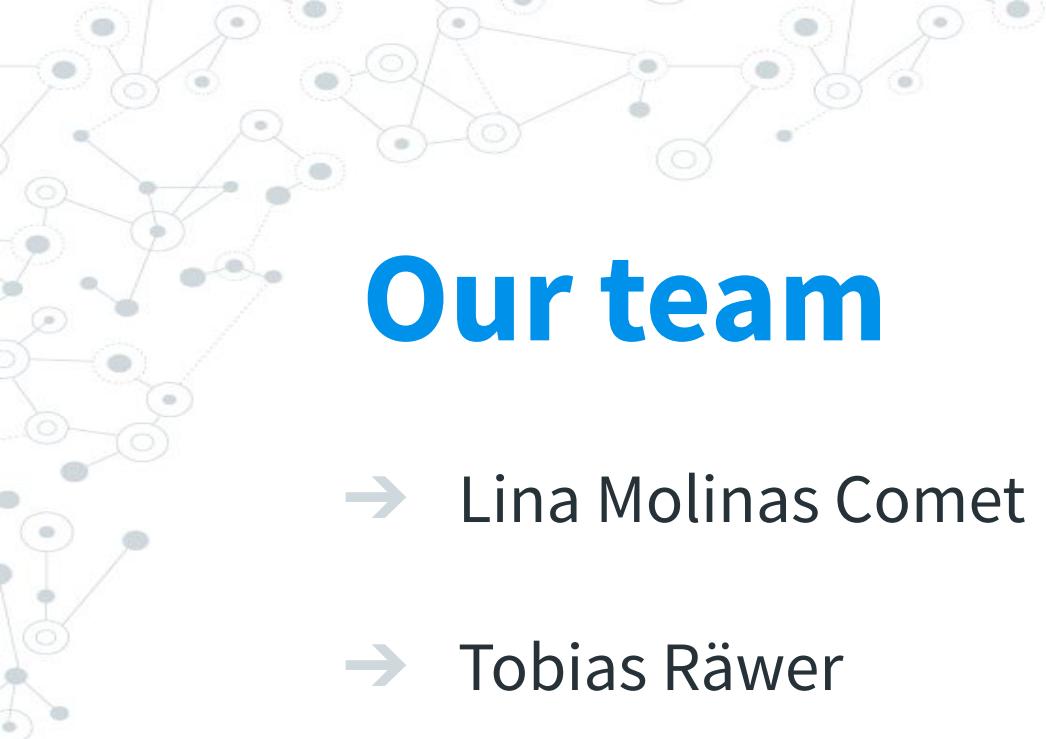


Ontology Matching

Knowledge Graphs Lab
15.07.2019





Our team

- Lina Molinas Comet
- Tobias Räwer
- Jonas Rülfing

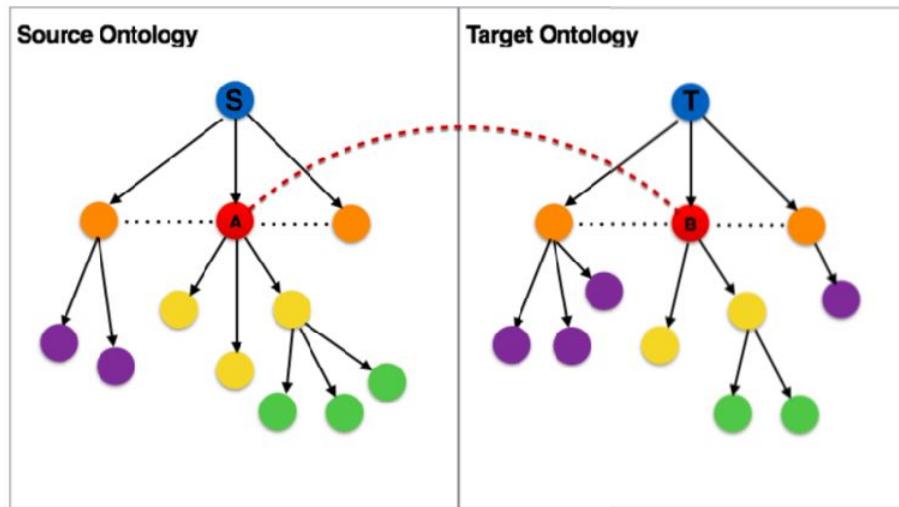


Agenda

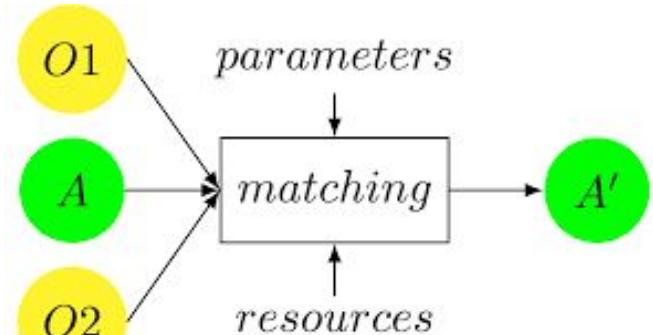
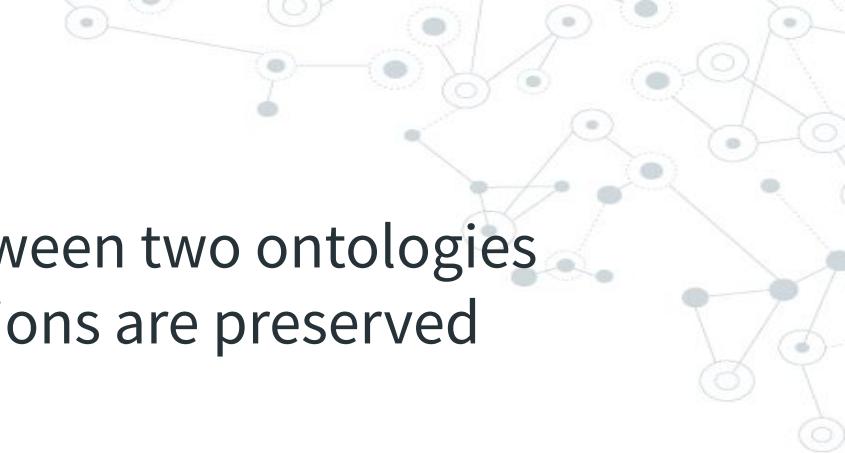
- Motivation
 - Main concepts (Recap)
 - Problem Definition
- Our Proposal
- Methods
 - Required Steps
- Evaluation
- Conclusion & Outlook

Ontology Matching (Recap)

Ontology matching is the relation between two ontologies where the concepts, roles and restrictions are preserved



Interactive Visualization Of Large Ontology Matching Results (Li & Stroe & Cruz)



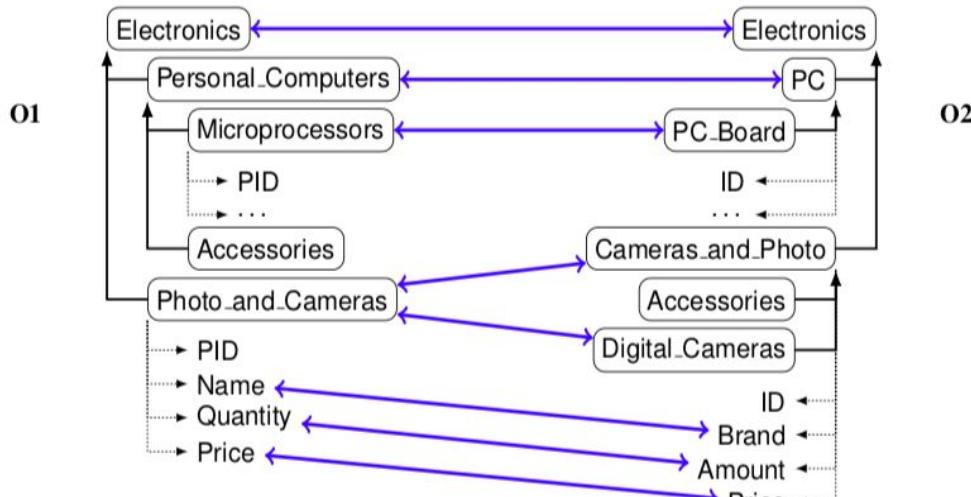
Ontology Matching: State of the Art and Future Challenges
(Shvaiko & Euzenat)

Why do we need Ontology Matching ? (Recap)

Because there is not a unified vocabulary/ontology:

- Partially overlap among ontologies
- Each ontology represents a own perspective
- Different languages

We need
Interoperability of
Data and Knowledge



Ten Challenges for Ontology Matching(Shvaiko & Euzenat)

Some problems in Ontology Matching (Recap)

| Techniques | Problems |
|---|---|
| String matching | Too many false positive results (based just on entity name/label) |
| String matching + logical reasoning | Reasoning highly depend on correctness of the expressions in the ontologies |
| Blocking (subdividing entities into mutually exclusive blocks) | Unavailability of expertise in every domain Rules are not well specified Not enough training data (supervised learning) |

BioPortal

BioPortal is an open repository of biomedical ontologies.

