Minim definition for chemmim testing of chembox data

This file contains a checklist to test presence of some key attributes in chembox chemical data extracted from dbpedia. The checklist is not currently a complete assessment of chemmim completeness (@@ref), but has been constructed to illustrate the key testing patterns. It was used as part of some evaluatrion work to compare with results produced by Matthew Gamble using his SPIN-based MIM framework.

Prefixes:

Prefix	URI
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
owl	http://www.w3.org/2002/07/owl#
xsd	http://www.w3.org/2001/XMLSchema#
xml	http://www.w3.org/XML/1998/namespace
rdfg	http://www.w3.org/2004/03/trix/rdfg-1/
ore	http://www.openarchives.org/ore/terms/
ao	http://purl.org/ao/
dcterms	http://purl.org/dc/terms/
ro	http://purl.org/wf4ever/ro#
minim	http://purl.org/minim/minim#

Checklists:

Target	Purpose	Model	Description
*	complete	#minim_model	Checklist for sampling of chemmim attributes in chembox data

Model: #minim_mode — Model for chemmim attributes in chembox data

Items:	Level	Rule	
010	MUST	#req_inchi	Requirement for value to exist in metadata
020	SHOULD	#req_chemspider	Requirement for value to exist for each match of pattern
030	MAY	#req_synonym	Requirement for value to be aggregated for each match of pattern

Define rules to test individual requirements

Rule: #req_inchi

This requirement ensures exactly one chembox:StdInChI value is defined on the target resource, and that its value is a string literal. A SPARQL filter is used to check the type of the defined value.

ForEach:
```
?targetres chembox:StdInChI ?value .
FILTER ( datatype(?value) = xsd:string )
```

NOTE: ?targetres is pre-bound to the target resource of the checklist evaluation; i.e. the chemical whose description is being tested. (This checklist can be used with an RO containing multiple chemical

Min: 1
Max: 1
Pass: A single InChI value is present for %(targetres)s
Fail: InChI not present, or multiple InChI values present, for %(targetres)s

Rule: #req_chemspider

This requirement ensures at least one chembox:ChemSpriderId value is defined on the target resource, and that its value is an integer of a string containing a numeral. A SPARQL filter is used to check the type

ForEach: ?targetres chembox:ChemSpiderID ?value .
FILTER (str(xsd:integer(?value)))

NOTE: ?targetres is pre-bound to the target resource of the checklist evaluation; i.e. the chemical whose description is being tested. (This checklist can be used with an RO containing multiple chemical descriptions.)

The FILTER is currently commented out because of a SPARQL query processor problem, which may now be fixed.

"FILTERs eliminate any solutions that, when substituted into the expression, either result in an effective boolean value of false or produce an error"
-- http://www.w3.org/TR/rdf-sparql-query/#tests.

Further, the str() of any valid integer is a non-blank string, which in SPARQL yields an equivalent boolean value (EBV) of True. Thus, only

Min: 1
Pass: A ChemSpiderId value is present for %(targetres)s
Fail: No ChemSpiderId is present for %(targetres)s

Rule: #req_synonym

Check for existence of a synonym for the target chemical. This requirement tests for the presence of a chembox:OtherNames value.

Exists: ?targetres chembox:OtherNames ?value
Pass: A synonym is present for %(targetres)s
Fail: No synonym is present for %(targetres)s

End: