A Systematic Analysis of Term Reuse and Term Overlap across Biomedical Ontologies

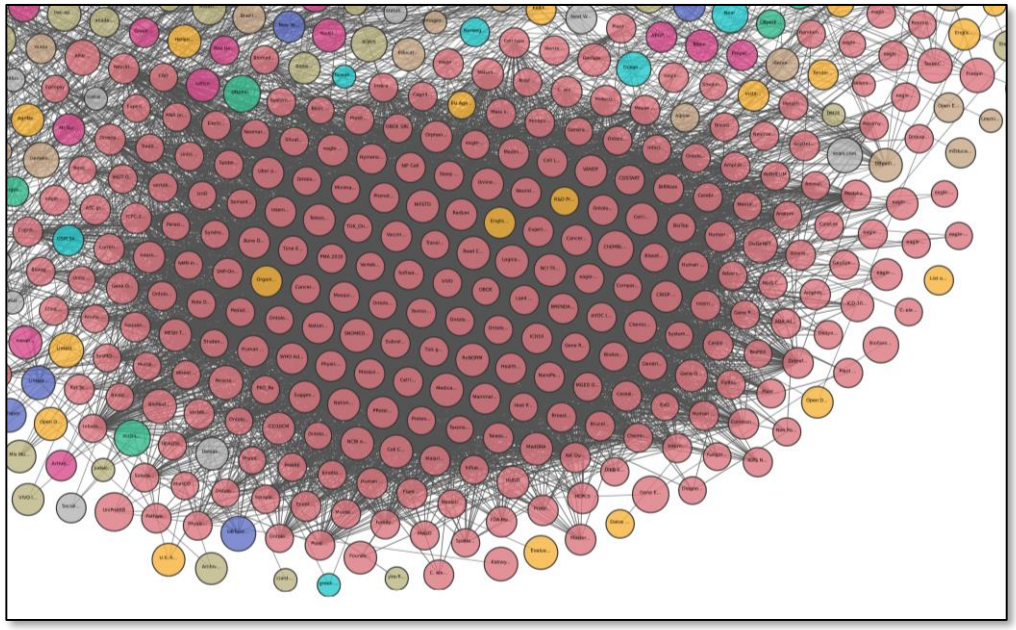
Maulik Kamdar, Tania Tudorache, Mark Musen

Stanford Center for Biomedical Informatics Research
maulikrk@stanford.edu

17th International Semantic Web Conference (ISWC) 2018

11th October 2018

Semantic web 8.6 (2017): 853-871

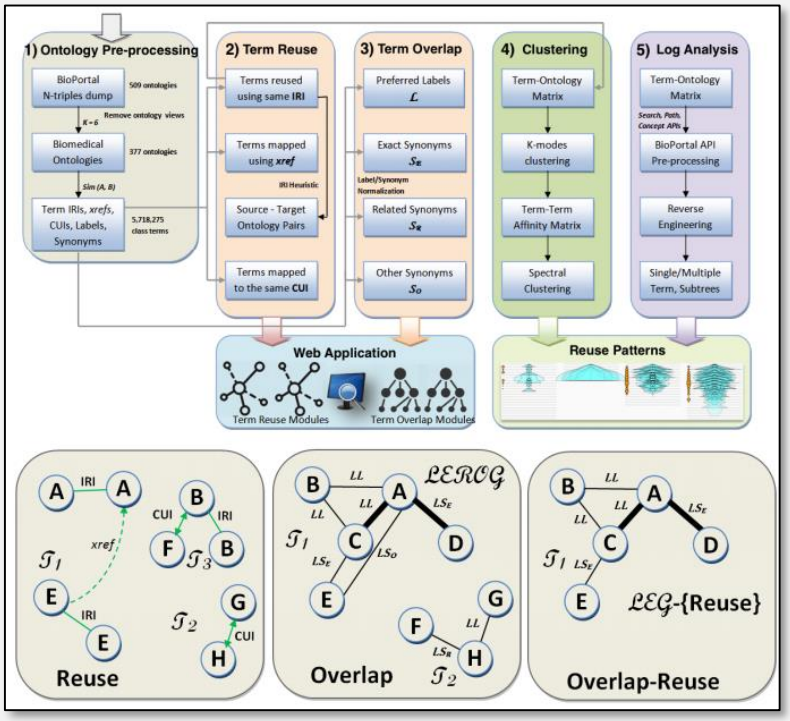


GO the Gene Ontology · efo · ChEBI · NCI thesaurus · SNOMED CT · FMA ontology

…

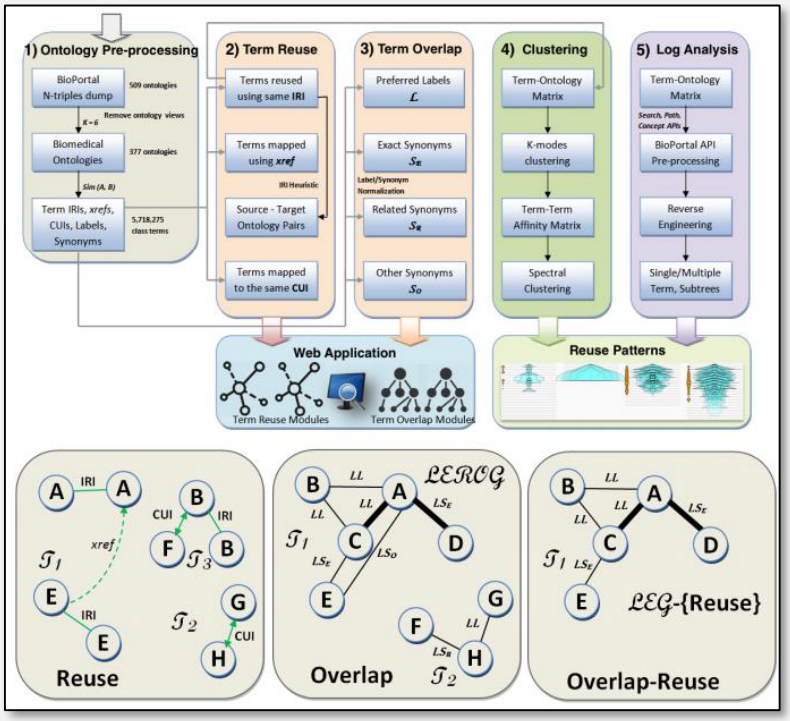
Uses: Knowledge management, decision support, semantic search, data annotation, data integration, reasoning …

