Semantics

Matthew J. Graham
CACR

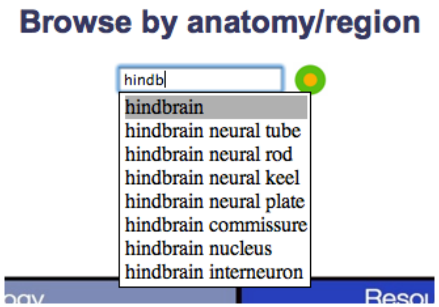
Methods of Computational Science
Caltech, 2011 May 10

matthew graham

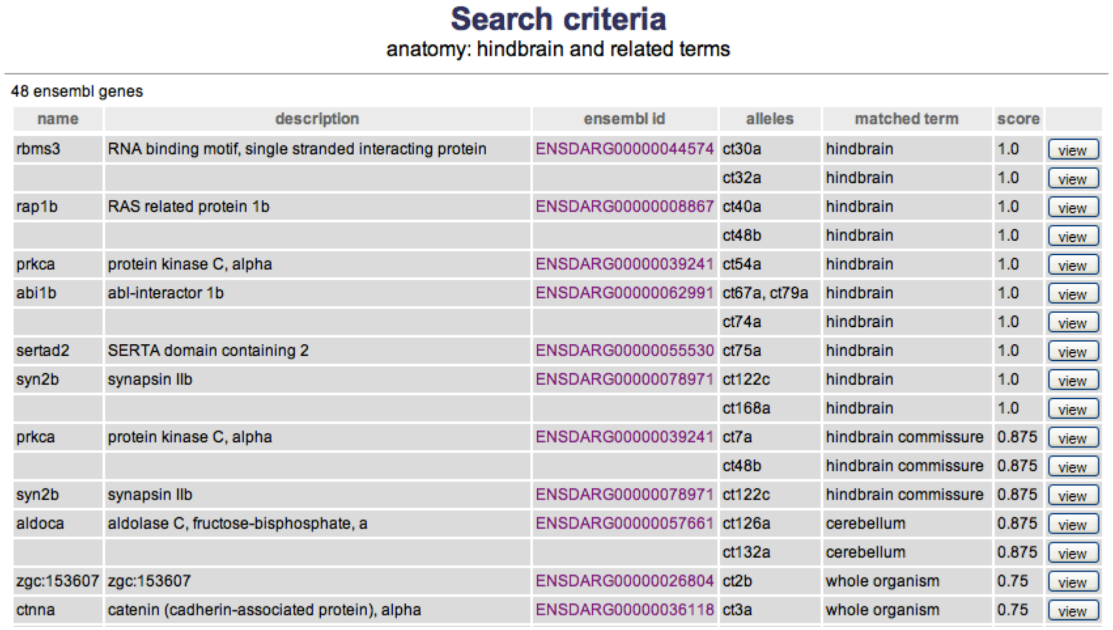
semantic web

- The future of the Internet (Web 3.0)
- Decentralized platform for distributed knowledge
- A web of databases
- Logical pieces of meaning that can be mechanically manipulated by a machine
- Vocabularies for making assertions about things
- Data consistency and inferencing
- Smart applications

