Information Extraction
with
Linked Life Data
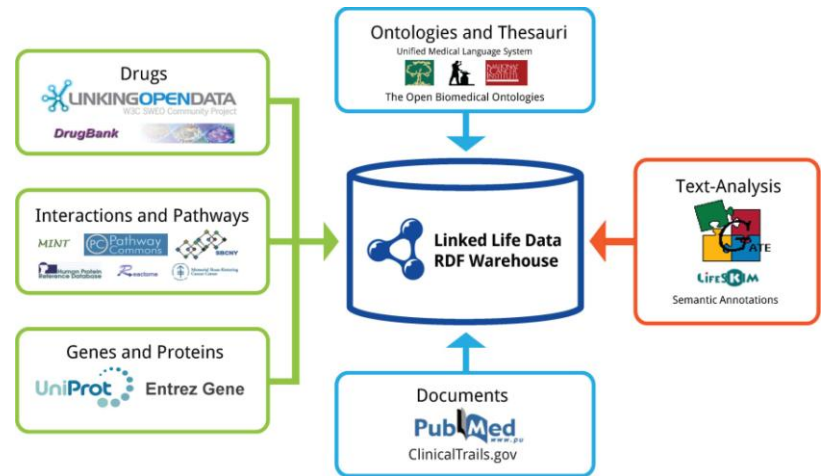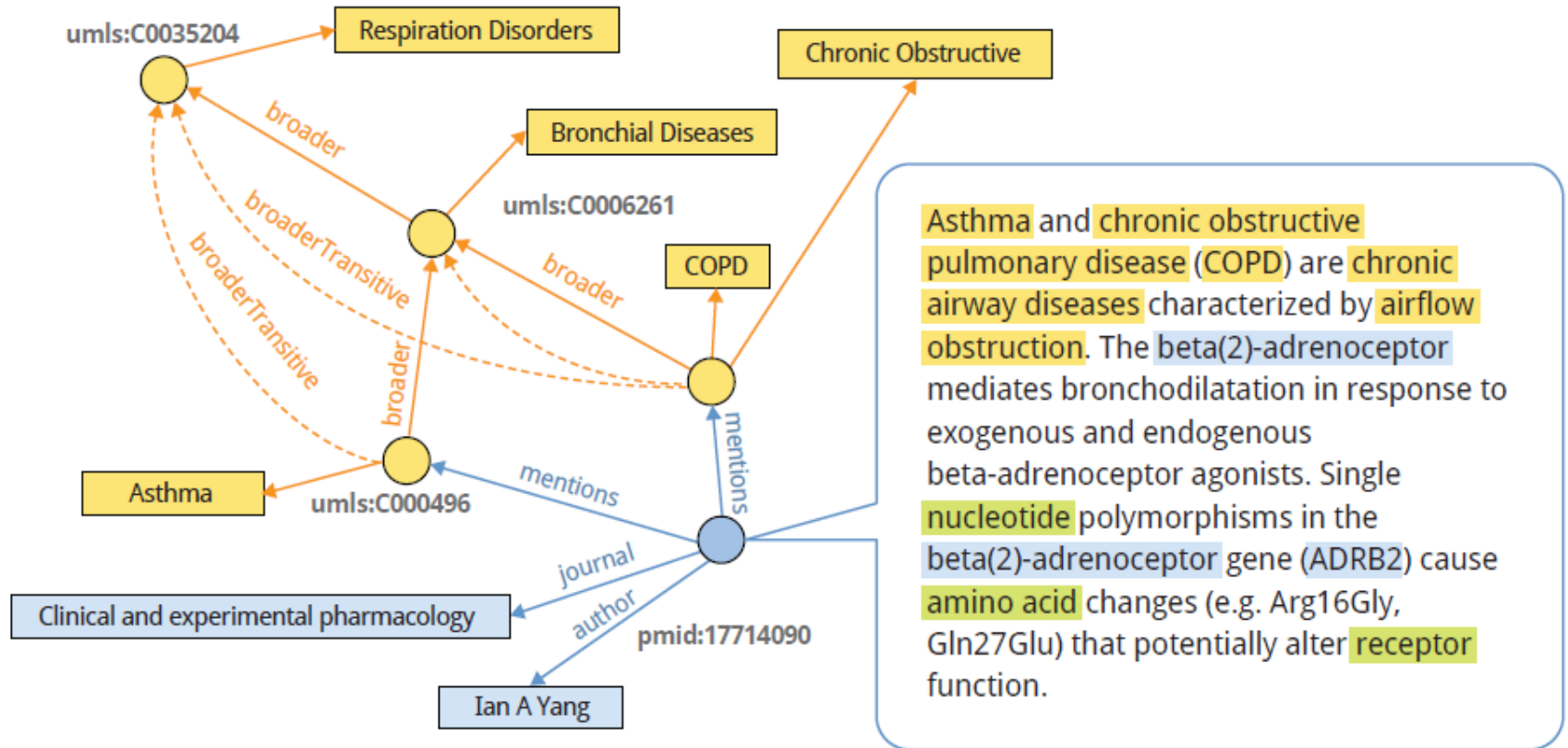
www.linkedlifedata.com