What this talk is about …

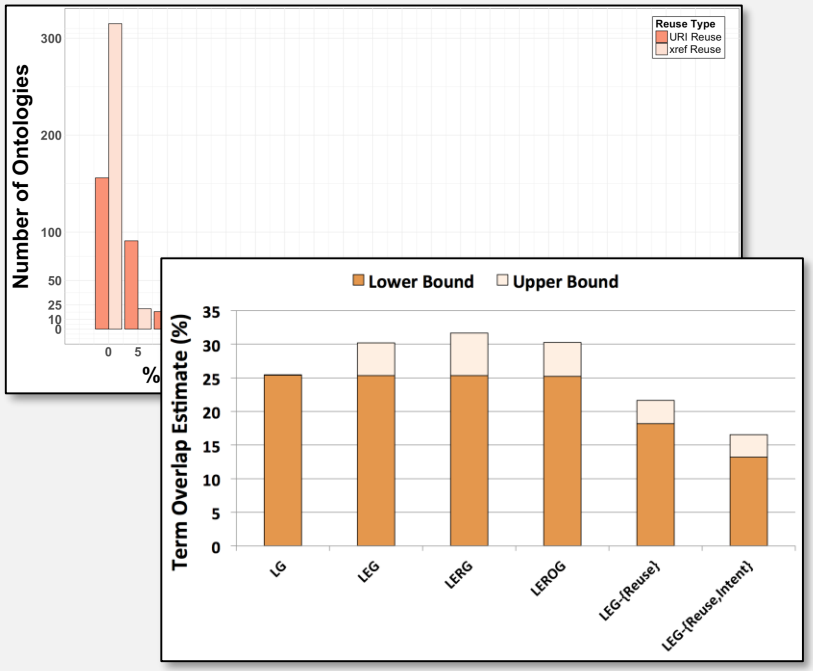


Methods to detect reuse and overlap across biomedical ontologies

What this talk is about …

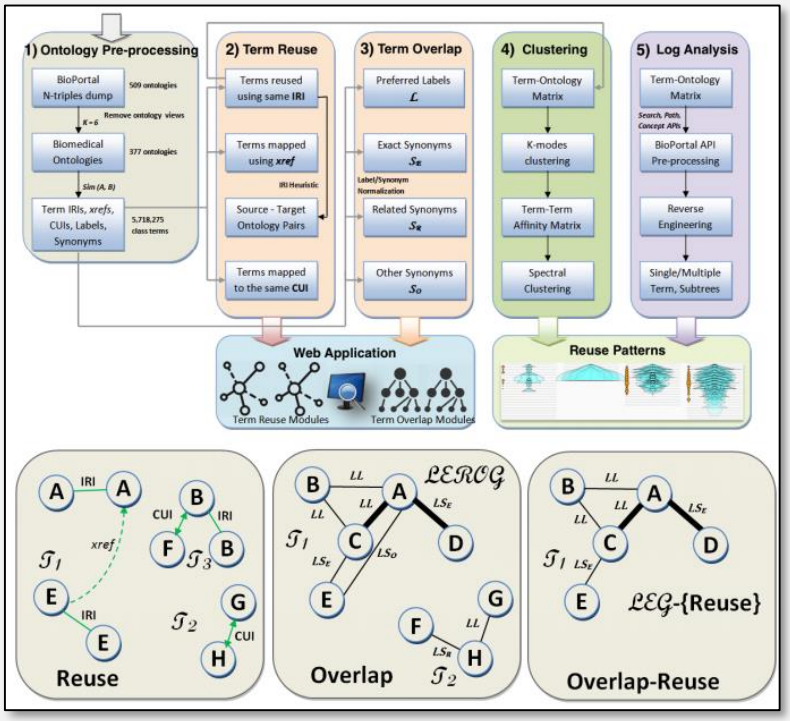


Methods to detect reuse and overlap across biomedical ontologies

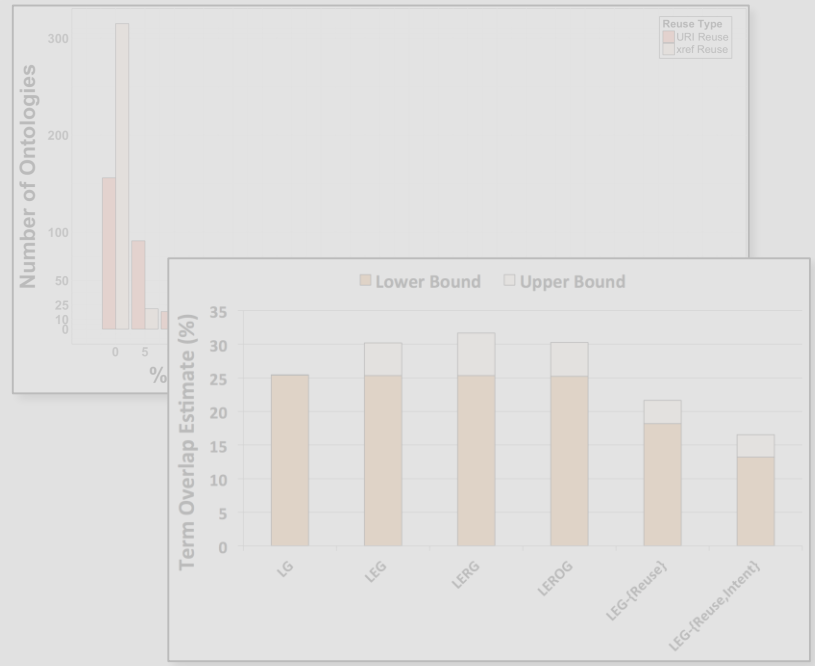


- ~9% Term Reuse
- Only popular or upper-level ontologies reused
- 25-31% Term Overlap

What this talk is about …

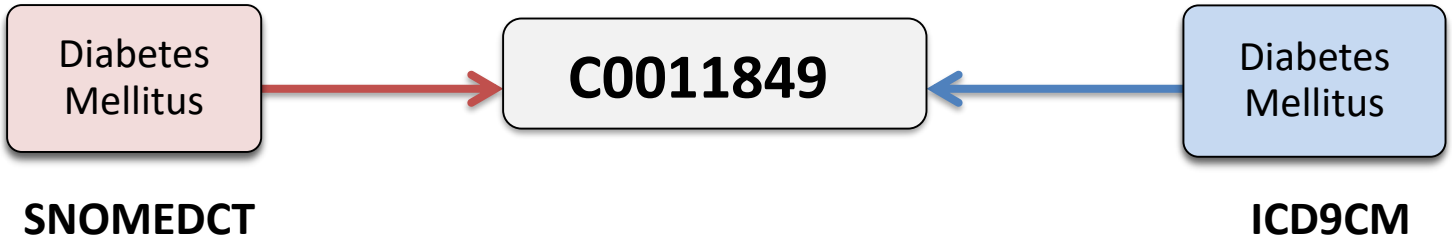


Methods to detect reuse and overlap across biomedical ontologies



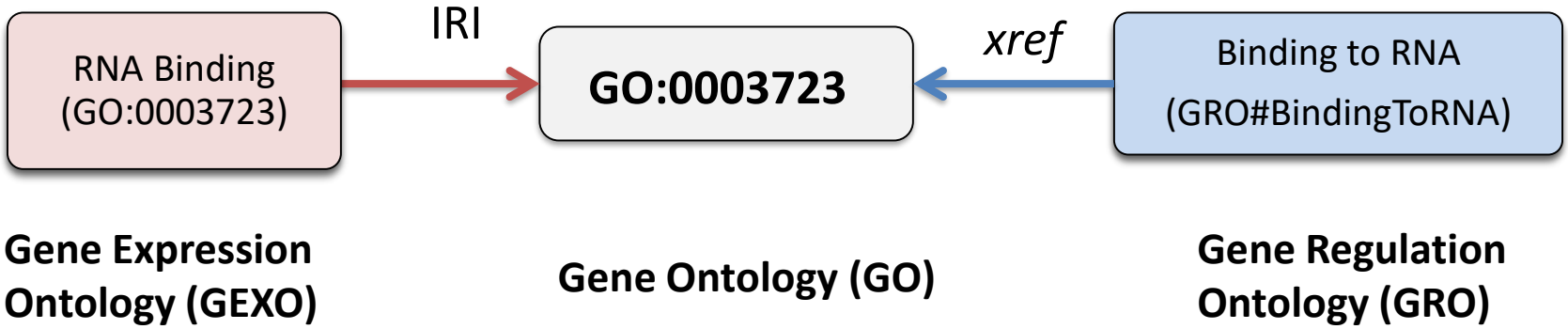
- ~9% Term Reuse
- Only popular or upper-level ontologies reused
- 25-31% Term Overlap

Unified Medical Language System (UMLS)

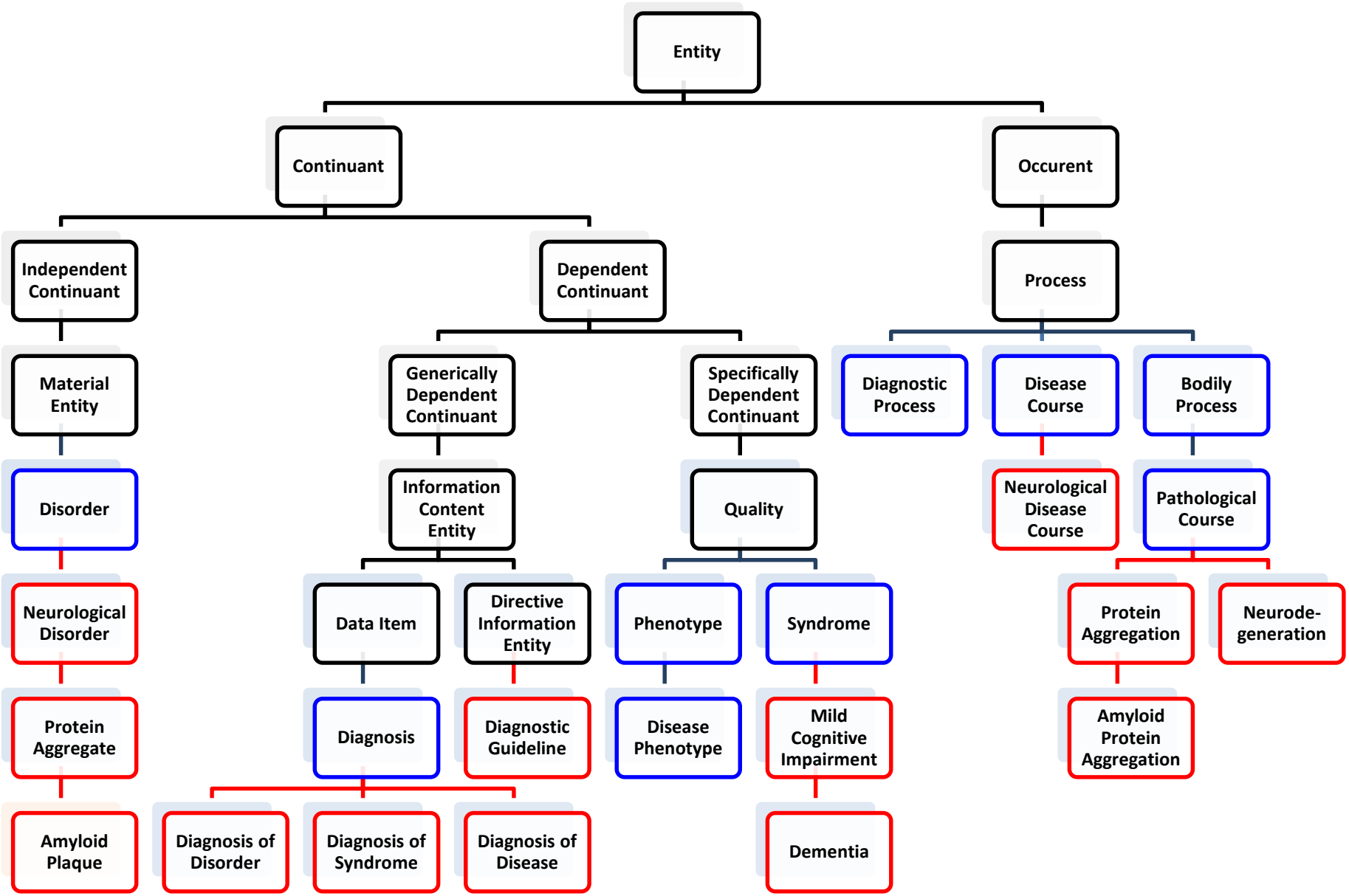


Unified Medical Language System (UMLS)