And it provides:

- a service to access them
- a set of tools to work with them

The screenshot shows the BioPortal homepage with a dark blue header containing the BioPortal logo, navigation links (Ontologies, Search, Annotator, Recommender, Mappings, Resource Index), and a 'Login' button. Below the header is a search bar with placeholder text 'Search...'. The main area is titled 'Browse' and features a list of ontologies. Each entry includes the ontology name, a brief description, the date it was uploaded (4/30/19), and three small boxes indicating the number of notes, projects, and datasets. The entries shown are:

- Current Procedural Terminology (CPT)**
Current Procedural Terminology
Uploaded: 4/30/19
Notes: 1, Projects: 1, Datasets: 13,996
- Medical Dictionary for Regulatory Activities Terminology (MedDRA) (MEDDRA)**
MedDRA is an international medical terminology with an emphasis on use for data entry, retrieval, analysis, and display
Uploaded: 4/30/19
Notes: 1, Projects: 10, Datasets: 71,982
- SNOMED CT (SNOMEDCT)**
SNOMED Clinical Terms
Uploaded: 4/30/19

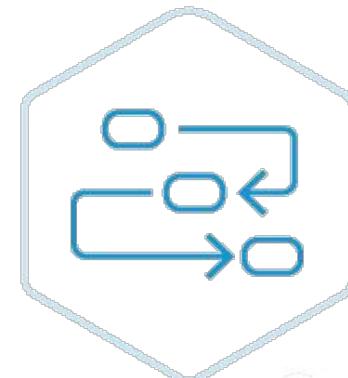
Ontology Mappings in the BioPortal

Lexical OWL Ontology Matcher (LOOM):

- Simple lexical algorithm
- Compares preferred names and synonyms of concepts in both ontologies

Manual integrated mappings:

- Added by domain experts
- Possible by adding through the Web interface or Rest API



Problem definition

There are not quality measures of the ontology matches among entities in the BioPortal

How do we measure how suitable are those matches?



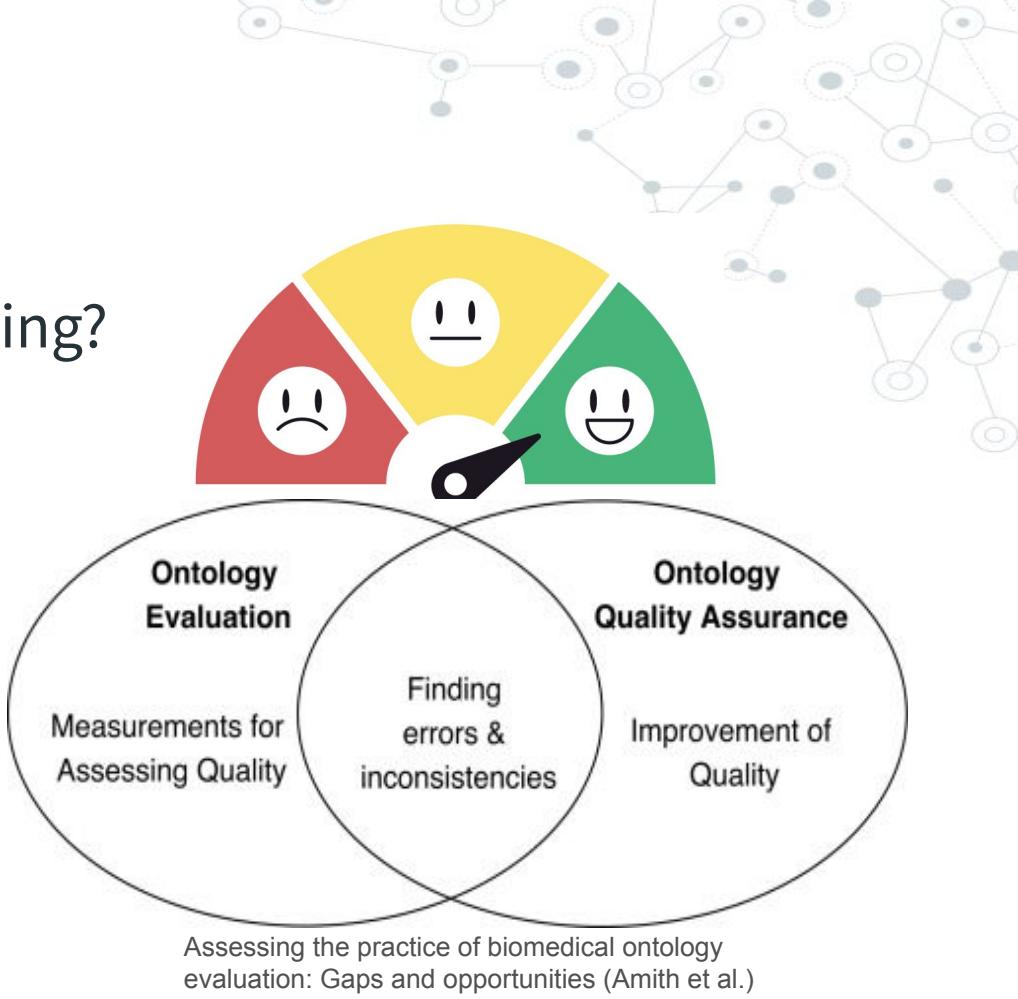
QUALITY MEASURE

Metrics ?

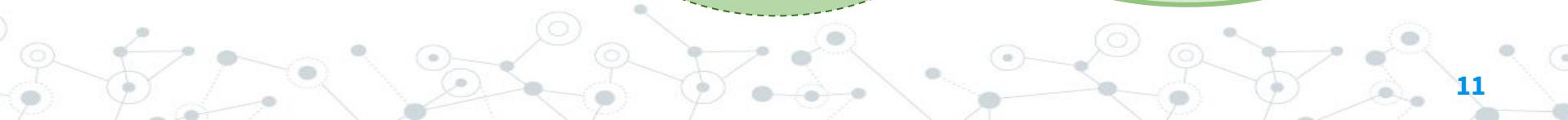
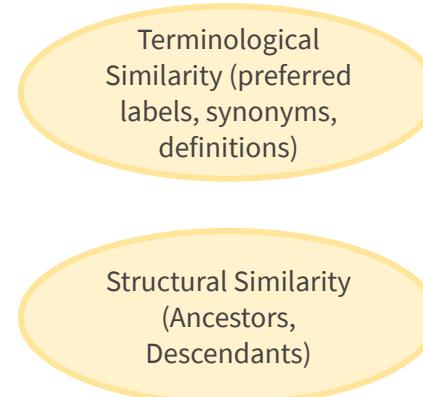
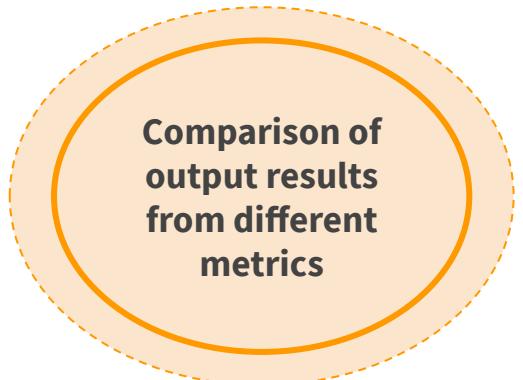
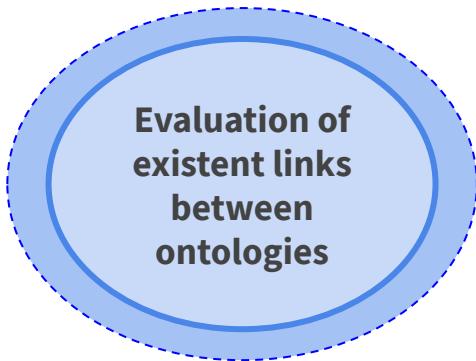
| | |
|------------------------------------|--------|
| Classes | 14,020 |
| Individuals | 0 |
| Properties | 8 |
| Maximum depth | 15 |
| Maximum number of children | 45 |
| Average number of children | 3 |
| Classes with a single child | 1,467 |
| Classes with more than 25 children | 26 |
| Classes with no definition | 858 |

Mappings Evaluation

- How good is one matching?
- How good is the overall mapping?
- Are there loops among the mappings?



