Experiences in Reusing Knowledge sources using Protégé and Prompt. An application in a medical domain

María Taboada
University of Santiago de Compostela.
Spain.

Introduction

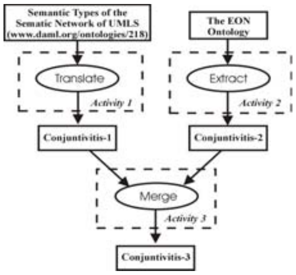
- Goal:
  - Modelling a Knowledge Base (KB) in a clinical domain

Conjunctivitis

- Guideline American Academy of Ophthalmology (AAO)
  - http://www.aao.org/aao/education/library/ppp.cfm

Introduction

Development of the domain ontology



Development of the ontology ontology

- We have chosen PROMPT for several reasons:
  - It allows us to merge source ontologies into a resulting ontology
  - It analyses source concepts, properties (including restriction on value properties) and relationships
  - It is interactive with the user, allowing us to accept or reject the suggestions

Development of the domain ontology

- Three types of conflicts identified by PROMPT:
  - Multiple paths from a same concept
  - Not defined concepts exist
  - Different defined types for a same concept

Experiences and Difficulties using Prompt

- PROMPT automatically identifies identical concepts from different sources for merging.
- However, if the concepts are not identical, it is necessary to manually identify them.
- So, the reuse time could be reduced if PROMPT uses a synonym dictionary.

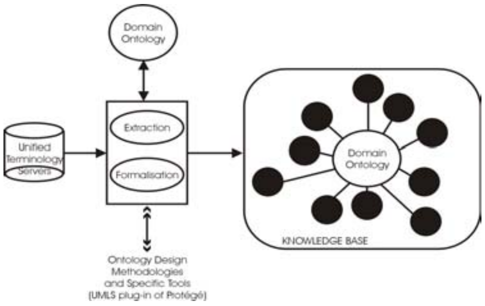
Experiences and Difficulties using Prompt

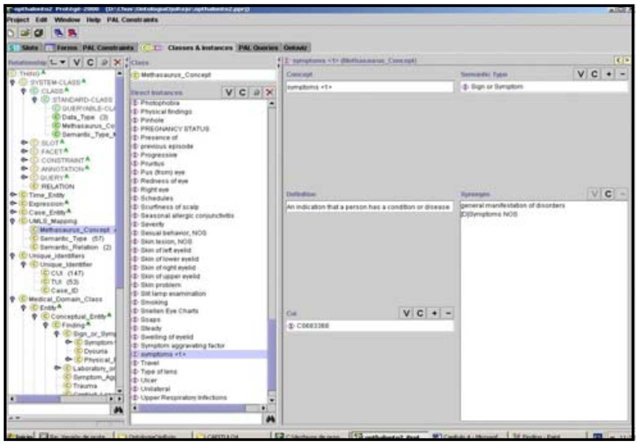
- PROMPT identifies dangling references and, for each of them, it makes a suggestion:
  - Copying a concept that does not exist
- If you accept the suggestion, PROMPT automatically copies the concept
- But if you reject it, PROMPT does not facilitate any operation
- We identified two possible operations:
  - Removing the property that made the dangling reference
  - Changing the restriction on the property values

Experiences and Difficulties using Prompt

- We have detected that we made some operations repeatedly.
- Example
  - Each time PROMPT identified the conflict 'Multiple paths from a concept'
    - We removed all parents relative to the most external nodes in the hierarchy
- An improved alternative:
  - Configuring operations when a conflict is detected
  - Take into account the particular heuristic of the each case.

Specialisation of the core ontology


Domain Ontology
Unified Terminology Servers
Extraction
Formalisation
Domain Ontology
KNOWLEDGE BASE
Ontology Design Methodologies and Specific Tools (UMLS plug-in of Protégé)


Project Edit Window Help PAL Constraints
Slots Forms PAL Constraints Classes & Instances PAL Queries Ontoviz
Relationship
Class
Direct Instances
Concept
Semantic Type
Definition
Synonym
Cui
V C + −

THING
SYSTEM-CLASS
CLASS
STANDARD-CLASS
QUERYABLE-CLA
Data_Type (3)
Methasaurus_Co
Semantic_Type_
SLOT
FACET
CONSTRAINT
ANNOTATION
QUERY
RELATION
Time_Entity
Expression
Case_Entity
UMLS_Mapping
Methasaurus_Concept
Semantic_Type (57)
Semantic_Relation (2)
Unique_Identifiers
Unique_Identifier
CUI (147)
TUI (53)
Case_ID
Medical_Domain_Class
Entity
Conceptual_Entity
Finding
Sign_or_Symp
Symptom
Dysuria
Physical_
Laboratory_or
Symptom_Ag
Trauma

Class: Methasaurus_Concept

Direct Instances: Photophobia, Physical findings, Pinhole, PREGNANCY STATUS, Presence of, previous episode, Progressive, Pruritus, Pus (from) eye, Redness of eye, Right eye, Schedules, Scurfiness of scalp, Seasonal allergic conjunctivitis, Severity, Sexual behavior, NOS, Skin lesion, NOS, Skin of left eyelid, Skin of lower eyelid, Skin of right eyelid, Skin of upper eyelid, Skin problem, Slit lamp examination, Smoking, Snellen Eye Charts, Soaps, Steady, Swelling of eyelid, Symptom aggravating factor, symptoms <1>, Travel, Type of lens, Ulcer, Unilateral, Upper Respiratory Infections

symptoms <1> (Methasaurus_Concept)