Open Biomedical Ontologies (OBO) Foundry

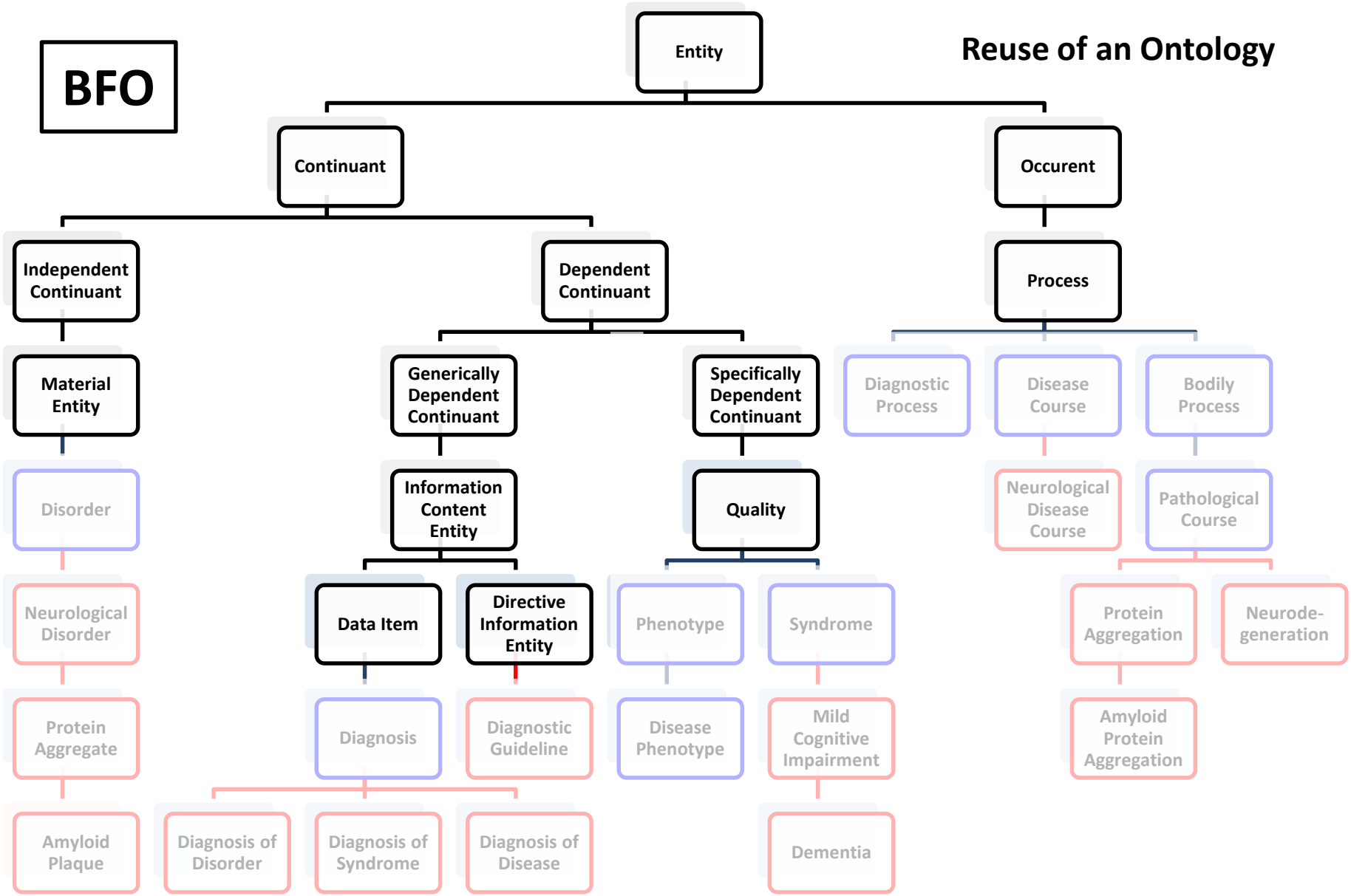


Neurological Disease Ontology



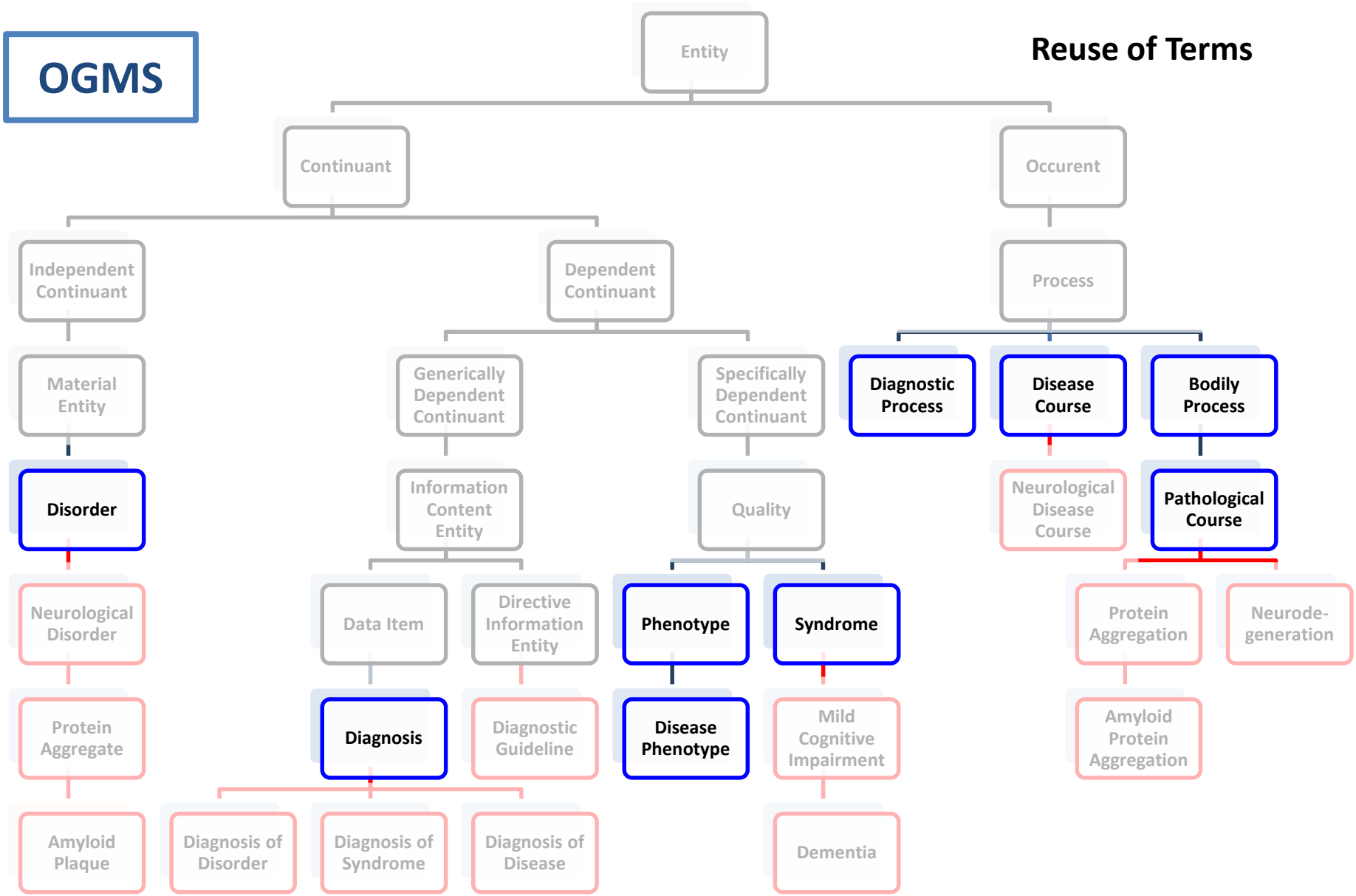
Ghazvinian, Amir, et al. "How orthogonal are the OBO Foundry ontologies?." J. Biomedical Semantics 2.S-2 (2011): S2.

Neurological Disease Ontology



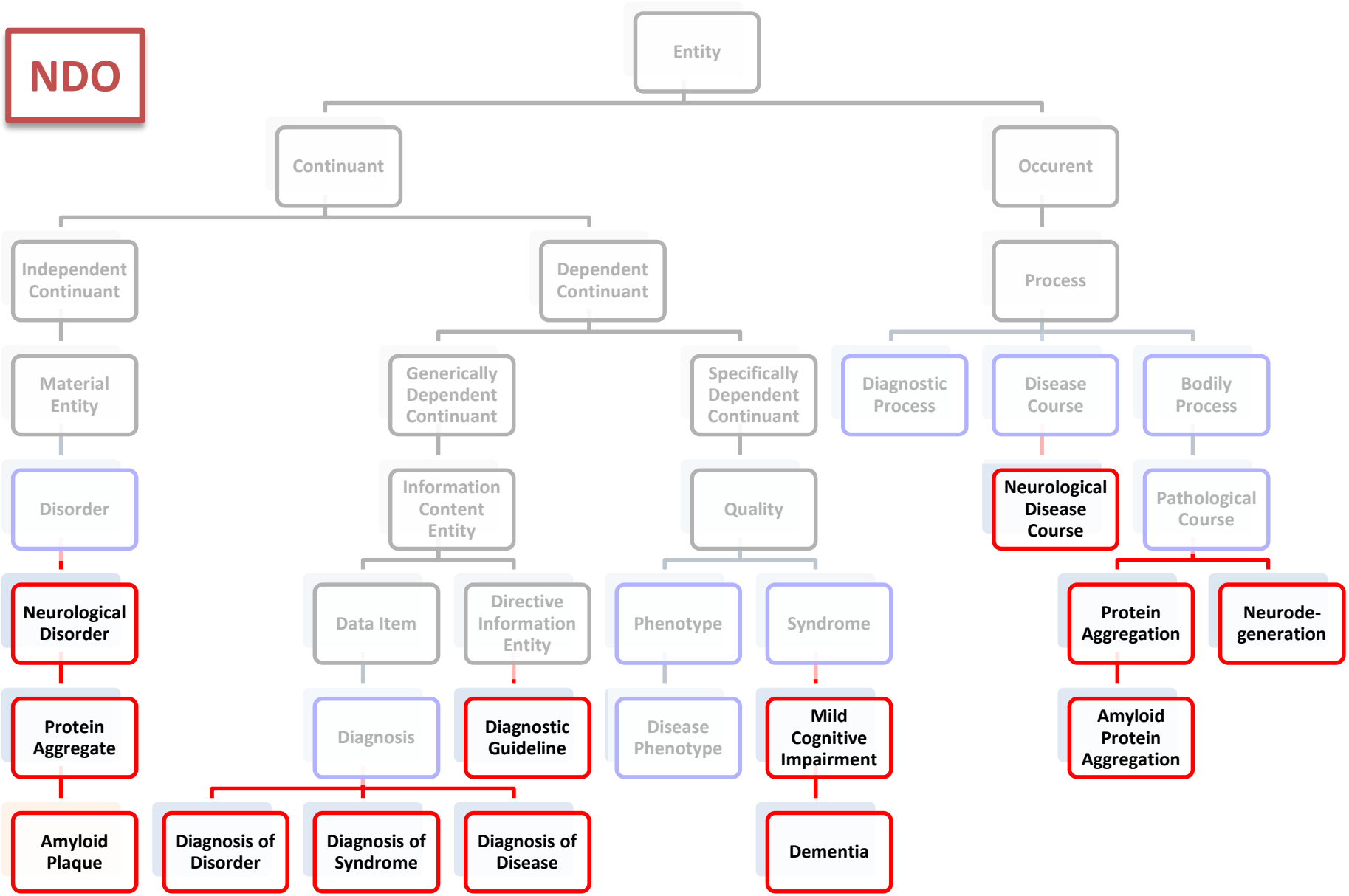
Ghazvinian, Amir, et al. "How orthogonal are the OBO Foundry ontologies?." J. Biomedical Semantics 2.S-2 (2011): S2.