Our proposal



Our steps to reach our goal



- ◎ **Selection of a subset of ontologies**
- ◎ **Access data** by using the BioPortal REST API
- ◎ **Evaluation of existing mappings**
 - Try out different metrics
 - Evaluate and refine the used metrics against the Silver Mappings provided from Pistoia

Selected Ontologies



Human Phenotype Ontology

Last uploaded: June 3, 2019

[Summary](#) [Classes](#) [Properties](#) [Notes](#) [Mapping](#)

Jump to:

- ↳ All
 - ↳ Blood group
 - ↳ Clinical course
 - ↳ Clinical modifier
 - ↳ Frequency
 - ↳ Mode of inheritance
 - ↳ Past medical history
 - ↳ Phenotypic abnormality
 - ↳ Abnormal cellular phenotype
 - ↳ Abnormality of blood and blood-forming tissue
 - ↳ Abnormality of connective tissue
 - ↳ Abnormality of head or neck
 - ↳ Abnormality of limbs
 - ↳ Abnormality of metabolism/homeostasis
 - ↳ Abnormality of prenatal development or birth
 - ↳ Abnormality of the breast
 - ↳ Abnormality of the cardiovascular system
 - ↳ Abnormality of the digestive system
 - ↳ Abnormality of abdomen morphology
 - ↳ Abnormality of digestive system morphology
 - ↳ Abnormality of digestive system physiology
 - ↳ Abdominal symptom
 - ↳ Abdominal distension

| HP URI | HP class label | MP URI | MP class label | Vote | Exact |
|-----------------------|--|----------------------------|--|------|-------|
| ry.org/obo/HP_0004974 | coarctation of abdominal aorta | ibrary.org/obo/MP_0010525 | abdominal aorta coarctation | 4 | 0 |
| ry.org/obo/HP_0003363 | abdominal situs inversus | ibrary.org/obo/MP_0011249 | abdominal situs inversus | 4 | 1 |
| ry.org/obo/HP_0030684 | abnormal adiponectin level | ibrary.org/obo/MP_0004891 | abnormal adiponectin level | 4 | 1 |
| ry.org/obo/HP_0011733 | abnormality of adrenal physiology | ibrary.org/obo/MP_0002909 | abnormal adrenal gland physiology | 4 | 0 |
| ry.org/obo/HP_0004379 | abnormality of alkaline phosphatase activity | ibrary.org/obo/MP_0011583 | abnormal alkaline phosphatase activity | 4 | 0 |
| ry.org/obo/HP_0030333 | abnormal alpha-beta t cell morphology | ibrary.org/obo/MP_0012762 | abnormal alpha-beta t cell morphology | 4 | 1 |
| ry.org/obo/HP_0045056 | abnormal levels of alpha-fetoprotein | ibrary.org/obo/MP_0012600 | abnormal alpha-fetoprotein level | 4 | 0 |
| ry.org/obo/HP_0004337 | abnormality of amino acid metabolism | ibrary.org/obo/MP_0013242 | abnormal amino acid metabolism | 4 | 0 |
| ry.org/obo/HI | | | | 4 | 0 |
| ry.org/obo/HI | | | | 4 | 1 |
| ry.org/obo/HI | | | | 4 | 0 |
| ry.org/obo/HI | | | | 4 | 1 |
| ry.org/obo/HI | | | | 4 | 1 |
| ry.org/obo/HI | | | | 4 | 0 |
| ry.org/obo/HI | | | | 4 | 1 |
| ry.org/obo/HI | | | | 4 | 0 |
| ry.org/obo/HI | | | | 4 | 1 |
| ry.org/obo/HI | | | | 4 | 0 |
| ry.org/obo/HP_0012440 | abnormal biliary tract morphology | ibrary.org/obo/MP_0005083 | abnormal biliary tract morphology | 4 | 1 |
| ry.org/obo/HP_0040210 | abnormal level of biotin | ibrary.org/obo/MP_0005320 | abnormal biotin level | 4 | 0 |
| ry.org/obo/HP_0011028 | abnormality of blood circulation | ibrary.org/obo/MP_0002128 | abnormal blood circulation | 4 | 0 |
| ry.org/obo/HP_0012415 | abnormal blood gas level | ibrary.org/obo/MP_0002329 | abnormal blood gas level | 4 | 1 |
| ry.org/obo/HP_0000002 | abnormality of body height | ibrary.org/obo/MP_0001253 | abnormal body height | 4 | 0 |
| ry.org/obo/HP_0045081 | abnormality of body mass index | ibrary.org/obo/MP_00020182 | abnormal body mass index | 4 | 0 |
| ry.org/obo/HP_0004323 | abnormality of body weight | ibrary.org/obo/MP_0001259 | abnormal body weight | 4 | 0 |
| ry.org/obo/HP_0011862 | abnormal bone collagen fibril morphology | ibrary.org/obo/MP_0011642 | abnormal bone collagen fibril morphology | 4 | 1 |
| ry.org/obo/HP_0004210 | abnormal bone mineral density | ibrary.org/obo/MP_0010110 | abnormal bone mineral density | 4 | 0 |



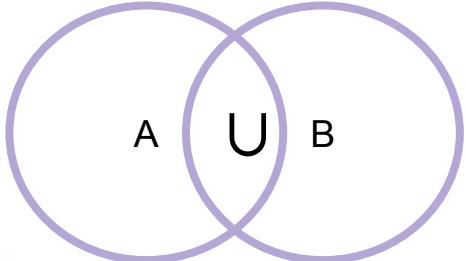
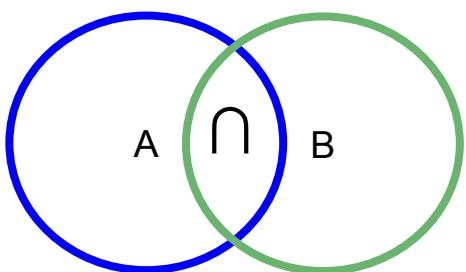


Applied Metrics

Jaccard Similarity

(Preferred Labels, Definitions, Synonyms)

Property Value
of Ontology A and
Ontology B



$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Jaccard Similarity Value

Jaccard Similarity (Preferred Labels, Definitions, Synonyms)

Mammalian Phenotype Ontology

Last uploaded: July 8, 2019

Summary Classes Properties Notes Mappings Widgets

Jump to:

- ⊕ mammalian phenotype
- ⊕ adipose tissue phenotype
- ⊕ behavior/neurological phenotype
- ⊕ cardiovascular system phenotype
- ⊕ cellular phenotype
- ⊕ craniofacial phenotype
- ⊕ digestive/alimentary phenotype
- ⊕ embryo phenotype
- ⊕ endocrine/exocrine gland phenotype
- ⊕ growth/size/body region phenotype
- ⊕ hearing/vestibular/ear phenotype
- ⊕ hematopoietic system phenotype
- ⊕ homeostasis/metabolism phenotype
- ⊕ immune system phenotype
- ⊕ integument phenotype
- ⊕ limbs/digits/tail phenotype
- ⊕ liver/biliary system phenotype

Details

Visualization

Notes (0)

Class Mappings (5)

Jaccard Similarity Value: 1

Preferred Name Lafora bodies

Synonyms

Lafora body accumulation

Human Phenotype Ontology

Last uploaded: June 3, 2019

Summary Classes Properties Notes Mappings Widgets

Jump to:

All

- ⊕ Blood group
- ⊕ Clinical course
- ⊕ Clinical modifier
- ⊕ Frequency
- ⊕ Mode of inheritance
- ⊕ Past medical history
- ⊕ Phenotypic abnormality
 - ⊕ Abnormal cellular phenotype
 - ⊕ Abnormality of blood and blood-forming tissue
 - ⊕ Abnormality of connective tissue
 - ⊕ Abnormality of head or neck

Details

Visualization

Notes (0)

Class Mappings (13)



Preferred Name

Lafora bodies

Definitions

An intraneuronal inclusion body composed of acid mucopolysaccharides.

ID

http://purl.obolibrary.org/obo/HP_0100318

created_by

doelkens

creation_date

2010-08-10T03:37:06Z

database_cross_reference

SNOMEDCT_US:87554006

UMLS:C0333749

Jaccard Similarity

(Preferred Labels, Definitions, Synonyms)

Mammalian Phenotype Ontology

Last uploaded: July 8, 2019

Summary Classes Properties Notes Mappings Widgets

Jump to:

- ↳ mammalian phenotype
- ↳ adipose tissue phenotype
- ↳ behavior/neurological phenotype
- ↳ cardiovascular system phenotype
- ↳ cellular phenotype
- ↳ craniofacial phenotype
- ↳ digestive/alimentary phenotype
- ↳ abnormal digestive system morphology
- ↳ abnormal digestive system physiology
- ↳ abnormal defecation
- ↳ abnormal digestion
- ↳ abnormal exocrine pancreas physiology
- ↳ abnormal feces composition
- melena**
- steatorrhea

Details Visualization Notes (0) Class Mappings (35) ⚙

Preferred Name melena

melanorrhea

melanorrhagia

Human Phenotype Ontology

Last uploaded: June 3, 2019

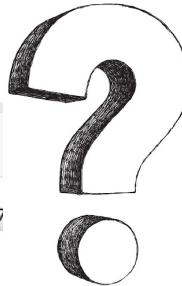
Summary Classes Properties Notes Mappings Widgets

Jump to:

All

- ↳ Blood group
- ↳ Clinical course
- ↳ Clinical modifier
- ↳ Frequency
- ↳ Mode of inheritance
- ↳ Past medical history
- ↳ Phenotypic abnormality
 - ↳ Abnormal cellular phenotype
 - ↳ Abnormality of blood and blood-forming tissue
 - ↳ Abnormality of connective tissue
 - ↳ Abnormality of head or neck

Jaccard Similarity Value: 0.666



Details Visualization Notes (

Preferred Name Melena

Synonyms Black feces

Definitions

The passage of blackish, tarry feces associated with gastrointestinal hemorrhage. in the colon long enough for it to be broken down by colonic bacteria. One degra stool with a blackish color. Thus, melena generally occurs with bleeding from the stomach ulcers or duodenal ulcers), since the blood usually remains in the gut for lower gastrointestinal bleeding.