smart apps



Browse by anatomy/region



Search criteria

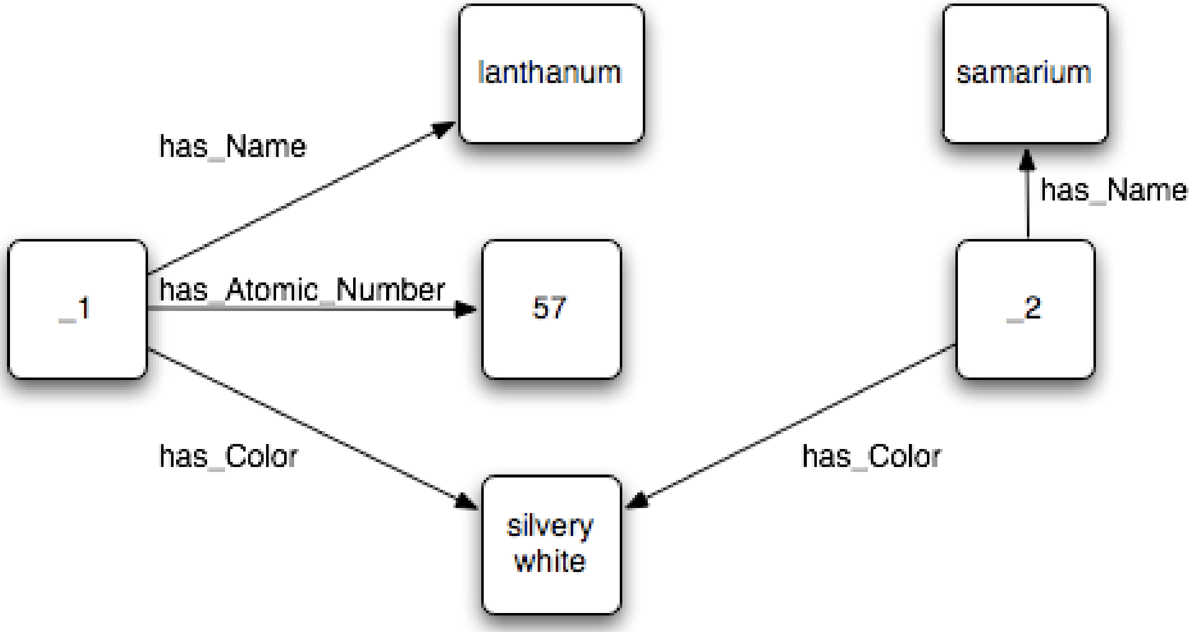
anatomy: hindbrain and related terms

48 ensembl genes

name	description	ensembl id	alleles	matched term	score	
rbms3	RNA binding motif, single stranded interacting protein	ENSDARG00000044574	ct30a	hindbrain	1.0	view
			ct32a	hindbrain	1.0	view
rap1b	RAS related protein 1b	ENSDARG00000008867	ct40a	hindbrain	1.0	view
			ct48b	hindbrain	1.0	view
prkca	protein kinase C, alpha	ENSDARG00000039241	ct54a	hindbrain	1.0	view
abi1b	abl-interactor 1b	ENSDARG00000062991	ct67a, ct79a	hindbrain	1.0	view
			ct74a	hindbrain	1.0	view
sertad2	SERTA domain containing 2	ENSDARG00000055530	ct75a	hindbrain	1.0	view
syn2b	synapsin IIb	ENSDARG00000078971	ct122c	hindbrain	1.0	view
			ct168a	hindbrain	1.0	view
prkca	protein kinase C, alpha	ENSDARG00000039241	ct7a	hindbrain commissure	0.875	view
			ct48b	hindbrain commissure	0.875	view
syn2b	synapsin IIb	ENSDARG00000078971	ct122c	hindbrain commissure	0.875	view
aldoca	aldolase C, fructose-bisphosphate, a	ENSDARG00000057661	ct126a	cerebellum	0.875	view
			ct132a	cerebellum	0.875	view
zgc:153607	zgc:153607	ENSDARG00000026804	ct2b	whole organism	0.75	view
ctnna	catenin (cadherin-associated protein), alpha	ENSDARG00000036118	ct3a	whole organism	0.75	view

knowledge as a graph

- Information is best expressed as a labelled, directed graph (entity-attribute-value data model)



resource description framework

- W3C standard for encoding knowledge
- A fact is expressed as a *subject-predicate-object* triple or statement:

