee and Ora Lassila. The semantic web. *Scientific American Magazine*, 2001, <https://www.scientificamerican.com/article.cfm?id=the-semantic-web>.
- [Bra09] Anthony J. Bradley. Don't be fooled web 3.0 doesn't exist, 2009.
- [Bur09] Oliver Burgert. Slides of tibm 2009 - vl surgical workflow, 2009.
- [Gar06] Gartner. Gartner's 2006 emergin technologies hype cycle highlights key technology themes, 2006.
- [Gru93] Thomas R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. 1993, <http://tomgruber.org/writing/onto-design.pdf>.
- [Gru09] T. R. Gruber. Ontology definition. 2009, <http://tomgruber.org/writing/ontology-definition-2007.htm>.
- [Her02] M. Chantal Pérez Hernández. Explotación de los córpora textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento, 2002.
- [Hoe10] Robert Hoehndorf. What is an upper level ontology?, April 2010.
- [JM10] Helen Parkinson James Malone. Reference and application ontologies, January 2010.
- [NHT09] NHTSA. Datalinkage - the crash outcome data evaluation system (codes) and applications to improve traffic safety decision-making. 2009, <http://www-nrd.nhtsa.dot.gov/Pubs/811181.pdf>.

- [O'R05] Tim O'Reilly. What is web 2.0, 09 2005.
- [RM11] Thomas Neumuth Raj Mudunuri, Oliver Burgert. Ontological modelling of surgical knowledge. page 11, 2011, <http://subs.emis.de/LNI/Proceedings/Proceedings154/gi-proc-154-61.pdf>.
- [Wik] Wikipedia. Ontology. Visited april 2011.