ID

http://purl.obolibrary.org/obo/HP_0002249

Jaccard Similarity

(Preferred Labels, Definitions, Synonyms)

Mammalian Phenotype Ontology

Last uploaded: July 8, 2019

Summary Classes Properties Notes Mappings Widgets

Jump to:

- mammalian phenotype
 - adipose tissue phenotype
 - behavior/neurological phenotype
 - cardiovascular system phenotype
 - cellular phenotype
 - craniofacial phenotype
 - digestive/alimentary phenotype
 - embryo phenotype
 - abnormal embryo development
 - abnormal embryo morphology
 - abnormal embryo size

Details Visualization Notes (0) Class Mappings (7) ⚙

Preferred Name abnormal neural tube morphology

Synonyms neural tube dysplasia

neural tube defects

abnormal neural tube morphology/development

central nervous system: neural tube defects

Definitions any structural anomaly of the hollow epithelial tube found on the dorsal side of the vertebrate embryo that develops

into the central nervous system (i.e. brain and spinal cord)

Jaccard Similarity Value: 0.96



Human Phenotype Ontology

Last uploaded: June 3, 2019

Summary Classes Properties Notes Mappings Widgets

Jump to:

- All
- Blood group
- Clinical course
- Clinical modifier
- Frequency
- Mode of inheritance
- Data type

Details Visualization Notes (0) Class Mappings (14) ⚙

Preferred Name Abnormal neural tube morphology

Definitions Any structural anomaly of the hollow epithelial tube found on the dorsal side of the vertebrate embryo that develops

into the central nervous system (i.e. brain and spinal cord).

ID http://purl.obolibrary.org/obo/HP_0410043

Jaccard Similarity

(Preferred Labels, Definitions, Synonyms)

Mammalian Phenotype Ontology

Last uploaded: July 8, 2019

Summary Classes Properties Notes Mappings Widgets

Jump to:

- ⊕ mammalian phenotype
- ⊕ adipose tissue phenotype
- ⊕ behavior/neurological phenotype
- ⊕ cardiovascular system phenotype
- ⊕ cellular phenotype
- ⊕ craniofacial phenotype
- ⊕ digestive/alimentary phenotype
- ⊕ embryo phenotype
- ⊕ endocrine/exocrine gland phenotype

Details Visualization Notes (0) Class Mappings (65) ⚙

| | |
|----------------|---|
| Preferred Name | dehydration |
| Definitions | excessive water loss from the body or from an organ or bodily part |
| ID | http://owl.obiobrain.org/obo/MR_0901439 |
| definition | excessive water loss from the body or from an organ or bodily part |

has_obo_namespace MPheno.ontology

Human Phenotype Ontology

Last uploaded: June 3, 2019

Summary Classes Properties Notes Mappings Widgets

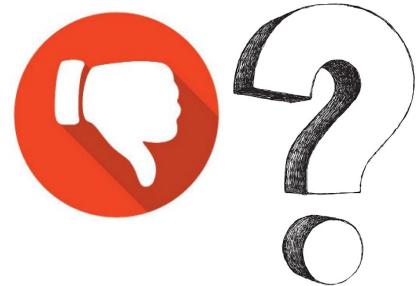
Jump to:

- ⊕ All
- ⊕ Blood group
- ⊕ Clinical course
- ⊕ Clinical modifier
- ⊕ Frequency
- ⊕ Haplotype of inheritance
- ⊕ Past medical history
- ⊕ Phenotypic abnormality
- ⊕ Abnormal cellular phenotype
- ⊕ Abnormality of blood and blood-forming tissue
- ⊕ Abnormality of connective tissue
- ⊕ Abnormality of head or neck
- ⊕ Abnormality of limbs
- ⊕ Abnormality of metabolism/homeostasis
- ⊕ Abnormal cellular physiology
- ⊕ Abnormal cerebrospinal fluid/metabolite concentration
- ⊕ Abnormal circulating metabolite concentration
- ⊕ Abnormal circulating nucleobase concentration
- ⊕ Abnormal enzyme/coenzyme activity
- ⊕ Abnormal erythrocyte sedimentation rate
- ⊕ Abnormal fat metabolism result
- ⊕ Abnormal homeostasis
- ⊕ Abnormal energy expenditure
- ⊕ Abnormal glucose homeostasis
- ⊕ Abnormal sweat homeostasis
- ⊕ Abnormal water homeostasis

Details Visualization Notes (0) Class Mappings (77) ⚙

| | |
|--------------------------|--|
| Preferred Name | Dehydration |
| Synonyms | |
| ID | library.org/obo/HP_0001944 |
| database_cross_reference | 34095006 |
| has_exact_synonym | S |
| has_obo_namespace | hpe |
| id | |
| label | Dehydration |
| notation | HP:0001944 |
| prefLabel | Dehydration |
| treeView | Abnormality of fluid regulation |
| subClassOf | Abnormality of fluid regulation |

Jaccard Similarity Value: 0



Jaccard Similarity

(Preferred Labels, Definitions, Synonyms)

Human Phenotype Ontology
Last uploaded: June 3, 2019

Summary Classes Properties Notes Mappings Widgets

Jump to:

- ⊕ All
 - ⊕ Blood group
 - ⊕ Clinical course
 - ⊕ Clinical modifier
 - ⊕ Frequency
 - ⊕ Mode of inheritance
 - ⊕ Past medical history
 - ⊕ Phenotypic abnormality
 - ⊕ Abnormal cellular phenotype
 - ⊕ Abnormality of blood and blood-forming tissue
 - ⊕ Abnormality of connective tissue
 - ⊕ Abnormality of head or neck
 - ⊕ Abnormality of limbs
 - ⊕ Abnormality of metabolism/homeostasis
 - ⊕ Abnormality of prenatal development or birth
 - ⊕ Abnormality of the breast
 - ⊕ Abnormality of the cardiovascular system
 - ⊕ Abnormality of the digestive system
 - ⊕ Abnormality of the ear
 - ⊕ Abnormal ear morphology
 - ⊕ Abnormality of the inner ear

| Details | Visualization | Notes (0) | Class Mappings (18) | ⊕ |
|--------------------------|--|-----------|---------------------|---|
| Preferred Name | Abnormal vestibulo-ocular reflex | | | |
| Synonyms | Abnormal vestibuloocular reflex | | | |
| Definitions | An abnormality of the vestibulo-ocular reflex (VOR). passive or active head movements in one direction a | | | |
| ID | http://purl.obolibrary.org/obo/HP_0007670 | | | |
| database_cross_reference | UMLS:C4021571 | | | |
| definition | An abnormality of the vestibulo-ocular reflex (VOR). passive or active head movements in one direction a | | | |
| has_exact_synonym | Abnormal vestibuloocular reflex | | | |
| has_obo_namespace | human_phenotype | | | |
| label | Abnormal vestibulo-ocular reflex | | | |



Jaccard Similarity Value: 0.94



Mammalian Phenotype Ontology

Last uploaded: July 8, 2019

Summary Classes Properties Notes Mappings Widgets

Jump to:

- ⊕ mammalian phenotype
 - ⊕ adipose tissue phenotype
 - ⊕ behavior/neurological phenotype
 - ⊕ cardiovascular system phenotype
 - ⊕ cellular phenotype
 - ⊕ craniofacial phenotype
 - ⊕ digestive/alimentary phenotype
 - ⊕ embryo phenotype
 - ⊕ endocrine/exocrine gland phenotype

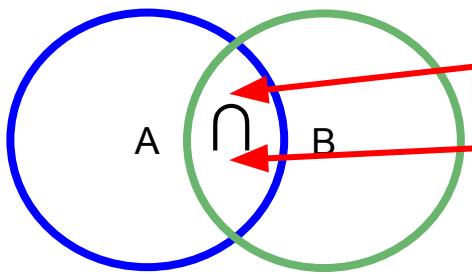
Details Visualization Notes (0) Class Mappings (10) ⊕

| | |
|----------------|----------------------------------|
| Preferred Name | abnormal vestibuloocular reflex |
| Synonyms | abnormal VOR |
| | abnormal ocular reflex |
| | abnormal vestibulo ocular reflex |
| | abnormal nystagmus |
| | abnormal eye reflex |