19/04/2011

Linked Life Data Overview

- Linked Life Data is a public and free RDF warehouse service

- Offers a SPARQL endpoint

- Integrates more than 25 popular biomedical data sources

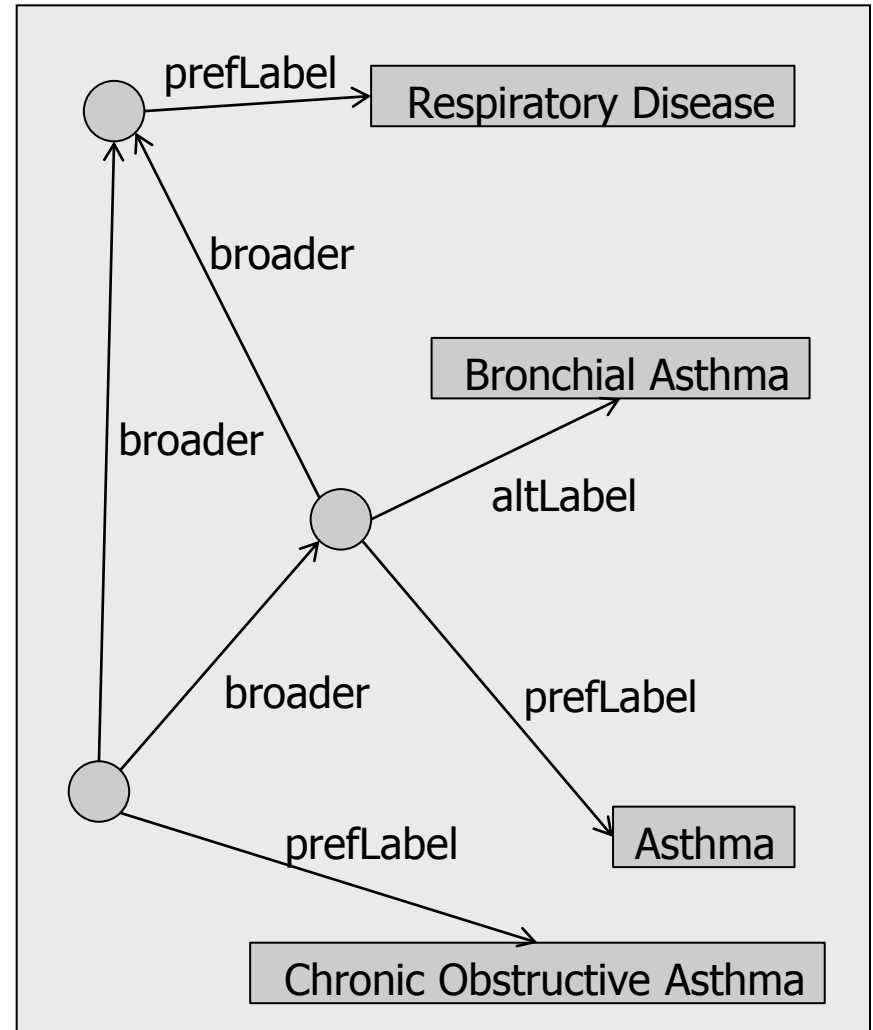- Specifies many cross data sources semantic mappings

- Exposes massive amounts of linked data
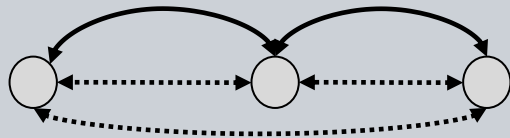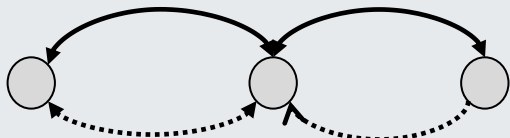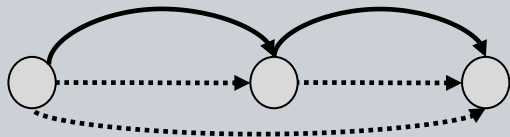

ontotext

Semantic Annotations with LLD

Simple Knowledge Organisation Schema (SKOS)