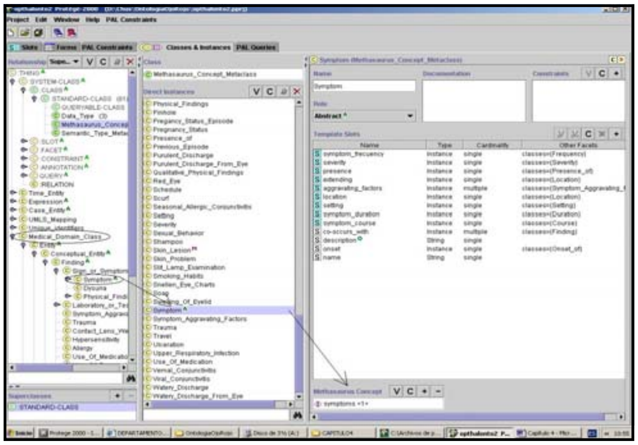
Concept: symptoms <1>

Semantic Type: Sign or Symptom

Definition: An indication that a person has a condition or disease

Synonym: general manifestation of disorders; [D]Symptoms NOS

Cui: C0683368



THING
SYSTEM-CLASS
CLASS
STANDARD-CLASS
QUERYABLE-CLASS
Data_Type (3)
Methasaurus_Concep
Semantic_Type_Meta
SLOT
FACET
CONSTRAINT
ANNOTATION
QUERY
RELATION
Time_Entity
Expression
Case_Entity
UMLS_Mapping
Unique_Identifiers
Medical_Domain_Class
Entity
Conceptual_Entity
Finding
Sign_or_Symptom
Symptom
Dysuria
Physical_Findi
Laboratory_or_Te
Symptom_Aggrava
Trauma
Contact_Lens_We
Hypersensitivity
Allergy
Use_Of_Medicatio

Superclasses: STANDARD-CLASS

Class: Methasaurus_Concept_Metaclass

Direct Instances: Physical_Findings, Pinhole, Pregancy_Status_Episode, Pregnancy_Status, Presence_of, Previous_Episode, Purulent_Discharge, Purulent_Discharge_From_Eye, Qualitative_Physical_Findings, Red_Eye, Schedule, Scurf, Seasonal_Allergic_Conjunctivitis, Setting, Severity, Sexual_Behavior, Shampoo, Skin_Lesion, Skin_Problem, Slit_Lamp_Examination, Smoking_Habits, Snellen_Eye_Charts, Soap, Swelling_Of_Eyelid, Symptom, Symptom_Aggravating_Factors, Trauma, Travel, Ulceration, Upper_Respiratory_Infection, Use_Of_Medication, Vernal_Conjunctivitis, Viral_Conjunctivitis, Watery_Discharge, Watery_Discharge_From_Eye

Symptom (Methasaurus_Concept_Metaclass)

Name: Symptom

Role: Abstract

Name	Type	Cardinality	Other Facets
symptom_frecuency	Instance	single	classes={Frequency}
severity	Instance	single	classes={Severity}
presence	Instance	single	classes={Presence_of}
extending	Instance	single	classes={Location}
aggravating_factors	Instance	multiple	classes={Symptom_Aggravating_
location	Instance	single	classes={Location}
setting	Instance	single	classes={Setting}
symptom_duration	Instance	single	classes={Duration}
symptom_course	Instance	single	classes={Course}
co-occurs_with	Instance	multiple	classes={Finding}
description	String	single	
onset	Instance	single	classes={Onset_of}
name	String	single	

Methasaurus Concept: symptoms <1>

Experiences and Difficulties using Protégé

- Meta-classes provides:
  - We are able to classify concepts from different sources in the same medical domain hierarchy
  - At the same time, we conserve the origin of the information.
  - We can easily distinguish between UMLS terminological information and essential properties of a class or instance
  - We do not overwhelm the user with excessive information
  - The maintaining of the KB is simplified

Experiences and Difficulties using UMLS

- We searched concepts using
  - The web based interaction tools provided by the UMLS Knowledge Source Server (UMLSKS)
  - The UMLS plug-in
- With regard to the search process
  - This way of displaying the information is more clear using the UMLS plug-in
    - The relevant information is located more quickly
    - The way of displaying the ordered list of search results and the narrow tree of a concept is more clear

Experiences and Difficulties using UMLS

- With regard to the import process
  - In Protégé it is quick
    - Selecting the concept to be imported
    - Selecting the place (class or instance) in the KB where the information will be imported
    - The plug-in automatically creates all slots and fills them with the search results.
    - You can directly import not only the searched concept, but also any concept in the narrow tree.

Experiences and Difficulties using UMLS

- With regard to the import process
  - The UMLS tab prevent us from importing the information directly into our UMLS Mapping class
    - The information can be imported as a class or instance
    - But, the definition of the imported slot is predefined
    - The place (class or instance) where the information will be imported is also predefined
    - This definition can be inadequate for all applications.

Experiences and Difficulties using UMLS

- For example, the *cui* slot is a single string and the slot *semantic type* is a multiple string.
- We decided to model the string *cui* separately from the class representing the UMLS information
  - The *cui* slot is an instance of the *CUI* subclass of the *Unique Identifier* class
  - We can verify whether an identifier is unique or not, by defining a single PAL constraint
- We represent the slot *semantic type* as an instance of the *Semantic Type* class.
- In our KB, Methasaurus concepts and Semantic types are related to each other.

Experiences and Difficulties using UMLS

- There is no means of selecting the information to be imported
  - It could be of interest to import only some slots.

THANK YOU
VERY MUCH!!!