Sigmoid Similarity

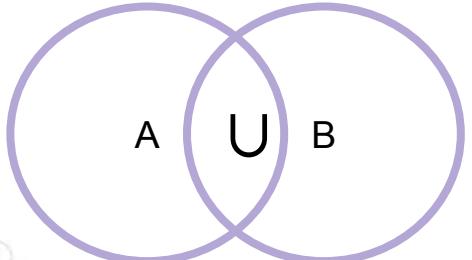
(Preferred Labels, Definitions, Synonyms)

Property Value
of Ontology A and
Ontology B



$$\text{SIM}_S(O_1, O_2) = \frac{e^{\text{CF}(O_1, O_2)} - 1}{(e^{\text{CF}(O_1, O_2)} + 1)(\text{DF}(O_1) + \text{DF}(O_2) + 1)}$$

Union - Intersection



Sigmoid Similarity Value

Sigmoid Similarity

(Preferred Labels, Definitions, Synonyms)

Mammalian Phenotype Ontology

Last uploaded: July 8, 2019

Summary Classes Properties Notes Mappings Widgets

Jump to:

- ⌚ mammalian phenotype
- ⌚ adipose tissue phenotype
- ⌚ behavior/neurological phenotype
- ⌚ cardiovascular system phenotype
- ⌚ cellular phenotype
- ⌚ craniofacial phenotype
- ⌚ digestive/alimentary phenotype
- ⌚ embryo phenotype
- ⌚ endocrine/exocrine gland phenotype
- ⌚ growth/size/body region phenotype
- ⌚ hearing/vestibular/ear phenotype
- ⌚ hematopoietic system phenotype
- ⌚ homeostasis/metabolism phenotype
- ⌚ immune system phenotype
- ⌚ integument phenotype
- ⌚ limbs/digits/tail phenotype
- ⌚ liver/biliary system phenotype

Details

Visualization

Notes (0)

Class Mappings (5)

Sigmoid Similarity Value: 0.99



Preferred Name Lafora bodies

Synonyms

Lafora body accumulation

Human Phenotype Ontology

Last uploaded: June 3, 2019

Summary Classes Properties Notes Mappings Widgets

Jump to:

All

- ⌚ Blood group
- ⌚ Clinical course
- ⌚ Clinical modifier
- ⌚ Frequency
- ⌚ Mode of inheritance
- ⌚ Past medical history
- ⌚ Phenotypic abnormality
 - ⌚ Abnormal cellular phenotype
 - ⌚ Abnormality of blood and blood-forming tissue
 - ⌚ Abnormality of connective tissue
 - ⌚ Abnormality of head or neck

Details

Visualization

Notes (0)

Class Mappings (13)

Preferred Name

Lafora bodies

Definitions

An intraneuronal inclusion body composed of acid mucopolysaccharides.

ID

http://purl.obolibrary.org/obo/HP_0100318

created_by

doekens

creation_date

2010-08-10T03:37:06Z

database_cross_reference

SNOMEDCT_US:87554006
UMLS:C0333749

Sigmoid Similarity

(Preferred Labels, Definitions, Synonyms)

Mammalian Phenotype Ontology

Last uploaded: July 8, 2019

Summary Classes Properties Notes Mappings Widgets

Jump to:

- ↳ mammalian phenotype
- ↳ adipose tissue phenotype
- ↳ behavior/neurological phenotype
- ↳ cardiovascular system phenotype
- ↳ cellular phenotype
- ↳ craniofacial phenotype
- ↳ digestive/alimentary phenotype
- ↳ abnormal digestive system morphology
- ↳ abnormal digestive system physiology
- ↳ abnormal defecation
- ↳ abnormal digestion
- ↳ abnormal exocrine pancreas physiology
- ↳ abnormal feces composition
- melena**
- steatorrhea

Details Visualization Notes (0) Class Mappings (35) ⚙

Preferred Name melena

melanorrhea

melanorrhagia

Sigmoid Similarity Value: **0.321**

Human Phenotype Ontology

Last uploaded: June 3, 2019

Summary Classes Properties Notes Mappings Widgets

Jump to:

- ↳ All
- ↳ Blood group
- ↳ Clinical course
- ↳ Clinical modifier
- ↳ Frequency
- ↳ Mode of inheritance
- ↳ Past medical history
- ↳ Phenotypic abnormality
 - ↳ Abnormal cellular phenotype
 - ↳ Abnormality of blood and blood-forming tissue
 - ↳ Abnormality of connective tissue
 - ↳ Abnormality of head or neck

Details Visualization Notes (

Preferred Name Melena

Black feces

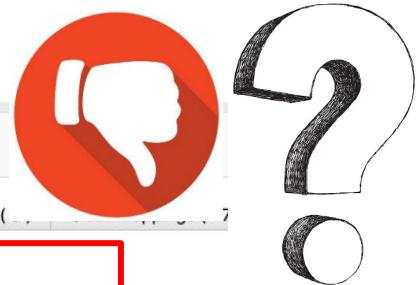
Synonyms

Definitions

ID

The passage of blackish, tarry feces associated with gastrointestinal hemorrhage. in the colon long enough for it to be broken down by colonic bacteria. One degra stool with a blackish color. Thus, melena generally occurs with bleeding from the stomach ulcers or duodenal ulcers), since the blood usually remains in the gut for lower gastrointestinal bleeding.

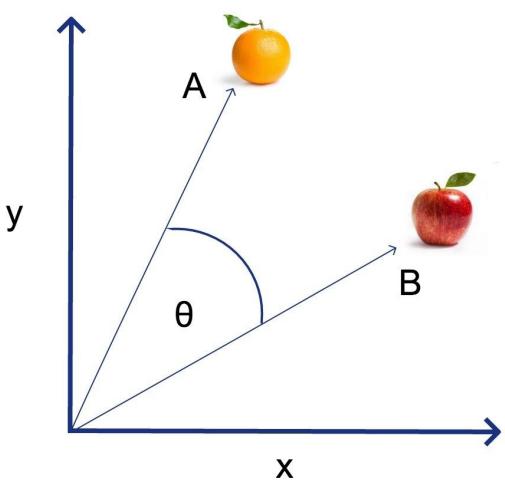
http://purl.obolibrary.org/obo/HP_0002249



Cosine Similarity

(Preferred Labels, Definitions, Synonyms)

$$\text{similarity}(A, B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}}$$



Count frequency of words occurrence (Vector1, Vector2)

Count Number of Different words (Vector1, Vector2)

Num Similar words,
Num Different words

Cosine Similarity Value

Cosine Similarity

(Preferred Labels, Definitions, Synonyms)

Mammalian Phenotype Ontology

Last uploaded: July 8, 2019

Summary Classes Properties Notes Mappings Widgets

Jump to:

- ⊕ mammalian phenotype
- ⊕ adipose tissue phenotype
- ⊕ behavior/neurological phenotype
- ⊕ cardiovascular system phenotype
- ⊕ cellular phenotype
- ⊕ craniofacial phenotype
- ⊕ digestive/alimentary phenotype
- ⊕ embryo phenotype
- ⊕ endocrine/exocrine gland phenotype
- ⊕ growth/size/body region phenotype
- ⊕ hearing/vestibular/ear phenotype
- ⊕ hematopoietic system phenotype
- ⊕ homeostasis/metabolism phenotype
- ⊕ immune system phenotype
- ⊕ integument phenotype
- ⊕ limbs/digits/tail phenotype
- ⊕ liver/biliary system phenotype

Details

Visualization

Notes (0)

Class Mappings (5)



Preferred Name

Lafora bodies

Synonyms

Lafora body accumulation

Cosine Similarity Value: 1



Human Phenotype Ontology

Last uploaded: June 3, 2019

Summary

Classes

Properties

Notes

Mappings

Widgets

Jump to:

All

- ⊕ Blood group
- ⊕ Clinical course
- ⊕ Clinical modifier
- ⊕ Frequency
- ⊕ Mode of inheritance
- ⊕ Past medical history
- ⊕ Phenotypic abnormality
 - ⊕ Abnormal cellular phenotype
 - ⊕ Abnormality of blood and blood-forming tissue
 - ⊕ Abnormality of connective tissue
 - ⊕ Abnormality of head or neck

Details

Visualization

Notes (0)

Class Mappings (13)



Preferred Name

Lafora bodies

Definitions

An intraneuronal inclusion body composed of acid mucopolysaccharides.