- SKOS is a data model

- Serialized as RDF graph

- Published on the web in a to be shared between applications

- Efficient structuring of terms in thesauri

- Allows the identification of the concepts by URLs

SKOS Semantics to Map Across Different Thesauri

Relationship	Semantics	Example
Exact match	Transitive equivalence	
Close match	Equivalent only for search purposes	
Broader match	Generalization of a concept	
Narrower match	Specialization of a concept	Inverse of broader match
Related	Unspecified relation (no real semantics)	

ontotext

Asthma http://linkedlifedata.com/resource/umls/id/C0004096

MSH: A form of bronchial disorder with three distinct components: airway hyper-responsiveness (RESPIRATORY HYPERSENSITIVITY), airway INFLAMMATION, and intermittent AIRWAY OBSTRUCTION. It is characterized by spasmodic contraction of airway smooth muscle, WHEEZING, and dyspnea (DYSPNEA, PAROXYSMAL). CSP: form of bronchial disorder associated with airway obstruction, marked by recurrent attacks of paroxysmal dyspnea, with wheezing due to spasmodic contraction of the bronchi. NCI: A chronic respiratory disease manifested as difficulty breathing due to the narrowing of bronchial passageways.

View as Triples   Download in JSON | RDF | N3/Turtle | N-Triples

Alternative labels (1 | 2 | 3)

- ASTHMA BRONCHIAL
  COSTART
- Bronchial Asthma
  National Drug File - Reference Terminology
  MedlinePlus
  MeSH
- Asthma
  Clinical Classifications Software
  Clinical Classifications Software
  ICPC
- Asthma [Disease/Finding]
  National Drug File - Reference Terminology
- ASTHMA, BRONCHIAL
  DXplain

Type

- Disease or Syndrome
- Pathologic Function
- Natural Phenomenon or Process
- Phenomenon or Process
- Event
- Biologic Function

Asthma http://linkedlifedata.com/resource/umls/id/C0004096

MSH: A form of bronchial disorder with three distinct components: airway hyper-responsiveness (RESPIRATORY HYPERSENSITIVITY), airway INFLAMMATION, and intermittent AIRWAY OBSTRUCTION. It is characterized by spasmodic contraction of airway smooth muscle, WHEEZING, and dyspnea (DYSPNEA, PAROXYSMAL). CSP: form of bronchial disorder associated with airway obstruction, marked by recurrent attacks of paroxysmal dyspnea, with wheezing due to spasmodic contraction of the bronchi. NCI: A chronic respiratory disease manifested as difficulty breathing due to the narrowing of bronchial passageways.

Different formats

View as Triples Download in JSON | RDF | N3/Turtle | N-Triples

Concept definition

Alternative labels (1 | 2 | 3)

- ASTHMA BRONCHIAL
  COSTART
- Bronchial Asthma
  National Drug File - Reference Terminology
  MedlinePlus
  MeSH
- Asthma
  Clinical Classifications Software
  Clinical Classifications Software
  ICPC
- Asthma [Disease/Finding]
  National Drug File - Reference Terminology
- ASTHMA, BRONCHIAL
  DXplain

Concept labels

Type

- Disease or Syndrome
- Pathologic Function
- Natural Phenomenon or Process
- Phenomenon or Process
- Event
- Biologic Function

Concept types

Querying LLD web interface with SPARQL

- LLD supports a database like web interface for executing SPARQL queries, where:

- You can test you query

- Populate directly from the database your gazetteer (e.g. LKB Gazetteer)

SPARQL Query

Append predefined namespaces:
bbc-pont, biopax-2, cpath, dailymed, dailymed-drugs, dailymed-ingredient, dailymed-instance, dbp-ont, dbp-prop, dbpedia, dbpedia-page, dbpedia-resource, dbtune, dc, dc-term, diseasome, diseasome-diseases, diseasome-gene, diseasome-instance, drug-category, drug-target, drugbank, drugbank-class, drugbank-drug, drugtype, entrez-gene, entrez-goterm, entrezgene, factbook, fb, ff, foaf, gene-rif, geo-ont, geo-pos, geonames, gr, lhgdn, lifeskim, lingvoj, linkedct, linkedct-condition, linkedct-intervention, lld, music-ont, nytimes, oasis, om, opencyc, opencyc-en, ot, owl, protonkm, protons, protont, protonu, pubmed, pubmed-article, pubmed-author, pubmed-mesh, pubmed-meshheading, pubmed-qualifier, rdf, rdfs, sider, sider-drugs, skos, skos-xl, sw-vocab, ub, umbel, umbel-ac, umbel-en, umbel-sc, umls, umls-concept, umls-label, umls-semnetwork, uniprot, uniprot-protein, uniprot-unstable, wordn-sc, wordnet, wordnet16, xsd, yago

