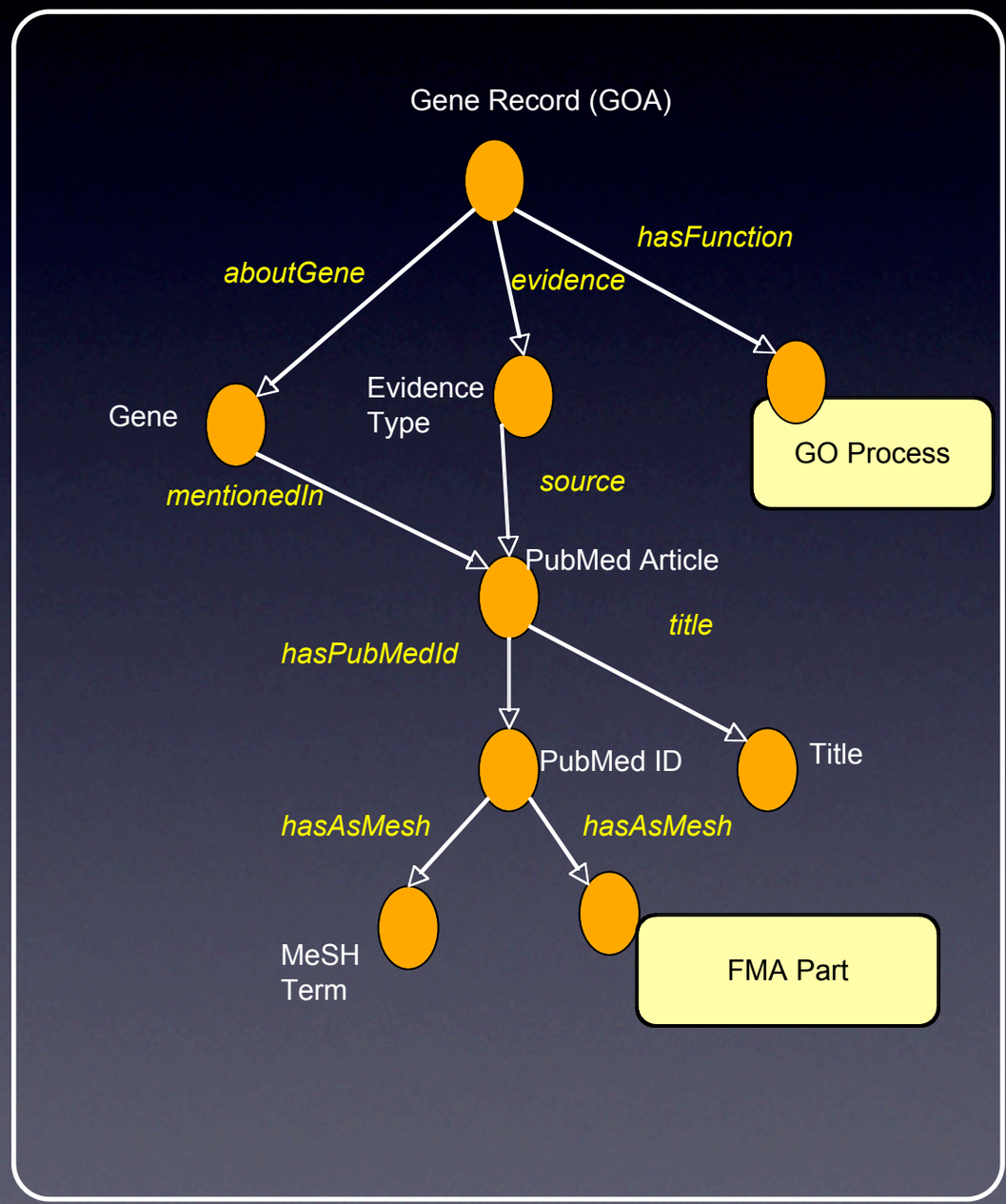
Logic based querying of integrated life sciences data

Julian Dolby, IBM Research
Achille Fokoue, IBM Research
Aditya Kalyanpur, IBM Research
Li Ma, IBM Research
Chintan Patel, Columbia University
Alan Ruttenberg, Science Commons
Edith Schonberg, IBM Research
Kavitha Srinivas, IBM Research

Can logic-based querying help?

- Building an integrated view of life sciences data (e.g., HCLS Banff Demo) is critical.
- Once integrated view is built, can logic-based querying provide additional value-add?
- In life sciences domain, many emerging ontologies that might help with semantic querying of data.
- Link the GOA and PubMed data to FMA, GO to see if logic based querying adds value.