ID

http://purl.obolibrary.org/obo/HP_0100318

created_by

doekens

creation_date

2010-08-10T03:37:06Z

database_cross_reference

SNOMEDCT_US:87554006

UMLS:C0333749

Cosine Similarity

(Preferred Labels, Definitions, Synonyms)

Mammalian Phenotype Ontology

Last uploaded: July 8, 2019

Summary Classes Properties Notes Mappings Widgets

Jump to:

- ⊖ mammalian phenotype
- ⊕ adipose tissue phenotype
- ⊕ behavior/neurological phenotype
- ⊕ cardiovascular system phenotype
- ⊕ cellular phenotype
- ⊕ craniofacial phenotype
- ⊕ digestive/alimentary phenotype
- ⊕ abnormal digestive system morphology
- ⊕ abnormal digestive system physiology
- ⊕ abnormal defecation
- ⊕ abnormal digestion
- ⊕ abnormal exocrine pancreas physiology
- ⊕ abnormal feces composition
- melena**
- steatorrhea

Details Visualization Notes (0) Class Mappings (35) ⚙

Preferred Name melena

melanorrhea

melanorrhagia

Human Phenotype Ontology

Last uploaded: June 3, 2019

Summary Classes Properties Notes Mappings Widgets

Jump to:

- ⊖ All
- ⊕ Blood group
- ⊕ Clinical course
- ⊕ Clinical modifier
- ⊕ Frequency
- ⊕ Mode of inheritance
- ⊕ Past medical history
- ⊕ Phenotypic abnormality
- ⊕ Abnormal cellular phenotype
- ⊕ Abnormality of blood and blood-forming tissue
- ⊕ Abnormality of connective tissue
- ⊕ Abnormality of head or neck

Details Visualization Notes (0) Class Mappings (47) ⚙

Preferred Name Melena

Synonyms Black feces

Definitions

The passage of blackish, tarry feces associated with gastrointestinal hemorrhage in the colon long enough for it to be broken down by colonic bacteria. One degrades stool with a blackish color. Thus, melena generally occurs with bleeding from the stomach ulcers or duodenal ulcers), since the blood usually remains in the gut for lower gastrointestinal bleeding.

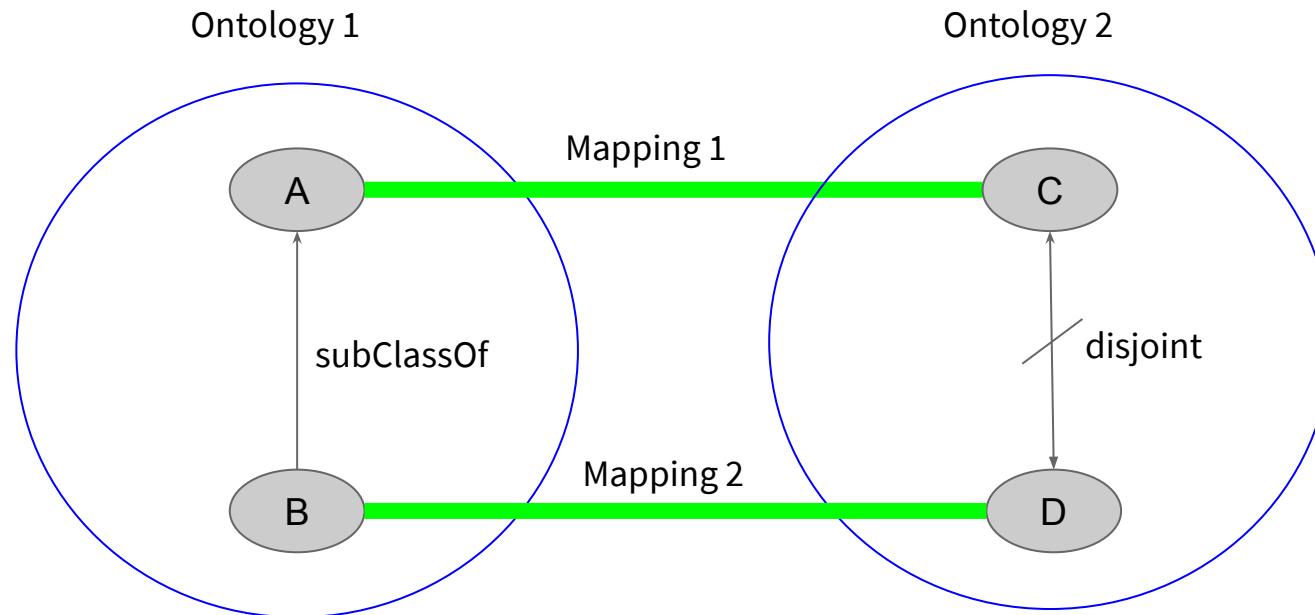
ID

http://purl.obolibrary.org/obo/HP_0002249

Cosine Similarity Value: 1

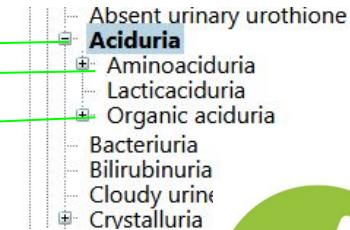
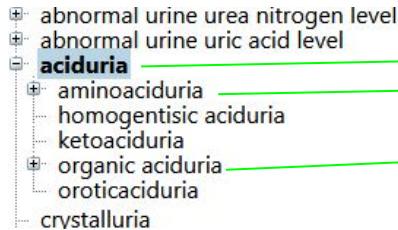


Children Metric

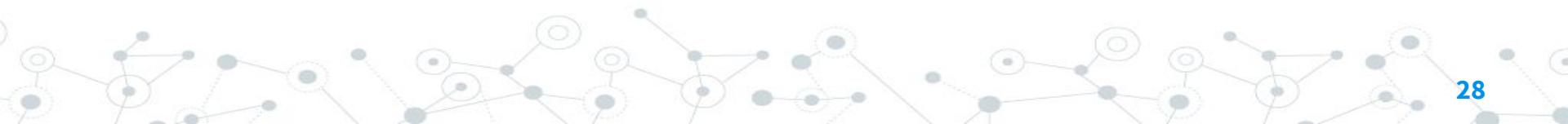


- **Logical inconsistent:** D should be a subClassOf C, if both ontologies are consistent and mappings are valid

Children Metric



- **Logical consistent:** all subclasses of Aciduria where a mapping exists, are subclasses in both ontologies



Children Metric

- vision/eye phenotype
- abnormal eye morphology
 - abnormal anterior eye segment morphology
 - abnormal anterior uvea morphology
 - abnormal aqueous drainage system morphology
 - abnormal ciliary body morphology**
 - abnormal ciliary body pigmentation
 - abnormal ciliary epithelium morphology
 - abnormal ciliary muscle morphology
 - abnormal ciliary process morphology
 - absent ciliary body
 - ciliary body coloboma
 - ciliary body hypoplasia
 - abnormal iris morphology
 - abnormal ciliary zonule morphology
 - abnormal conjunctiva morphology
 - abnormal conjunctival sac morphology
 - abnormal cornea morphology
 - abnormal eye anterior chamber morphology
 - abnormal eye posterior chamber morphology
 - abnormal lens morphology
 - abnormal tear film morphology
 - anterior staphylooma
 - abnormal cornea limbus morphology

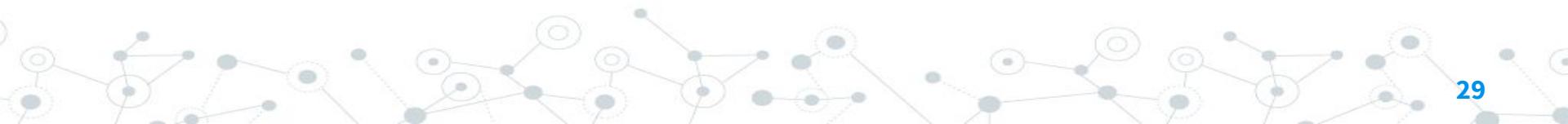
| | |
|-------------------|--|
| | the processes radiate from the ciliary muscle and give attachments to ligaments supporting the lens of the eye |
| has_exact_synonym | ciliary body abnormalities |
| has_obo_namespace | MPheno.ontology |
| id | MP:0005099 |
| in_subset | http://purl.obolibrary.org/obo/mp#Sanger_Terms http://purl.obolibrary.org/obo/mp#IMPC |
| label | abnormal ciliary body morphology |
| notation | MP:0005099 |
| prefLabel | abnormal ciliary body morphology |

- Abnormal posterior eye segment morphology
- Abnormal sclera morphology
- Abnormal uvea morphology
- Abnormal choroid morphology
- Abnormal ciliary body morphology**
 - Ciliary body melanoma
 - Hypoplasia of the ciliary body
 - Abnormality iris morphology
 - Aplasia/Hypoplasia affecting the uvea
 - Uveitis
 - Abnormality of globe location
 - Abnormality of globe size
 - Abnormality of the vasculature of the eye
 - Aplasia/Hypoplasia affecting the eye
 - Coloboma**
 - Chorioretinal coloboma
 - Ciliary body coloboma
 - Iris coloboma
 - Lens coloboma
 - Optic nerve coloboma
 - Retinal coloboma
 - Neoplasm of the eye
 - Phthisis bulbi
 - Spontaneous rupture of the globe



| | |
|-------------------|-----------------------------|
| has_obo_namespace | human_phenotype |
| id | HP:0012776 |
| label | Abnormal ciliary morphology |
| notation | HP:0012776 |
| prefLabel | Abnormal ciliary morphology |
| treeView | Abnormal uvea morphology |
| subClassOf | Abnormal uvea morphology |