Neurological Disease Ontology



Ghazvinian, Amir, et al. "How orthogonal are the OBO Foundry ontologies?." J. Biomedical Semantics 2.S-2 (2011): S2.

Neurological Disease Ontology



Ghazvinian, Amir, et al. "How orthogonal are the OBO Foundry ontologies?." J. Biomedical Semantics 2.S-2 (2011): S2.

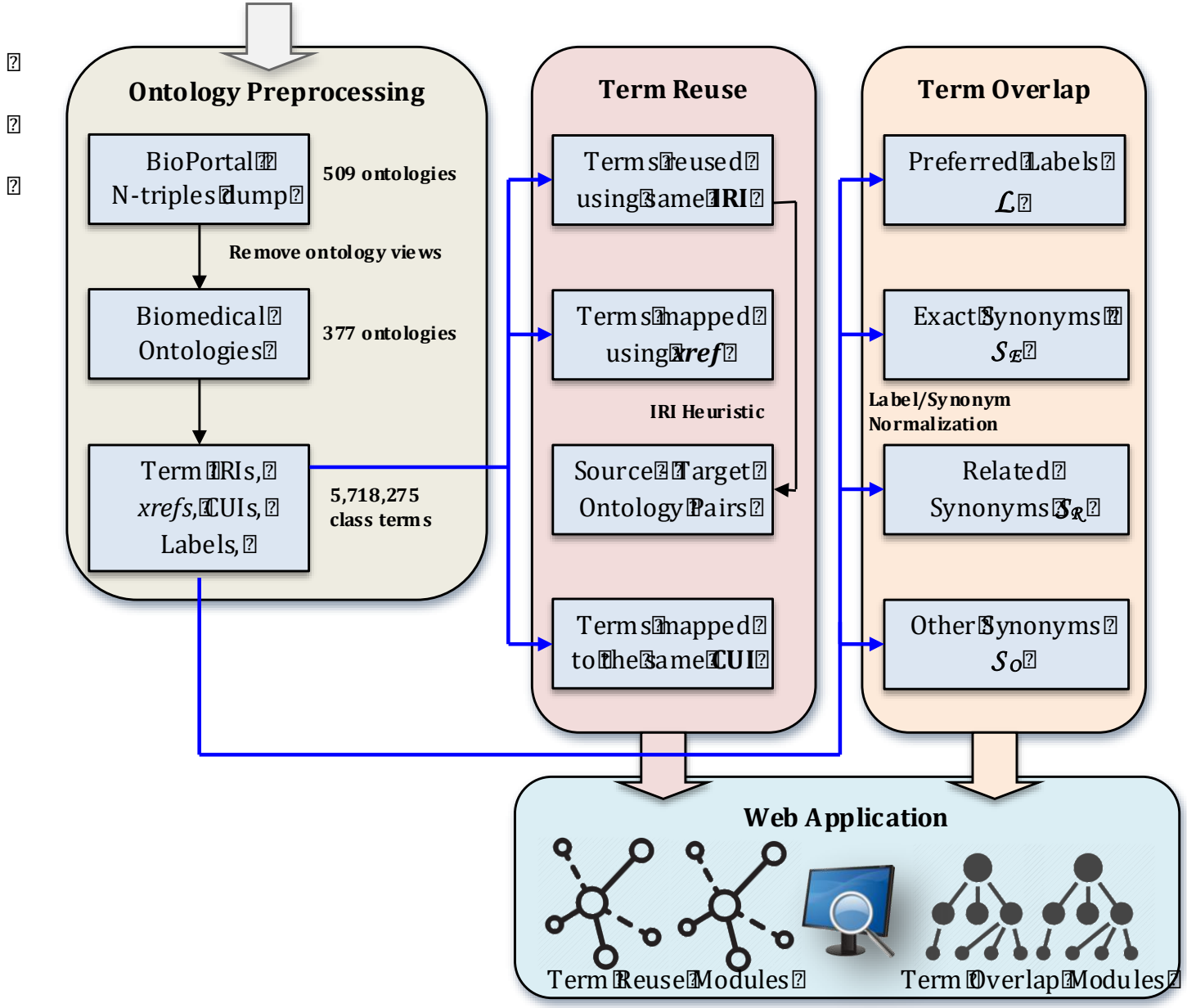
Term Reuse and Term Overlap

- **Term Reuse:** The same term is present in two or more ontologies either by direct use of the same IRI, or via explicit references or mappings.

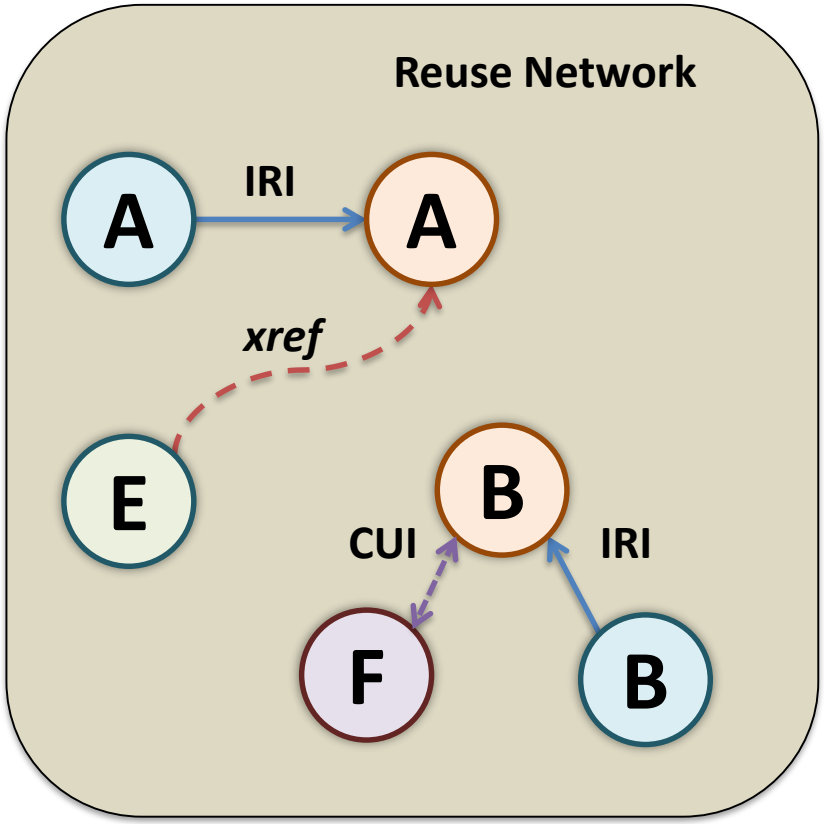
- **Term Overlap:** Two terms in different ontologies are similar, when compared using their labels or synonyms.

- **Term Overlap – Term Reuse**: Two terms in different ontologies are similar but do not have same IRI or explicit references or mappings.

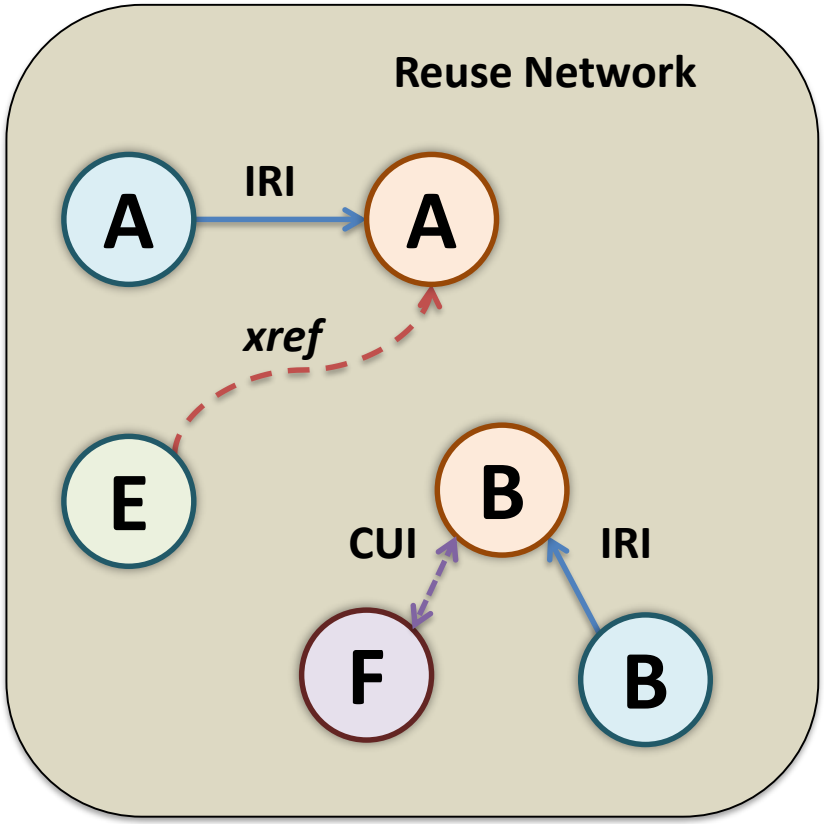
Reuse and Overlap across Biomedical Ontologies