_1 has_Atomic_Number 57

- Subjects, predicates and objects are given as names for entities
- Names are URIs
- Objects can also be given as text values, called literal values, which may or may not be typed using XML Schema datatypes

n3/turtle

- URIs are expressed as <…> or abbreviated with namespaces
- Statement is just written as subject URI, predicate URI and object URI/literal value with a period:

```
@prefix pt: <http://www.example.org/PeriodicTable#> .
<http://www.example.org/PeriodicTable#La> pt:name "lanthanum" .
```

- Multiple statements with the same subject (and predicate) can be grouped together with a semicolon (commas)

```
<http://www.example.org/PeriodicTable#La> pt:name "lanthanum" ;
    pt:atomicNumber 57 ; pt:color "silvery white" .
```

rdf/xml

Resource nodes:

- Statement subjects and objects
- Usually have rdf:about attribute
- Can only contain property nodes

Property nodes:

- Literal values
- Reference to an object resource using rdf:resource attribute
- Resource node

rdf/xml example

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:pt="http://www.example.org/PeriodicTable#">
   <rdf:Description rdf:about="http://www.example.org/PeriodicTable#La">
        <pt:name> lanthanum <pt:name>
        <pt:atomicNumber> 57 </pt:atomicNumber>
        <pt:color> silvery white </pt:color>
    </rdf:Description>
</rdf:RDF>
```

For comparison:

```
@prefix pt: <http://www.example.org/PeriodicTable#> .
<http://www.example.org/PeriodicTable#La> pt:name "lanthanum" ;
    pt:atomicNumber 57 ; pt:color "silvery white" .
```

rdfa

- Embed knowledge in an (X)HTML document

- Uses attributes applicable to all elements:

-- **about**
-- **rel** and **rev**
-- **href, src** and **resource**
-- **property**
-- **content**
-- **datatype**
-- **typeof**

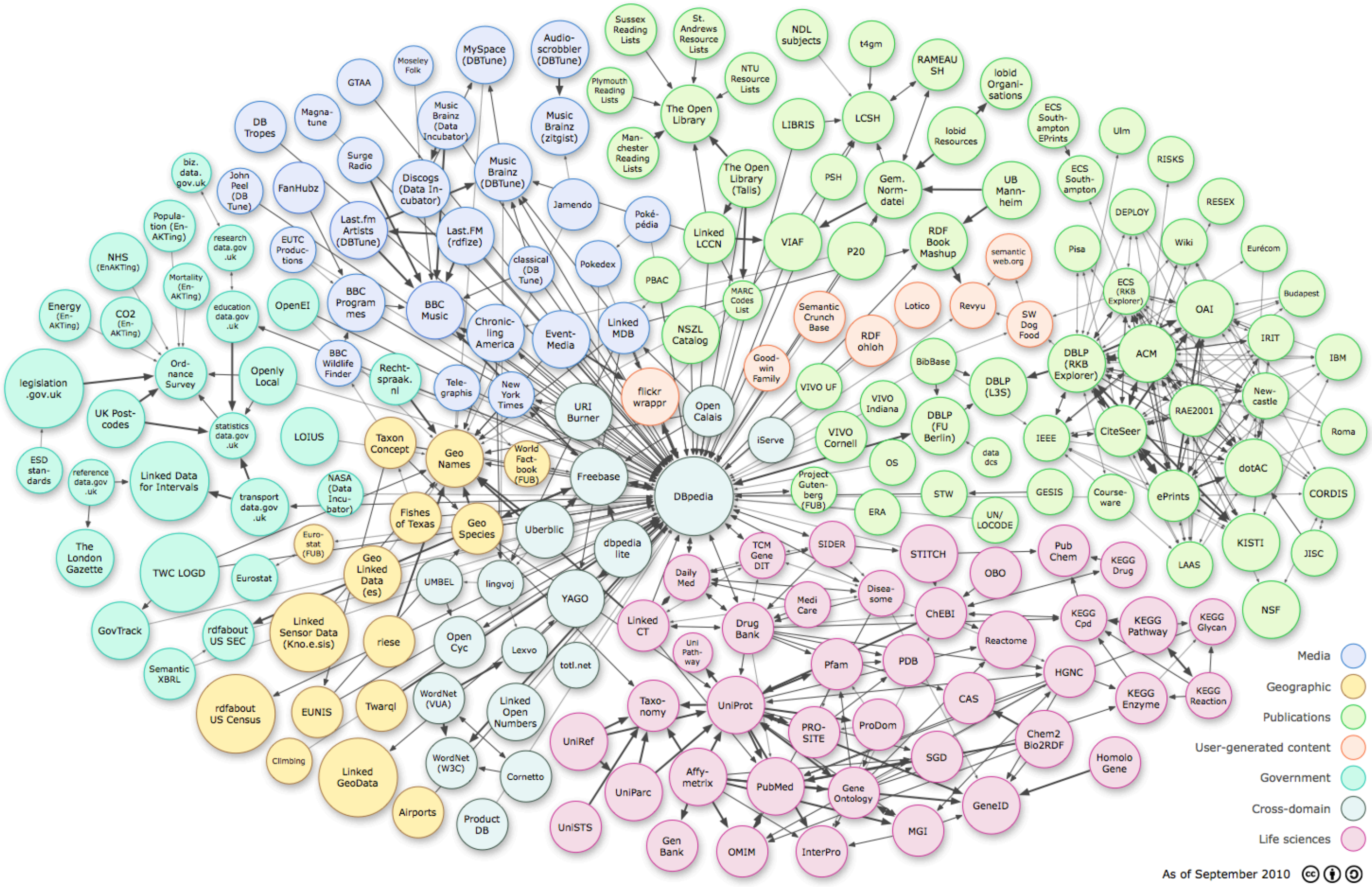
rdfa example