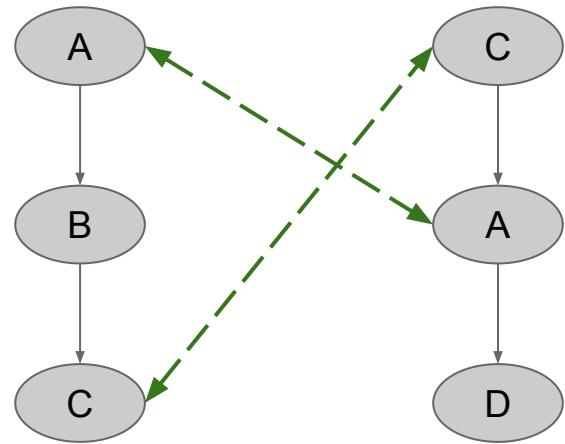
Logical inconsistent: ciliary body coloboma is subclass of abnormal ciliary body morphology in MP, but not in HP



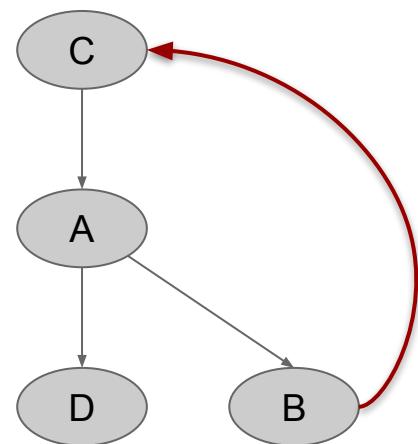
Loop Detection

Ontology A

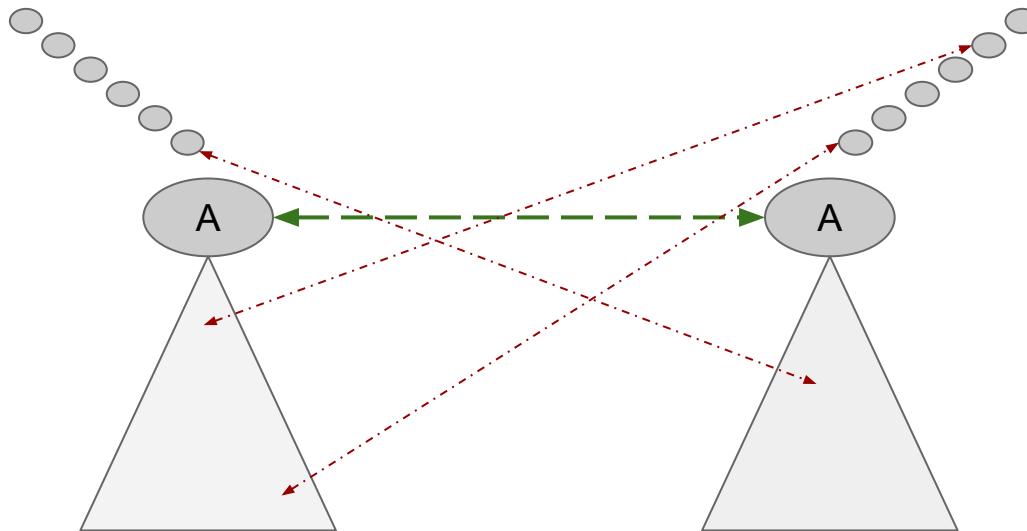
Ontology B



$A \cup B$



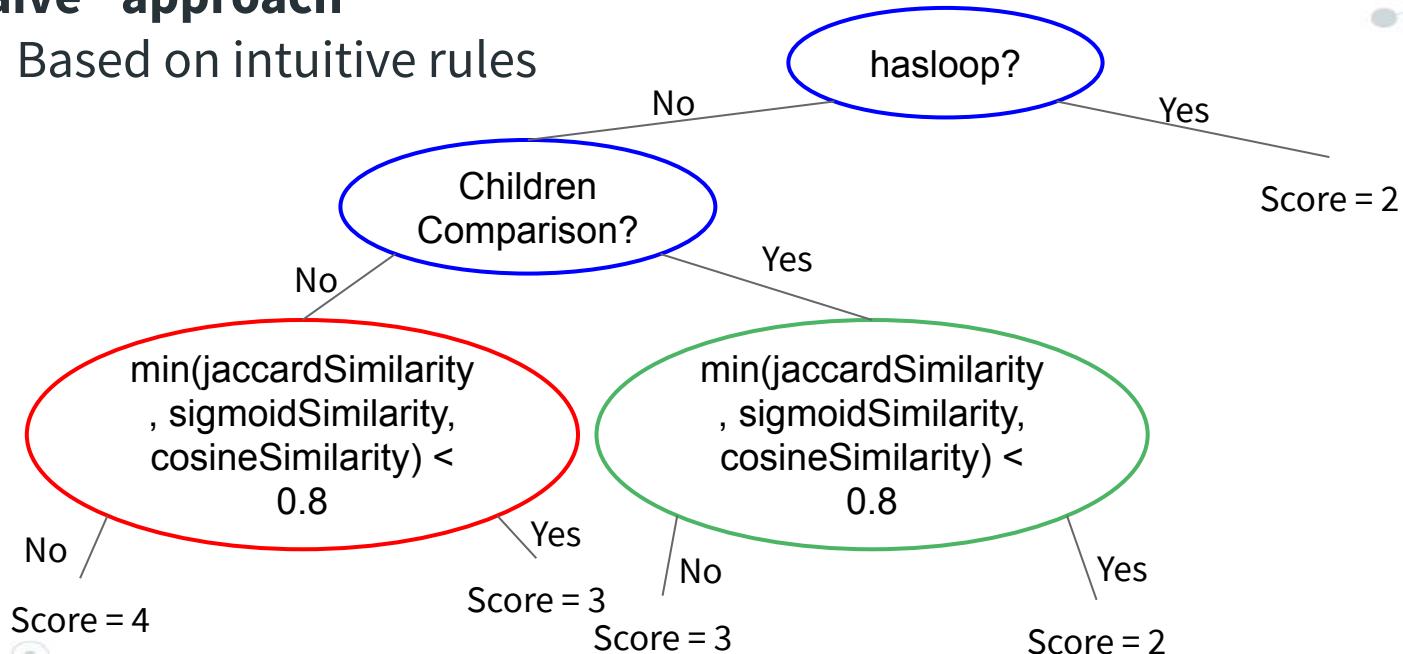
Our Loop Detection



Calculate a Confidence Score after applying Metrics

“Naive” approach

- Based on intuitive rules



Calculate a Confidence Score after applying Metrics

- **Machine Learning approaches**
 - Based on logistic regression
 - Based on random forest
- **Ground truth** provided by silver standard (Pistoia et al.)

| jaccardSimilarity | sigmoidSimilarity | cosineSimilarity | jaccardSubtree | cosineSubtree | descendantsEvaluation | hasLoop | silverMapping |
|-------------------|-------------------|------------------|----------------|---------------|-----------------------|---------|---------------|
| 0.900000 | 0.499877 | 1.0 | 1.0 | 1.000000 | 0 | 0 | 4 |
| 0.909091 | 0.499955 | 1.0 | 0.0 | 0.000000 | 0 | 0 | 4 |
| 0.950000 | 0.500000 | 1.0 | 0.0 | 0.269333 | 0 | 0 | 4 |
| 0.833333 | 0.333303 | 1.0 | 1.0 | 1.000000 | 0 | 0 | 4 |
| 0.909091 | 0.499955 | 1.0 | 1.0 | 1.000000 | 0 | 0 | 4 |

Features

Labels

Use 60% of mappings as training set and 40% as test set



Evaluation of Results

How did we evaluate our results?

- Evaluation of mappings between Human Phenotype Ontology (HP) and Mammalian Phenotype Ontology (MP)
 - **Ground truth scores** provided by silver standard (Pistoia et al.)
 - **Compare those scores** to our scores by **naive approach** vs **machine learning techniques**
- **Test set contains 750 mappings** we evaluated
- **Problem:** silver standard from 2017 -> new mappings were created, bad mappings deleted
- **Only 557 mappings we could compare against silver standard**
 - 522 with score 4
 - 7 with score 3
 - 28 with score 2