Semantic querying of the integrated data



Query not by keywords, but by subclasses and parts of a given GO process:
e.g., *neuron development* expands to subprocesses such as *dendrite development*

Similarly, expand anatomical terms to their subparts e.g., *heart* expands to subparts such as *left ventricle*

Challenges

- Instance data set 300M RDF triples
- Provide a web search, must operate in web time.
- Build an intuitive key word based UI.
- FMA: Challenging for DL Reasoners because it defines deep part-of hierarchies, and has-part hierarchies.
  - ✳ Use only the part-of hierarchy to reason
  - ✳ Incorporate a fast, EL++ algorithm into SHER (reasoner for very large datasets). Complete, even if negation exists in the data, but not in web time.

A view of the service

SHER Life Sciences Demo - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://pasta.watson.ibm.com:8080/SHERService/SHERHCLSDemo.html#WHY%3A%20SUBGO(http%3A%2F%2Fpurl.org%2Fobo%2Fowl%2FGO%23

IBM Savings | Net Benefits | IBM HR Central Place | OWL 1.1 | owl-dev | Intern Offers | Voicemail | OWLED Review | PBC/IDP | IBM BluePages | SHER | Meebo | Room Reserve | GWT | GSA

SHER Life Sciences Demo

Find PubMed Articles... Search Stop

...containing MeSH term(s):

Specify MeSH Term

Lung (D008168)
Remove

...related to Anatomical Part(s):

Specify Part (using FMA Ontology)

...annotated by Gene Ontology (GO) Annotation(s)

Specify GO Annotation

GO Process Concept(s): [respiratory tube development (GO_0030323)]
GO Evidence: []
Remove

Published between 1990 and 2007

Query Results

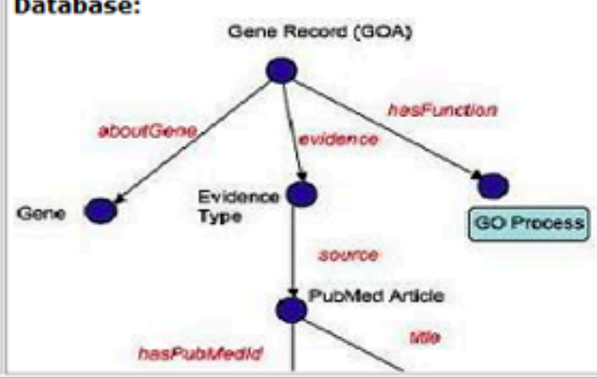
In addition to PubMed Articles, also show related: ☐ Anatomical Parts ☑ Gene/Gene Product Names ☑ GO Process Concepts

72 result(s) found in 28.444 seconds

10572052	The basic-helix-loop-helix protein pod1 is critically important for kidney and lung organogenesis. .	lung development	252856	?
15136295	Matrix GLA protein modulates branching morphogenesis in fetal rat lung. .	lung development	25333	?
7852342	Cloning and expression of glucocorticoid-induced genes in fetal rat lung fibroblasts. Transforming growth factor-beta 3. .	lung development	25717	?
14711803	Inhibition of CLC-2 chloride channel expression interrupts expansion of fetal lung cysts. .	lung development	29232	?
12508227	T-box gene products are required for mesenchymal induction of epithelial branching in the embryonic mouse lung. .	lung development	304514	?
15733665	Inactivation of tensin3 in mice results in growth retardation and postnatal lethality. .	alveolus development	319939	
14517255	Essential function of p300 acetyltransferase activity in heart, lung and small intestine formation. .	lung development	328572	?
7852342	Cloning and expression of glucocorticoid-induced genes in fetal rat lung fibroblasts. Transforming growth factor-beta 3. .	lung development	59086	?
16157679	l7Rn6 encodes a novel protein required for clara cell function in mouse lung development. .	lung development	67669	?
7852342	Cloning and expression of glucocorticoid-induced genes in fetal rat lung	lung development	81809	?

Result Explanation

Record matched the following graph pattern in the Database:



Found Direct Match for FMA Part / MeSH Term in Query

Found Indirect Match (alveolus development) for GO Process in Query (respiratory tube development)

Because...

lung development ⊆ ∃part_of.respiratory tube development

alveolus development ⊆ ∃part_of.lung development

part_of is Transitive

Done

Open issues

- Do we improve recall, but at the cost of precision?
- E.g., in FMA, it is possible to start with a higher level organ, and expand subparts to a cellular structure such as *ribosome*.
  - Granularity of part-of relations (FMA defines many part of relations, but the OWL file collapses all part-ofs)
  - Deeper problem when you have multiple inheritance? May be necessary to rank by depth of the hierarchy.