Homework #2 Solutions

1. [20 pts]
   a. PREFIX sp: <http://swat.cse.lehigh.edu/resources/onto/scipub.rdf#>
      SELECT ?title
   
   b. PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
      PREFIX sp: <http://swat.cse.lehigh.edu/resources/onto/scipub.rdf#>
      SELECT ?title ?type ?year
      WHERE { ?p a ?typeId .
        ?typeId rdfs:label ?type .
        FILTER (?year >= 2001 && ?year <= 2005) }

   c. PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
      PREFIX sp: <http://swat.cse.lehigh.edu/resources/onto/scipub.rdf#>
      PREFIX sw: <http://swat.cse.lehigh.edu/resources/onto/swpub.rdf#>
      SELECT ?title ?venue ?year ?topic
      WHERE { ?p sp:publishedYear ?year .
        FILTER regex(?venue,"ISWC") }
      ORDER BY DESC(?year) ?topic 

   d. PREFIX sp: <http://swat.cse.lehigh.edu/resources/onto/scipub.rdf#>
      SELECT ?authorName (COUNT(?p) AS ?numPapers)
      WHERE { ?p sp:authorList ?list .
        ?author sp:fullName ?authorName . }
      GROUP BY ?authorName
      ORDER BY ?authorName

2. [10 pts.]
   PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
   PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
   CONSTRUCT { ?y rdf:type ?r }
   WHERE { ?p rdf:range ?r .
        ?x ?p ?y }
3. [10 pts.]
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
CONSTRUCT {?u owl:sameAs ?v }
WHERE { ?u foaf:mbox ?x .
    ?v foaf:mbox ?x .
    FILTER (?u != ?v) }
Converting ?u and ?v to strings and then comparing with > is okay too.

4. The links generated by this approach will be high, but not perfect precision. Usually, two people with the same e-mail address will be the same, but there are situations when people in the same household may share the same e-mail, or people with a given role at an organization may have the same e-mail address at different points in time. The approach will have moderate recall, because different profiles (especially those reflecting the person at different points in time) might not have identical e-mail addresses. There are lots of other ways of determining if two people are the same: SSN, similar name and identical birthday, same name and address (although one must be careful about Sr./Jr./III, etc.). The advantage of the approach is that it will find many owl:sameAs statements. However, there are two disadvantages: first an imperfect precision means that occasionally a false owl:sameAs will be discovered, which will lead to faulty inferences. Second, it will miss many potential owl:sameAs statements by focusing only on one narrow form of identity.

5. See sample input files and output file on course web site.