Query:

```
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX lld: <http://linkedlifedata.com/resource/>

SELECT DISTINCT ?concept ?label
WHERE {
    ?top skos:prefLabel "Respiration Disorders".
    ?concept skos:broader ?top.
    ?concept skos:inScheme lld:umls.
    ?concept rdfs:label ?label.
}
```

☑ Include inferred

[Execute]

SPARQL Query

Results for PREFIX skos: <http://www.... (100 of 1632)

View as Exhibit | Download in JSON | SPARQL Results in XML | SPARQL Results in JSON

concept	label
umls-concept:C0001127	Acidosis, Respiratory
umls-concept:C0001127	RESPIRATORY ACIDOSIS
umls-concept:C0001127	respiratory acidosis
umls-concept:C0001127	Respiratory acidosis
umls-concept:C0001127	ACIDOSIS RESPIRATORY
umls-concept:C0001127	ACIDOSIS, RESPIRATORY
umls-concept:C0001127	Acidoses, Respiratory
umls-concept:C0001127	Respiratory Acidoses
umls-concept:C0001127	Respiratory Acidosis
umls-concept:C0001127	Hypercapnic Acidosis
umls-concept:C0001127	hypercapnic acidosis
umls-concept:C0001127	Acidosis, Respiratory [Disease/Finding]
umls-concept:C0001883	Airway Obstruction

ontotext

Querying SKOS Data

```
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX lld: <http://linkedlifedata.com/resource/>
```

Namespace prefixes

```
SELECT DISTINCT ?label ?concept ?top
WHERE {
    ?top skos:prefLabel "Respiration Disorders".
    ?concept skos:broader ?top.
    ?concept skos:inScheme lld:umls.
    ?concept rdfs:label ?label.
}
```

concept with this name

child concepts

part of UMLS

all their synonyms

Return all "Respiration Disorder" concepts in LLD and all their IDs and labels

ontotext

How to use LLD from Gate Developer?

- Use LKB Gazetteer part of Gate Developer's distribution

- Modify "query.txt" file with the results you have tested

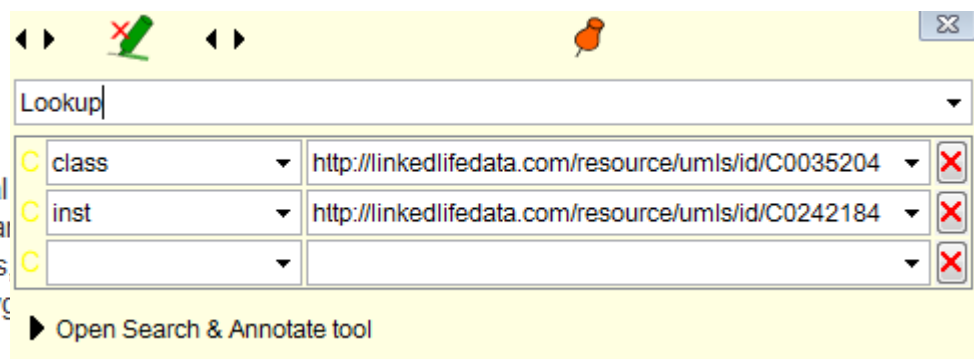- Point the gazetteer to the "repositoryURL" of:

http://linkedlifedata.com/sparql

The molecular basis for the renal compensation to respiratory acidosis and specifically the role of pendrin in this condition are unclear. Therefore, we studied the adaptation of the proximal tubule and the collecting duct to respiratory acidosis. Male Wistar-Hannover rats were exposed to either hypercapnia and hypoxia [8% CO(2) and 13% O(2) (hypercapnic, n = 6) or normal air (controls, n = 6)] in an environmental chamber for 10 days and were killed under the same atmosphere.

http://linkedlifedata.com/resource/umls/id/C0242184

Hypoxia

CSP: reduction of oxygen supply to tissue below physiological oxygen. NCI: Having too little oxygen. NCI: A decrease in the ar Symptoms range from mild (impaired judgment, memory loss, to severe (seizures and coma). NCI: Status of decreased oxyg blood, or tissues. -- 2003

Lookup

class   http://linkedlifedata.com/resource/umls/id/C0035204

inst   http://linkedlifedata.com/resource/umls/id/C0242184

▶ Open Search & Annotate tool

View as Triples   Download in JSON | RDF | N3/Turtle | N-Triples

ontotext

Conclusion