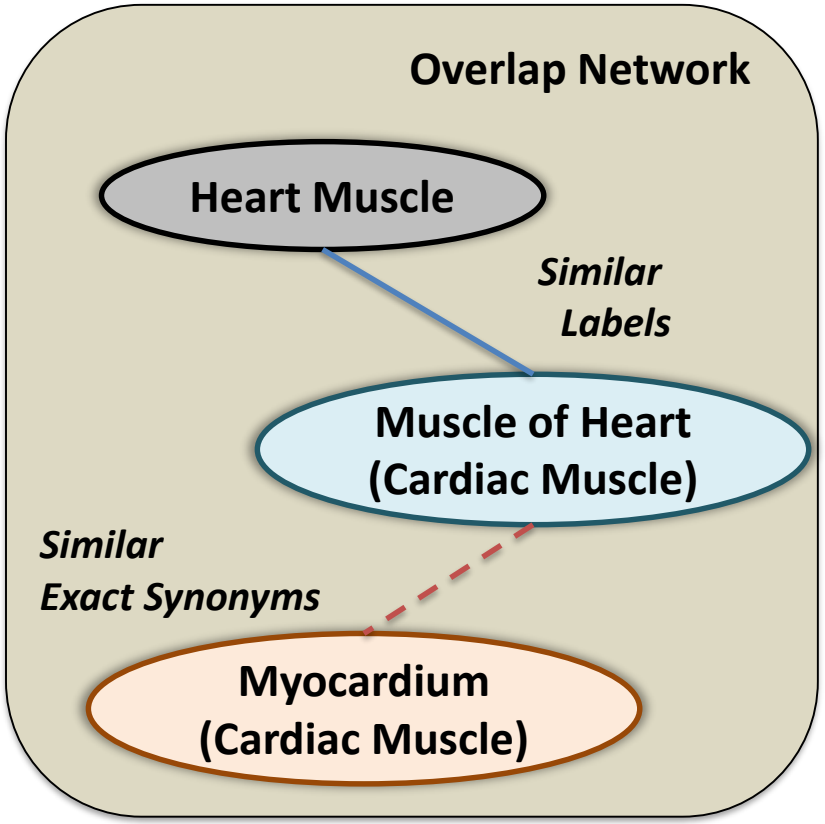


Reuse and Overlap Modules



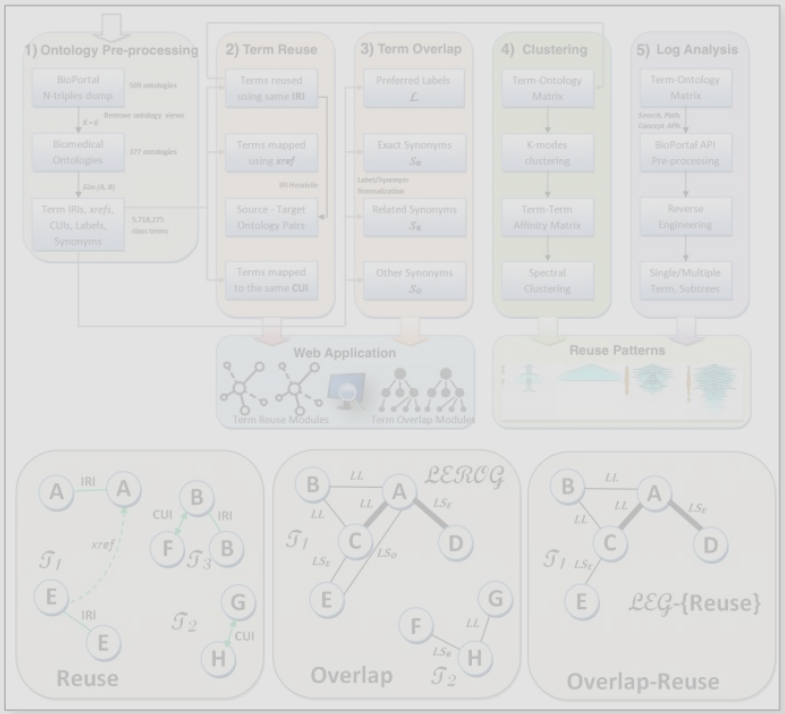
Reuse and Overlap Modules



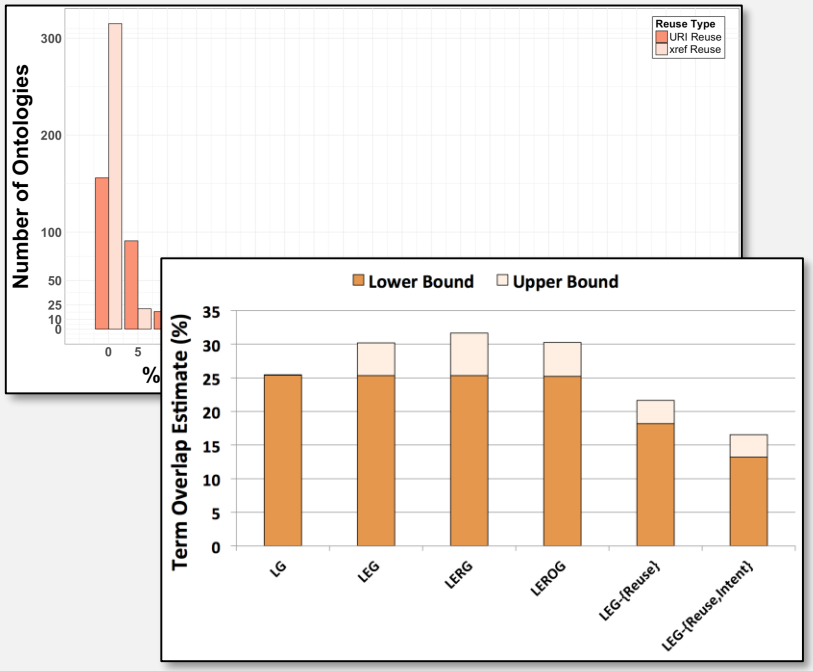


Different "Modules" can be extracted from these networks depending on the types of edges.