```
<html xmlns="http://www.w3.org/1999/xhtml"
      xmlns:pt="http://www.example.org/PeriodicTable#">
  <head>
    <title> Lanthanum </title>
    <base href="http://www.example.org/PeriodicTable#La" />
  </head>
  <body>
    <p>
      The element <span property="pt:name"> lanthanum </span> has an atomic
      weight of <span property="pt:atomicWeight"> 57 </span> and is
      <span property="pt:color"> silvery white </span> in color.
    </p>
  </body>
</html>
```

linked data

- A term coined by Tim Berners-Lee who outlined four principles:
  - Use URIs to identify things that you expose to the Web as resources
  - Use HTTP URIs so that people can locate and look up (dereference) these things
  - Provide useful information about the resource when its URI is dereferenced
  - Include links to other, related URIs in the exposed data

linked data web



linked data examples

- List all episodes of the HBO television series "The Sopranos" ordered by their air-date

- Find the official websites of companies with more than 50000 employees

- Find me things close to the Eiffel Tower

- Discover new drugs to treat Alzheimer's:
  - What proteins are involved in signal transduction **and** are related to pyramidal neurons?
    - Google: 223000 hits, 0 results
    - Linked healthcare data: 32 hits, 32 results

linked data tools

- Browsers:

  -- Tabulator (also has Firefox plugin)
  -- Disco
  -- Openlink Data Browser

- Libraries

  -- Semantic Web Client Library (Java)

- Exposing data

  -- d2r for exposing relational data in a db

concept schemes

- controlled vocabulary: a closed list of terms that can be used for classification
- taxonomy: a controlled vocabulary with hierarchy (term - sub-term)
- thesaurus: a taxonomy with broader/narrower term, synonymous term, top term, scope note and related term
- subject heading list, terminology, glossary, faceted classification

simple knowledge organization system

- W3C standard for expressing knowledge concept schemes in a machine-understandable way

- Set of RDF properties and RDFS classes to express the content and structure of a concept scheme as an RDF graph:

  - Collection
  - CollectableProperty
  - OrderedCollection
  - ConceptScheme
  - Concept
  - prefLabel
  - altLabel
  - hiddenLabel
  - definition
  - example
  - scopeNote
  - broader
  - narrower
  - related

skos example

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:skos="http://www.w3.org/2004/02/skos/core#"
 xml:base="http://www.astro.physik.uni-goettingen.de/~hessman/rdf/IVOAT">
    <skos:ConceptScheme rdf:about="">
        <skos:hasTopConcept rdf:resource="#temperatureScales"/>
    </skos:ConceptScheme>
    <skos:Concept rdf:about="#absoluteTemperatureScale">
        <skos:prefLabel>absolute temperature scale</skos:prefLabel>
        <skos:definition>absolute temperature scale</skos:definition>
        <skos:broader rdf:resource="#temperatureScales"/>
        <skos:narrower rdf:resource="#Kelvin"/>
        <skos:related rdf:resource="#celsiusTemperatureScale"/>
        <skos:related rdf:resource="#temperature"/>
    </skos:Concept>
</rdf:RDF>
```

