Semantic Query Generation from Eligibility Criteria in Clinical Trials

Chintan O. Patel, MS, James J. Cimino, MD
Department of Biomedical Informatics, Columbia University, New York, NY

Abstract

Towards the goal of automated eligibility determination for clinical trials from electronic health records, we propose a method to formulate Semantic Web based queries using the free-text eligibility criteria on clinicaltrials.gov.

Introduction

Screening eligible patients from electronic health records that match the clinical trials is a challenging problem. Towards this goal, one of the first steps is to represent the eligibility criteria as a formal query (such as SQL or Arden Syntax), which can be executed against the clinical data to identify the matching patients. We propose a Semantic Web based query model to encode the eligibility criteria as SPARQL[1] queries.

Methods

We have developed a tool (Figure 1) that allows browsing clinical trials from the clinicaltrails.gov website. When the user enters a clinicaltrials.gov URL, the corresponding eligibility criteria are extracted from the XML version of the clinical trials provided by the website. The eligibility criteria do not contain distinct structured sections for the "inclusion" and "exclusion" criteria, hence we use a set of regular expressions to identify the respective sections. Using the MetaMap (MMTx)[2] natural language processing tool from the NLM, we identify the matching Unified Medical Language System (UMLS) concepts from the free-text eligibility criteria. The Semantic Types of the concepts are used to filter the irrelevant concepts such as *Before* (qualifier), *condition* (attribute), *Scientific Study* etc. The remaining concepts are then mapped to a specific target terminology that has been used to code the clinical data. We restrict the UMLS concepts to SNOMED-CT as target terminology and populate them into the inclusion and exclusion lists respectively.

A description logic editor allows the user to manually formulate a complex expression using logical constructs. The relationships are loaded from a separate ontology file. A SPARQL query is generated based on the given complex concept constructed using the criteria. The user can edit the SPARQL to add the non-semantic sections of the query such as checking temporal constraints, gender and so on.

Future Work and Conclusions

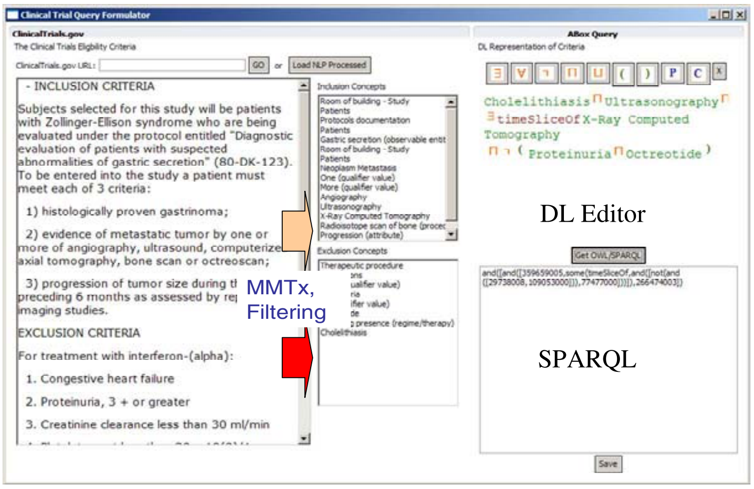
We plan to add functionality to allow automatic creation of logic queries based on the semantics of the concepts. Using a semantic query model for eligibility determination is critical to improving the matching results and the proposed standards based solution provides an important step in the direction.

Acknowledgements

The work was done at the IBM T. J. Watson Research Center as part of summer internship.

References

1. SPARQL,http://www.w3.org/TR/rdf-sparql-query/
2. MetaMap Transfer, http://mmtx.nlm.nih.gov/



- INCLUSION CRITERIA

Subjects selected for this study will be patients with Zollinger-Ellison syndrome who are being evaluated under the protocol entitled "Diagnostic evaluation of patients with suspected abnormalities of gastric secretion" (80-DK-123). To be entered into the study a patient must meet each of 3 criteria:

1) histologically proven gastrinoma;

2) evidence of metastatic tumor by one or more of angiography, ultrasound, computerize axial tomography, bone scan or octreoscan;

3) progression of tumor size during tl preceding 6 months as assessed by rep imaging studies.

EXCLUSION CRITERIA

For treatment with interferon-(alpha):

1. Congestive heart failure

2. Proteinuria, 3 + or greater

3. Creatinine clearance less than 30 ml/min

Room of building - Study
Patients
Protocols documentation
Patients
Gastric secretion (observable entit
Room of building - Study
Patients
Neoplasm Metastasis
One (qualifier value)
More (qualifier value)
Angiography
Ultrasonography
X-Ray Computed Tomography
Radioisotope scan of bone (procec
Progression (attribute)

Therapeutic procedure
ons
ualifer value)
ria
ifier value)
de
presence (regime/therapy)
Cholelithiasis

Cholelithiasis ⊓ Ultrasonography ⊓ ∃ timeSliceOf X-Ray Computed Tomography ⊓ ¬ (Proteinuria ⊓ Octreotide)

```
and([and([359659005,some(timeSliceOf,and([not(and
([29738008,109053000])),77477000])))]),266474003])
```

EligibilityConcept = ∃hasPossibleDisorder. Staph Aureus Bacteremia ∩ ¬ (∃ hasLabResult. (¬Staphylococuus Aureus ∩ ∃ hasLabSpecimen.Blood) ∩ ∃onMedication. ∃hasActiveIngredient. Warfarin)

```
select ?patients from PatientKB where
?patients rel:hasEvent ?Event
?Event rel:hasGender "Female"
?Event owl:EquivalentWith EligibilityConcept
```

Figure 1. Encoding eligibility criteria from cinicaltrials.gov into SPARQL