What this talk is about …

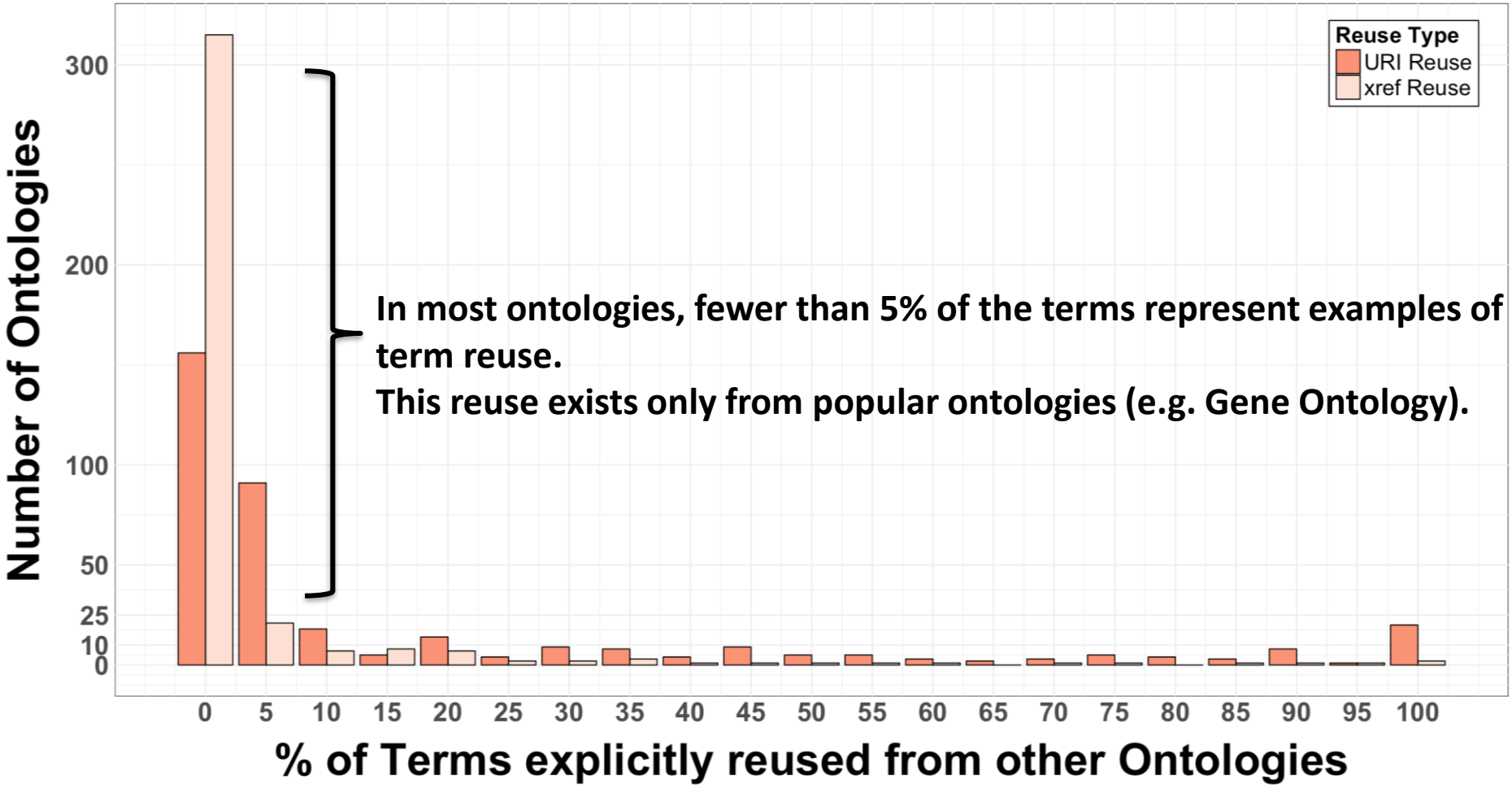


Methods to detect reuse and overlap across biomedical ontologies

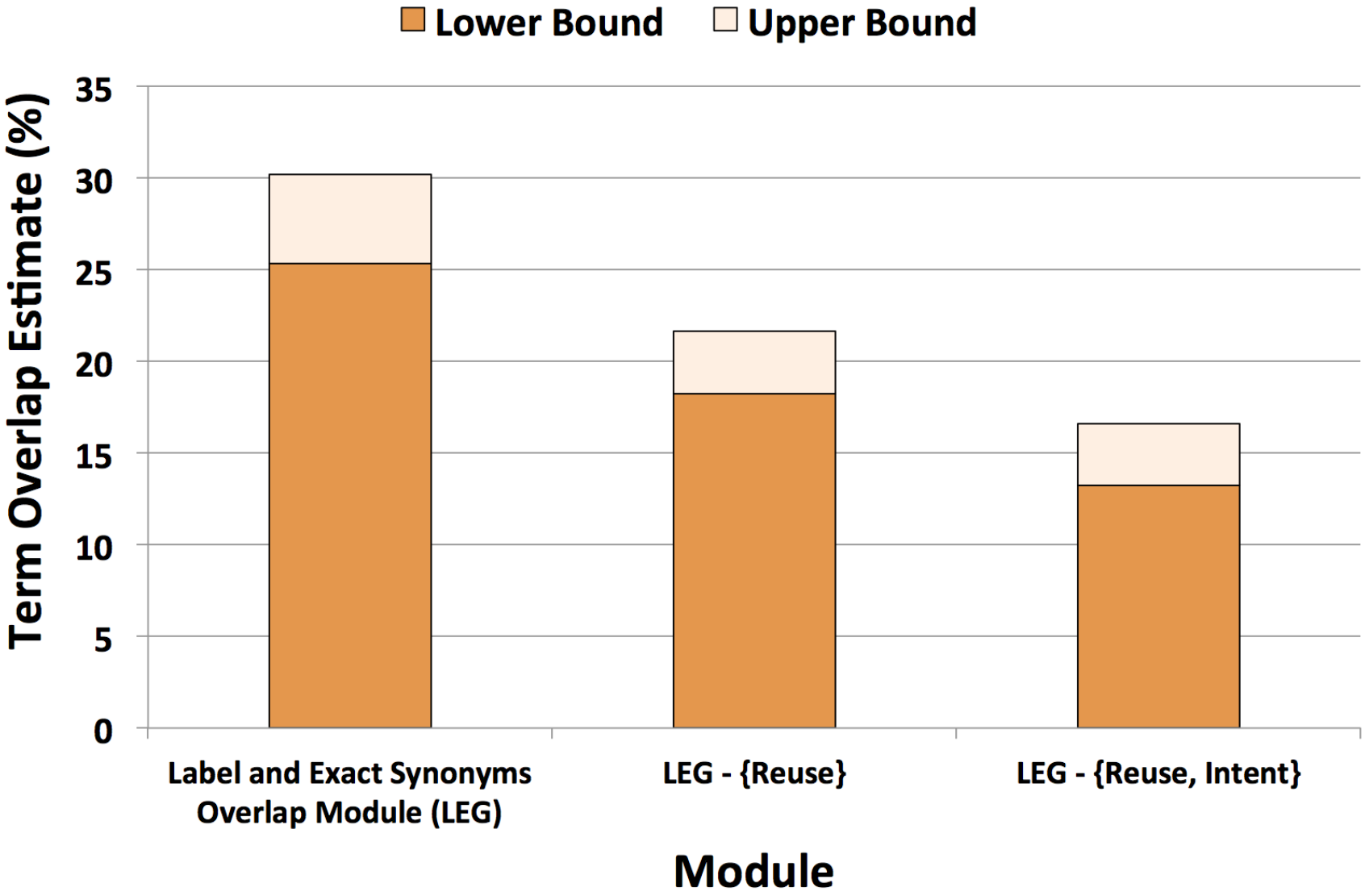


- ~9% Term Reuse
- Only popular or upper-level ontologies reused
- 25-31% Term Overlap

Minimal Reuse across biomedical Ontologies



25-31% Term Overlap exists across BioPortal



Ontology Engineers show an intent for reuse!

Ontology Engineers show an intent for reuse!

Different Versions

- **Version 1.0 / Version1.1 of Basic Formal Ontology (BFO)**

Ontology Engineers show an intent for reuse!

Different Versions

- **Version 1.0 / Version1.1 of Basic Formal Ontology (BFO)**

Different Notations

- **OBO:FMA_31396**
- OBO:owlapi/fma#FMA_31396
- OBO:owl/FMA#FMA_31396
- OBO:fma#Cartilage_of_inferior …

Ontology Engineers show an intent for reuse!

Different Versions

- **Version 1.0 / Version1.1 of Basic Formal Ontology (BFO)**

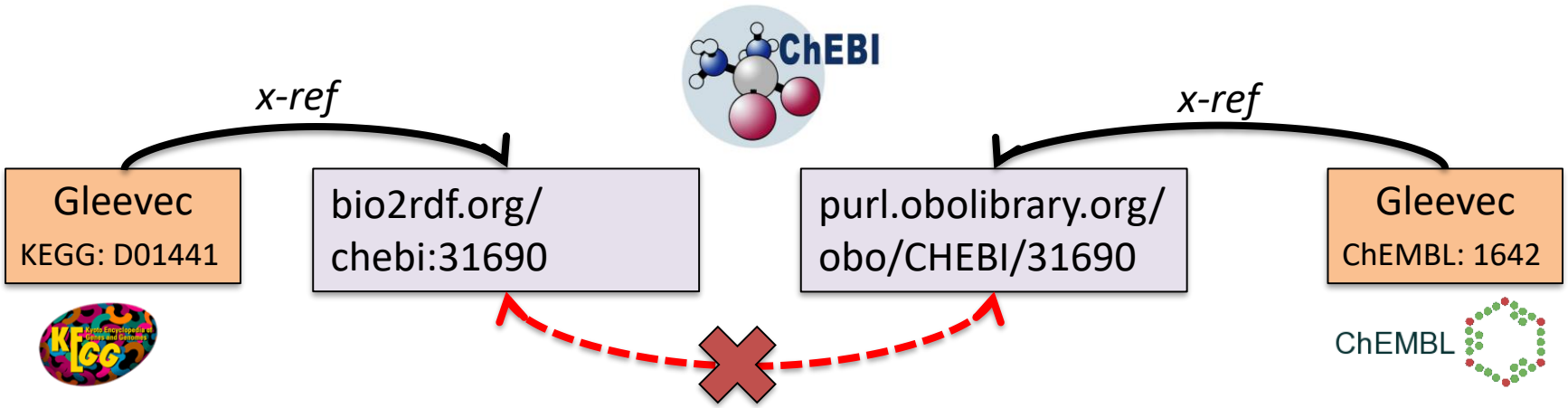
Different Notations

- **OBO:FMA_31396**
- OBO:owlapi/fma#FMA_31396
- OBO:owl/FMA#FMA_31396
- OBO:fma#Cartilage_of_inferior …

Different Domains

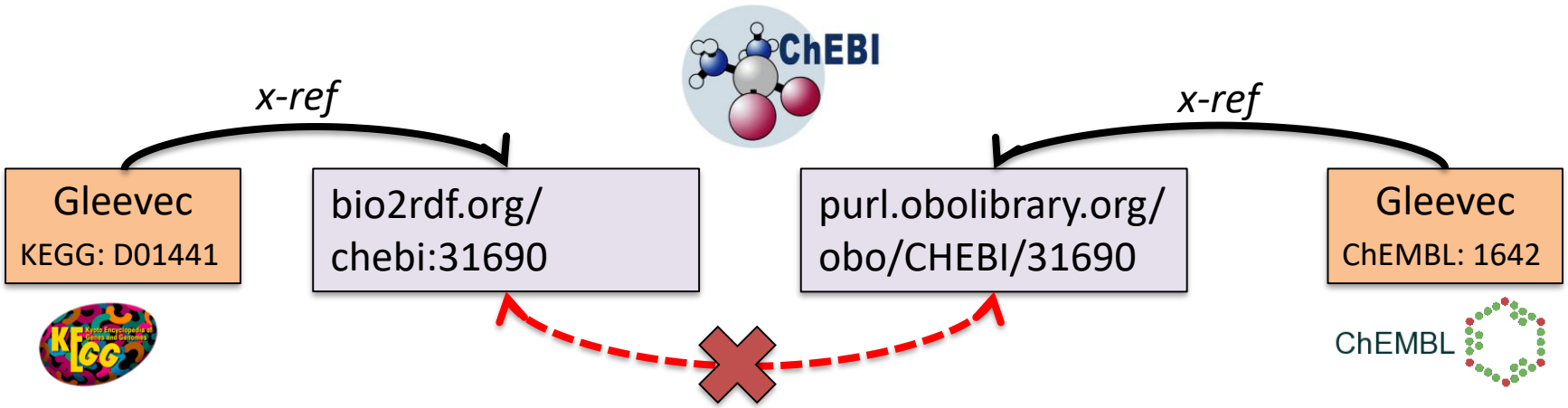
- http://ihtsdo.org/snomedct/
- http://purl.bioontology.org/ontology/SNOMEDCT