sparql protocol and rdf query language

- W3C standard for RDF query language and data access protocol
- SELECT, CONSTRUCT, ASK, DESCRIBE
- WHERE, FILTER, OPTIONAL, UNION
- ORDER BY, DISTINCT, REDUCED, LIMIT, OFFSET

sparql example

```
PREFIX table: <http://www.example.org/PeriodicTable#>
SELECT ?name ?symbol ?weight ?number ?color
FROM <http://www.example.org/PeriodicTable.owl>
WHERE
{
    ?uranium table:name "uranium".
    ?uranium table:atomicWeight ?uraniumWeight.
    ?element table:name ?name.
    ?element table:symbol ?symbol.
    ?element table:atomicWeight ?weight.
    ?element table:atomicNumber ?number.
    OPTIONAL { ?element table:color ?color }.
    FILTER ( ?weight > ?uraniumWeight )
}
ORDER BY ASC[?weight]
```

ontologies

- An ontology is a formal specification of a shared conceptualization

-- it is a data model that represents a set of concepts within a domain and the relationships between those concepts

- Ontologies generally describe:
  - Individuals: the basic or "ground level" objects
  - Classes: sets, collections, or types of objects
  - Attributes: properties, features, characteristics, or parameters that objects can have and share
  - Relations: ways that objects can be related to one another
  - Events: the changing of attributes or relations

web ontology language (owl)

- W3C standard for authoring ontologies

- Based on RDF (OWL semantically extends RDFS)

- Regarded as one of the fundamental technologies underpinning the Semantic Web

- OWL allows descriptions of:

  -- relations between classes (e.g. disjointness)
  -- cardinality (e.g. "exactly one")
  -- characteristics of properties (e.g. symmetry)
  -- enumerated classes

owl components

- Data is interpreted as:
  - a set of **individuals**
  - a set of **property assertions** relating the individuals to each other
  - a set of **axioms** placing constraints on sets of individuals (**classes**) and the types of relationships allowed between them

- For example, the *family* ontology:
  - "hasMother" is only present between two individuals when "hasParent" is also present
  - members of "HasTypeOBlood" are never related via "hasParent" to members of "HasTypeABBlood"
  - If Ada "hasMother" Anne and Ada is "HasTypeOBlood" then Anne is not "HasTypeABBlood"

inference

Computers provide reasoning services over a knowledge domain where the domain and the knowledge have been formally and rigorously specified and reasoning algorithms have been implemented in a way which that computer can apply. For example:

- Sheep only eat grass
- Grass is a plant
- Plants and parts of plants are disjoint from animals and parts of animals
- Vegetarians only eat things which are not animals or parts of animals

=> sheep are vegetarians!

inference examples

- relevant terms for stage:

  start some (succeeded_by some stage1) and end some (preceded_by some stage2)

- related terms for term:

  part_of some class1 or develops_from some class1 or has_part from some class1

- lowest common ancestor:

  superclassOf(class1 or class2)

software

- Jena (jena.sourceforge.net)
- Protege (protege.stanford.edu)
- cwm (www.w3.org/2000/10/swap/doc/cwm.html)