Results from applying metrics

| MP URI | HP URI | jaccardS | jaccard | jaccardSim | sigmo | sigmo | sigmo | cosine | cosineS | cosineS | jaccar | cosineS | deser | hasLoop | silverMapping | ourScore | |
|---------|---------|----------|---------|------------|-------|--------|-------|--------|---------|---------|--------|---------|-------|---------|---------------|----------|---|
| http:// | http:// | 0.9 | 0.8 | 0.53333333 | 0.499 | 0.166 | 0.124 | 1 | 0.61631 | 0.35355 | 1 | 1 | 1 | FALSE | FALSE | 4 | 4 |
| http:// | http:// | 0.944444 | 0.76 | 0 | 0.499 | 0.1428 | 0 | 1 | 0.72501 | 0 | 0 | 0.2535 | 0 | FALSE | FALSE | 0 | 4 |
| http:// | http:// | 0.909090 | 0 | 0.90909090 | 0.499 | 0 | 0.499 | 1 | 0 | 1 | 0 | 0 | 0 | FALSE | FALSE | 4 | 4 |
| http:// | http:// | 0.95 | 0.64285 | 0 | 0.499 | 0.0909 | 0 | 1 | 0.34554 | 0 | 0 | 0.2693 | 0 | FALSE | FALSE | 4 | 4 |
| http:// | http:// | 0.9375 | 0.68181 | 0.72222222 | 0.499 | 0.1248 | 0.166 | 1 | 0.47140 | 0.28867 | 0.0285 | 0.56355 | 0 | FALSE | FALSE | 0 | 4 |
| http:// | http:// | 0.833333 | 0.80769 | 0 | 0.333 | 0.166 | 0 | 1 | 0.71111 | 0 | 1 | 1 | 1 | FALSE | FALSE | 4 | 4 |
| http:// | http:// | 0.9375 | 0.68181 | 0.52323232 | 0.499 | 0.1176 | 0.100 | 1 | 0.57112 | 0 | 1 | 1 | 1 | FALSE | FALSE | 4 | 4 |
| http:// | http:// | 0.9375 | 0.68181 | 0.72222222 | 0.499 | 0.1176 | 0.100 | 1 | 0.28867 | 0 | 0.2703 | 0 | 0 | TRUE | TRUE | 2 | 4 |
| http:// | http:// | 0.933333 | 0.875 | 0 | 0.499 | 0.2498 | 0 | 1 | 0.24806 | 0 | 0.1764 | 0.9414 | 0 | FALSE | FALSE | 4 | 4 |
| http:// | http:// | 0.923076 | 0 | 0 | 0.499 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | FALSE | FALSE | 4 | 4 |
| http:// | http:// | 0.916666 | 0.81818 | 0.52631578 | 0.499 | 0.1999 | 0.099 | 1 | 0.50783 | 0 | 0 | 0 | 0 | FALSE | FALSE | 4 | 4 |
| http:// | http:// | 0.9375 | 0.69565 | 0.65 | 0.499 | 0.1248 | 0.124 | 1 | 0.46358 | 0.28867 | 0 | 0 | 0 | FALSE | FALSE | 2 | 4 |
| http:// | http:// | 0.714285 | 0.85714 | 0.29411764 | 0.328 | 0.2498 | 0.075 | 1 | 0.62994 | 0 | 0.25 | 0.8249 | 0 | FALSE | FALSE | 4 | 4 |
| http:// | http:// | 0.9375 | 0.54545 | 0.6 | 0.499 | 0.062 | 0.111 | 1 | 0.21170 | 0 | 0.0655 | 0.4039 | 1 | TRUE | FALSE | 0 | 3 |
| http:// | http:// | 0.833333 | 0.78260 | 0.52941176 | 0.333 | 0.166 | 0.111 | 1 | 0.41247 | 0.35355 | 1 | 1 | 1 | FALSE | FALSE | 4 | 4 |
| http:// | http:// | 0.818181 | 0.69565 | 0 | 0.333 | 0.1248 | 0 | 1 | 0.67419 | 0 | 1 | 1 | 1 | FALSE | FALSE | 4 | 4 |
| http:// | http:// | 0.8 | 0.80769 | 0.33333333 | 0.333 | 0.166 | 0.089 | 1 | 0.68640 | 0 | 0.3333 | 0.8860 | 1 | TRUE | FALSE | 0 | 3 |
| http:// | http:// | 0.941176 | 0.88 | 0.68421052 | 0.499 | 0.2498 | 0.142 | 1 | 0.78422 | 0 | 0 | 0.1967 | 1 | FALSE | FALSE | 0 | 4 |

Input for the calculation of score and Machine Learning



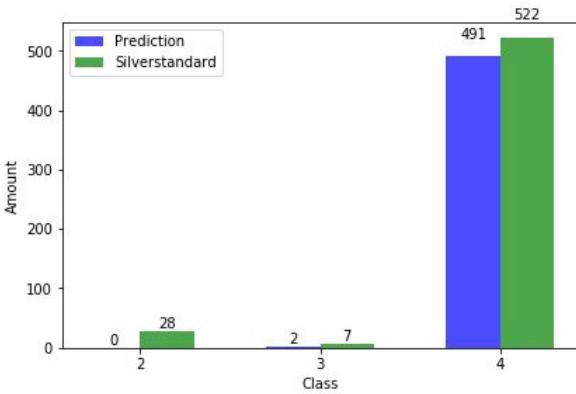
Well-known metrics

| Metrics | Evaluates |
|--------------------------------------|--|
| Precision (P_t) - Correctness | $\frac{\text{Number of correctly detected matching pairs}}{\text{Total number of detected pairs}}$ |
| Recall (R_t) - Completeness | $\frac{\text{Number of correctly detected matching pairs}}{\text{Total number of matching pairs given a set of ontology instances}}$ |
| F1-score ($F1_t$) | Aggregation of precision and recall |

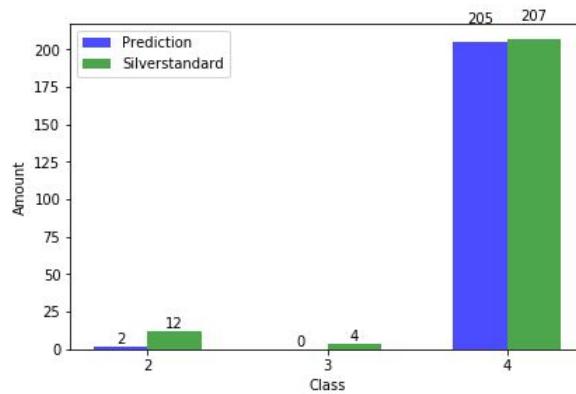
We evaluated our results based on them

| Method | Precision | Recall | F1-measure |
|---------------------|-----------|---------|------------|
| Naive | 0.88389 | 0.88509 | 0.88352 |
| Logistic Regression | 0.86165 | 0.92825 | 0.89371 |
| Random Forest | 0.90198 | 0.91479 | 0.90835 |

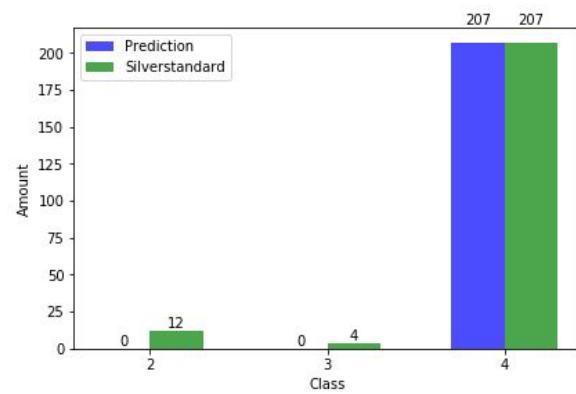
Class specific Results - True Positives



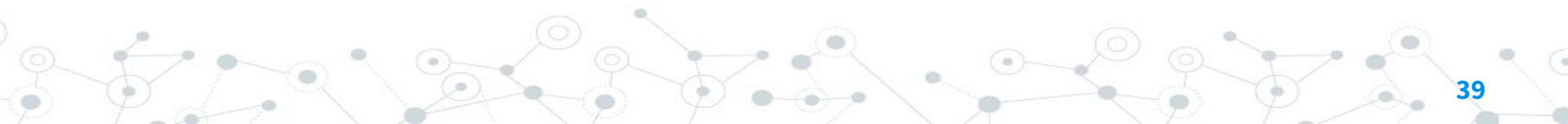
Naive



Random Forest



Logistic Regression





Conclusion & Outlook

What we have learnt...

- **Different metrics** to assess quality of matchings
- **Evaluation of results:** naive approach vs machine learning techniques



Problems we faced

- Big amounts of data
 - REST API very slow

During Evaluation:

- Ground truth was small
 - hard to apply machine learning
- Hard to evaluate quality of links manually, without domain knowledge

Outlook

- Extension to more ontologies
→ Ground truth needed
- More advanced evaluation of the results
- Apply metrics based on semantic

Thanks!

Any questions?

Literature

- What is an ontology? - Gruber - 1992
- Development 101: a guide to creating your first ontology - Noy, McGuinness - 2001
- Ontology instance linking: towards interlinked knowledge graphs - Helfin, Song - 2016
- Ten challenges for ontology matching - Shvaiko, Euzenat - 2008
- Modulo 5.2 Ontology Matching - Knowledge and Data VU Amsterdam - 2018
- A survey on ontology matching techniques - Hooi, Hassan, Shariff - 2014
- Measuring Similarity between Ontologies - Maedche, Staab - 2002
- Assessing Similarity Value between Two Ontologies - Ngom et al. - 2018
- How to Improve Jaccard's feature-based similarity measure - Likavec et al. - 2018
- Where to search top-K biomedical ontologies? - Oliveira et al. - 2018