Intent for Reuse: Use of similar, but not actual URIs/IRIs



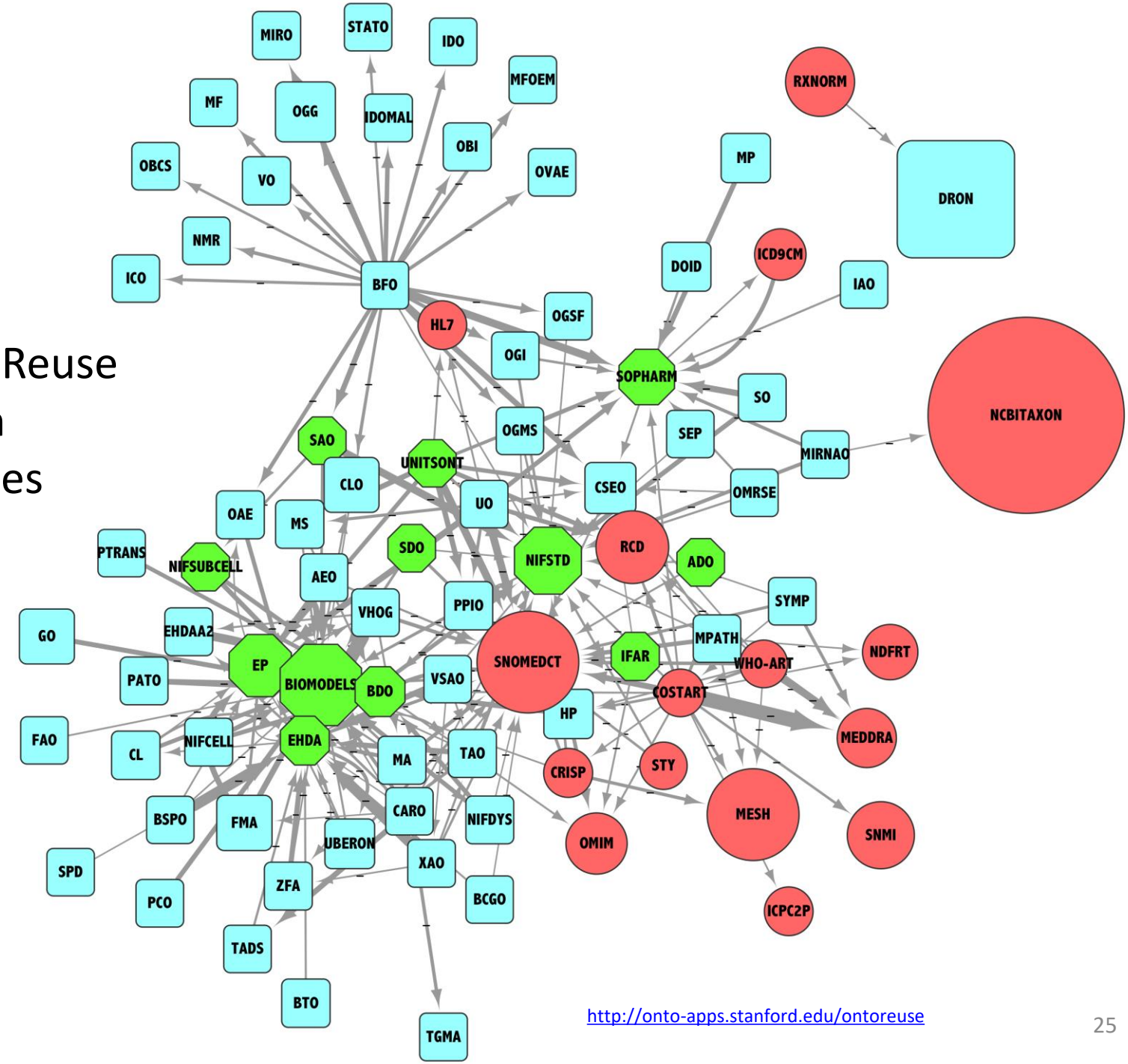
Retrieve Assays which identify potential Chemopreventive Agents that have *Molecular Weight < 300 g/mol* and *target* Estrogen Receptors present in Human

Intent for Reuse: Use of similar, but not actual URIs/IRIs



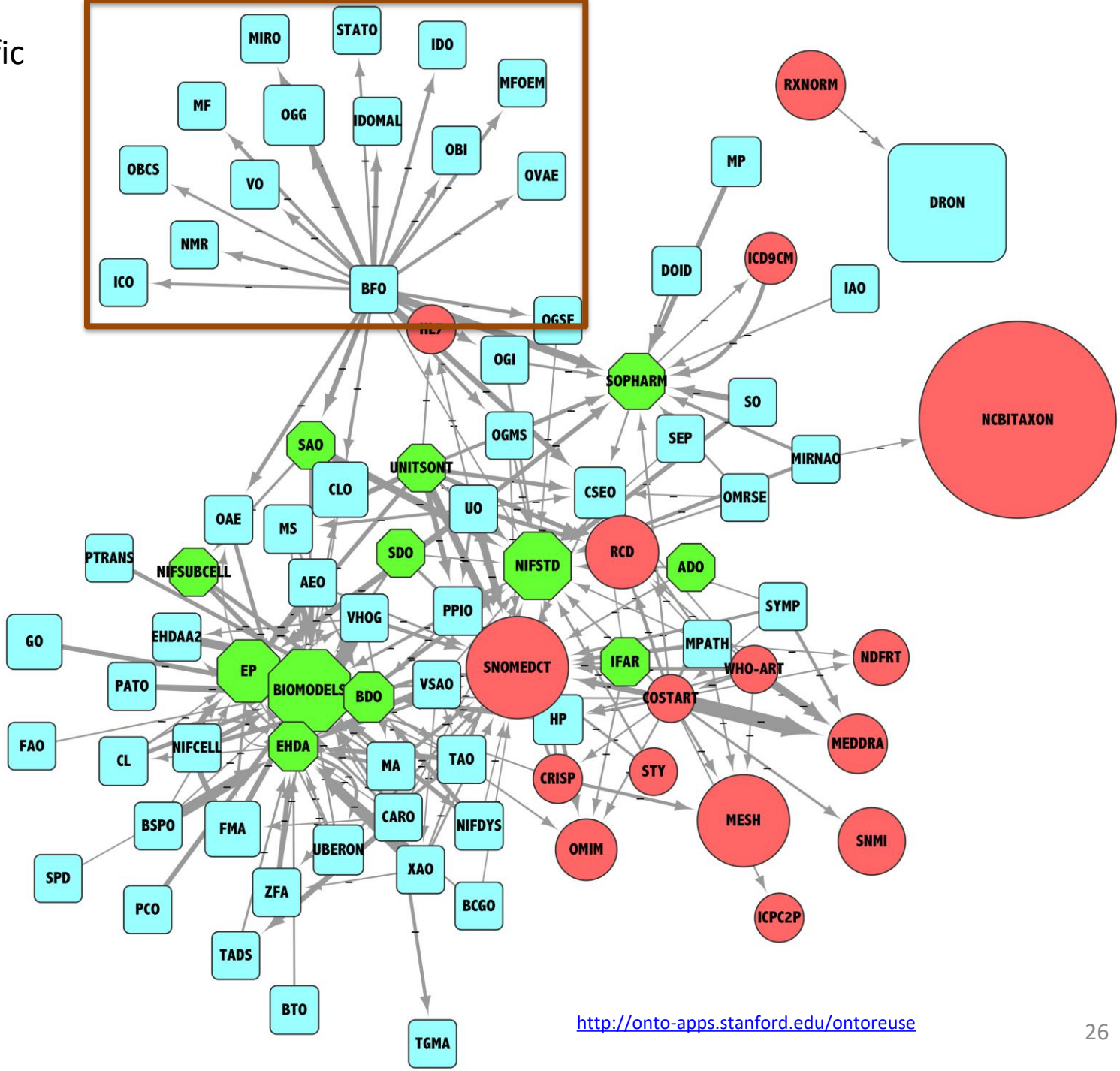
Origin	IRI Representation	"Reusing" Sources
ChEBI	purl.obolibrary.org/obo/CHEBI/31690 identifiers.org/chebi/CHEBI:31690 www.ebi.ac.uk/chebi/31690 bio2rdf.org/chebi:31690	PubChem, ExpresssionAtlas WikiPathways, PDB BioPax, ChEMBL Bio2RDF

Overlap-Reuse between Ontologies

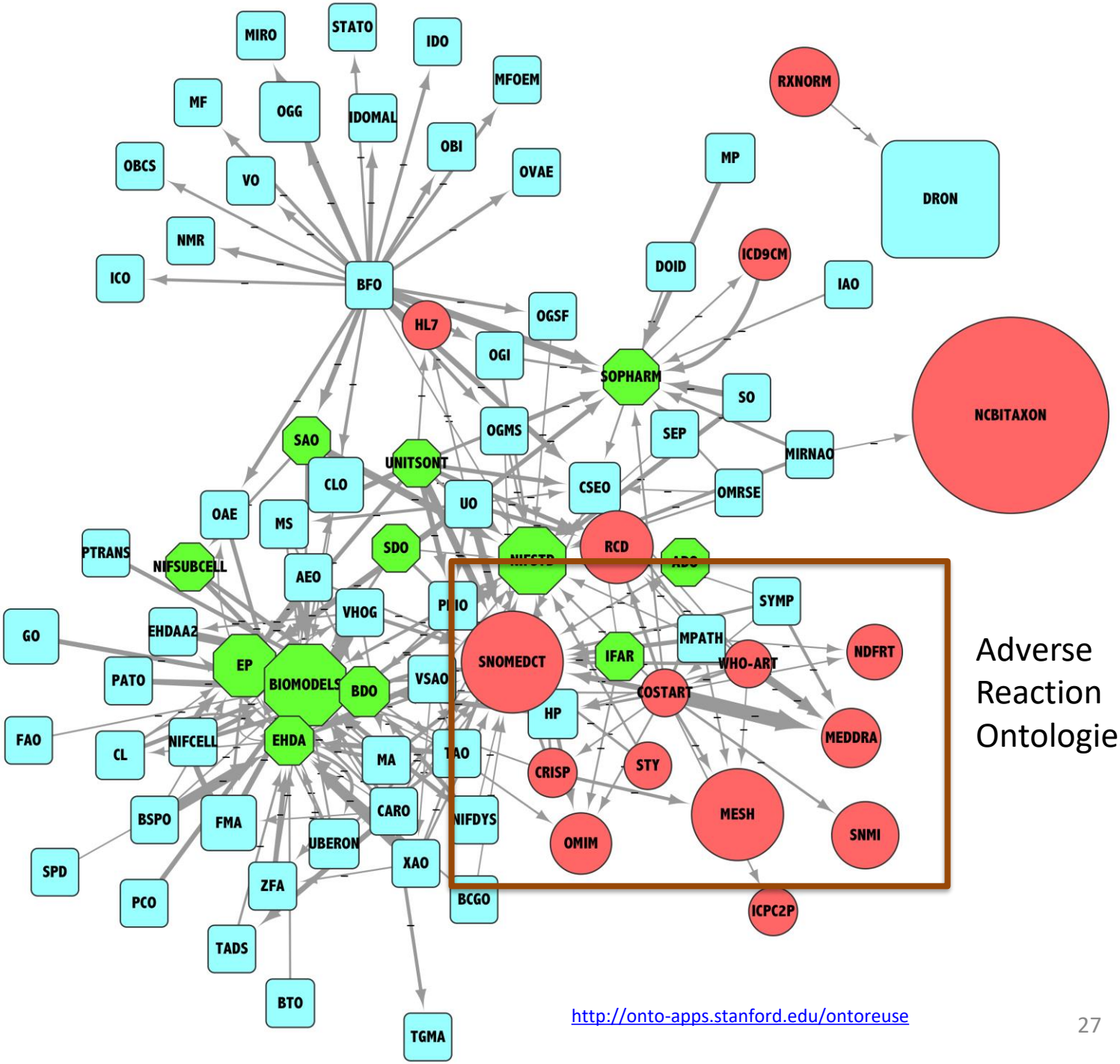


http://onto-apps.stanford.edu/ontoreuse

BFO-specific Overlap

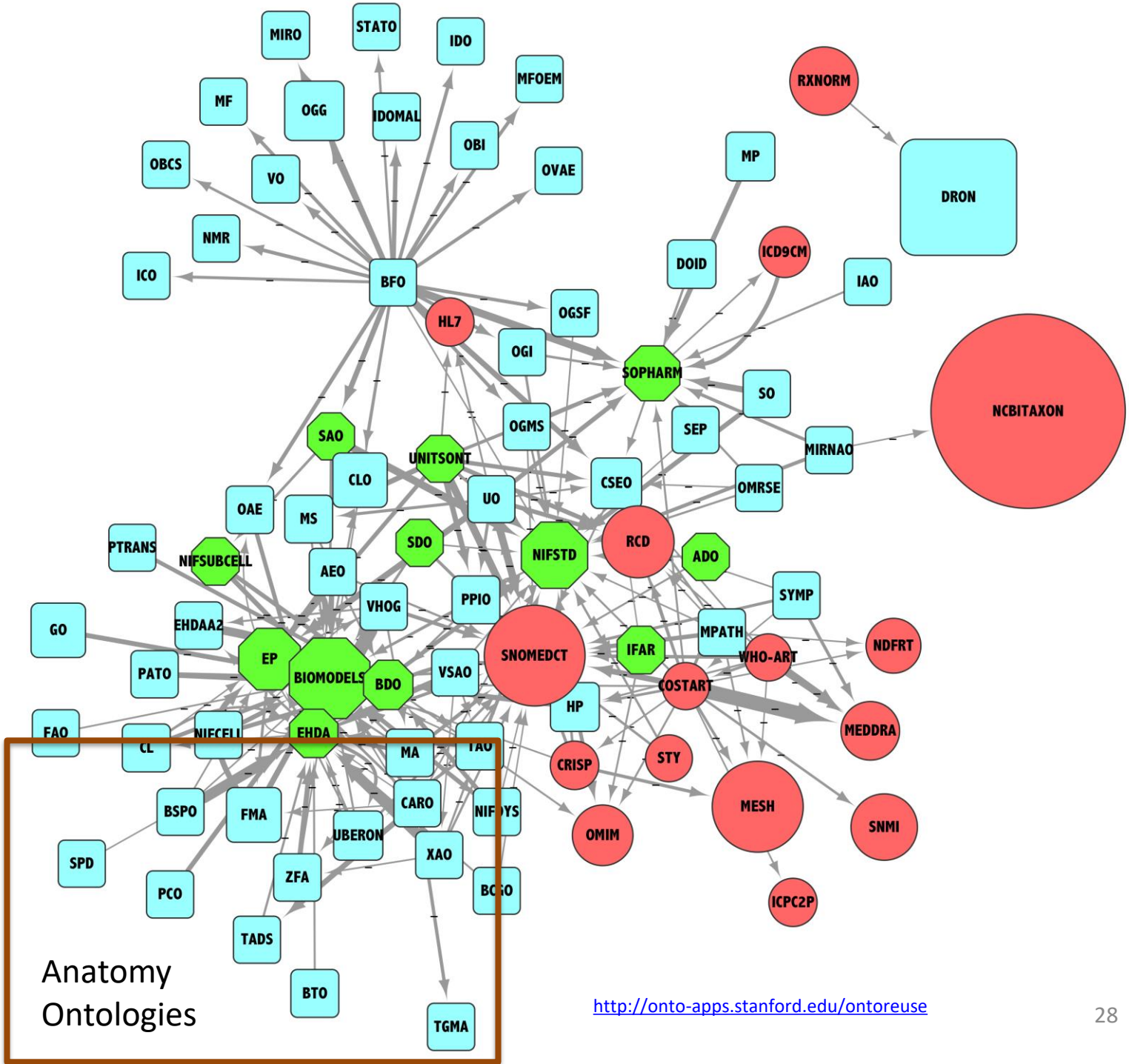


http://onto-apps.stanford.edu/ontoreuse



Adverse Reaction Ontologies

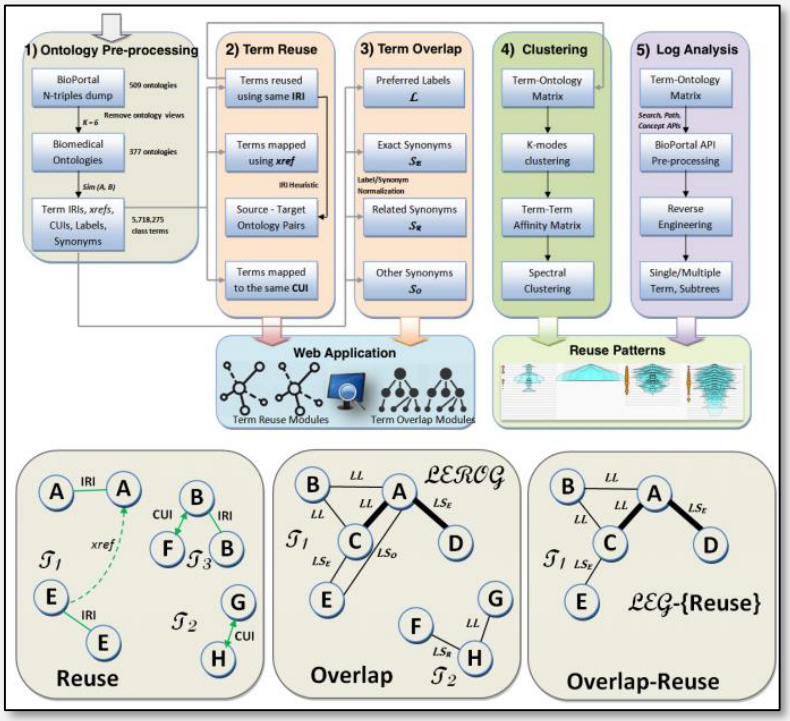
http://onto-apps.stanford.edu/ontoreuse



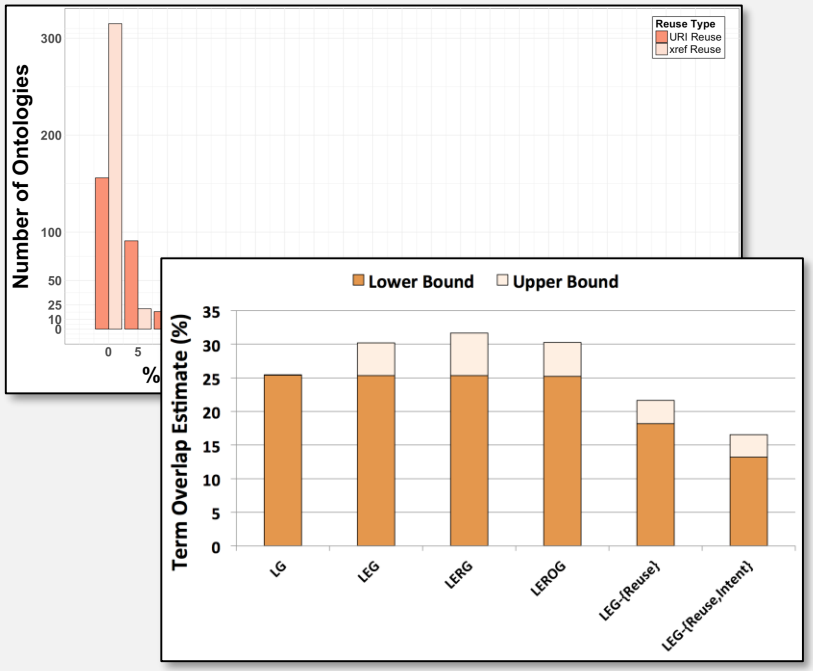
Anatomy Ontologies

http://onto-apps.stanford.edu/ontoreuse

This talk was about …



Methods to detect reuse and overlap across biomedical ontologies



- ~9% Term Reuse
- Only popular or upper-level ontologies reused
- 25-31% Term Overlap

Sad State of Affairs …

- Still far from achieving ideal term reuse, beyond upper level and popular ontologies
- Newer ontologies are always added in BioPortal
  - 700+ ontologies and counting …
- Incorrect representations without mappings do not provide advantages of Term Reuse!
- Without proper guidelines and semi-automated tools, the semantic chasm will widen!

Welcome to BioPortal, the world's most comprehensive repository of biomedical ontologies

Search for a class

Enter a class, e.g. Melanoma

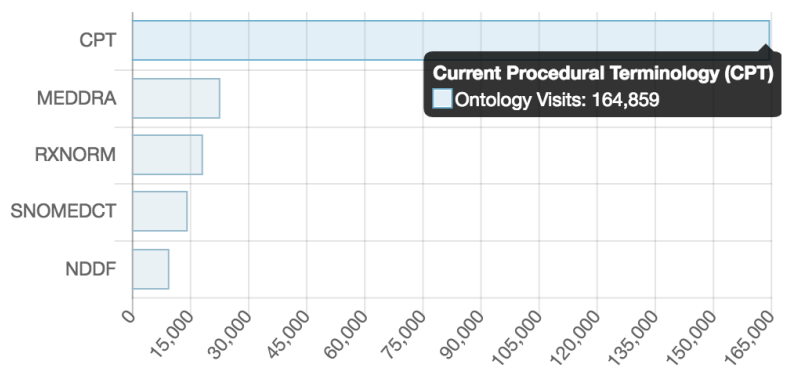
Advanced Search

Find an ontology

Start typing ontology name, then choose from list

Browse Ontologies

Ontology Visits (September 2018)



More

BioPortal Statistics

Ontologies	730
Classes	9,594,701
Resources Indexed	48
Indexed Records	39,537,360
Direct Annotations	95,468,433,792
Direct Plus Expanded Annotations	144,789,582,932

THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

maulikrk@stanford.edu

http://onto-apps.stanford.edu/ontoreuse

bmi