

# I 202: INFORMATION ORGANIZATION & RETRIEVAL FALL 2025

---

Class 8: Ontologies; Tags vs Faceted Categories

# Today's Outline

Ontology Definition

Ontology Examples

AI-based Ontologies

Tags vs Faceted Categories

Upcoming Assignment

# Exercise:

## Create Faceted Metadata for this Image



Photo by J. Hearst, jhearst.typepad.com

# Faceted Metadata Limitation: Shows Attributes, but not Relations between Attributes



Photo by J. Hearst, jhearst.typepad.com

Aquamarine  
Red  
Orange

Door  
Doorway  
Wall

- Does not show which color is associated with which object
- Ontologies can help with this limitation

# WHAT IS AN ONTOLOGY?

Gr: *onto*: being, *logy*: science, study-of

- An information structure that often includes **relationships, constraints, and rules.**
- A method for describing **knowledge** in such a way that it can be process by a computer.
- The branch of metaphysics dealing with the nature of being (we are not using this sense of the word)

# Taxonomy vs Ontology

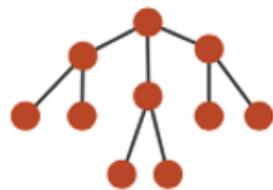
## Taxonomy

- Tree / Hierarchical in structure
- Relation is consistent throughout and signifies inclusion
  - *(type-of, part-of)*

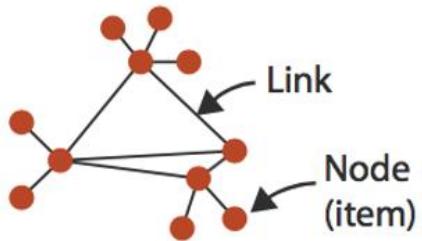
## Ontology

- Graph / Network in structure
- Relations can link objects to each other *without* inclusion
- Relations can represent attributes
- Often includes rules and constraints

# Tree vs. Network / Graph



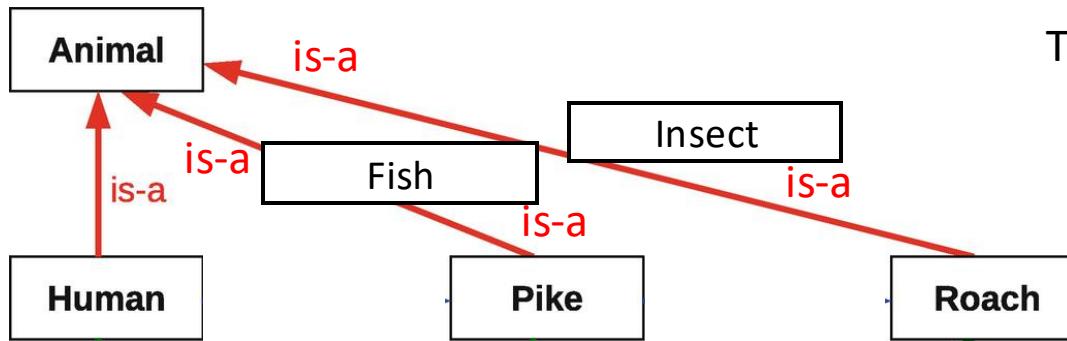
tree



graph

BASIS FOR COMPARISON	TREE	GRAPH
Root node	It has exactly one root node.	Graph doesn't have a root node.
Cycles	No cycles are permitted.	Graph can have cycles.
Complexity	Less complex	More complex comparatively
Number of edges	$n-1$ ( $n = \#$ of nodes)	Not defined
Model type	Hierarchical	Network

# ONTOLOGY EXAMPLE: INFERENCES FROM A TAXONOMY



This is a partial taxonomy

We can make inferences from taxonomies, using the implicit is-a rule:

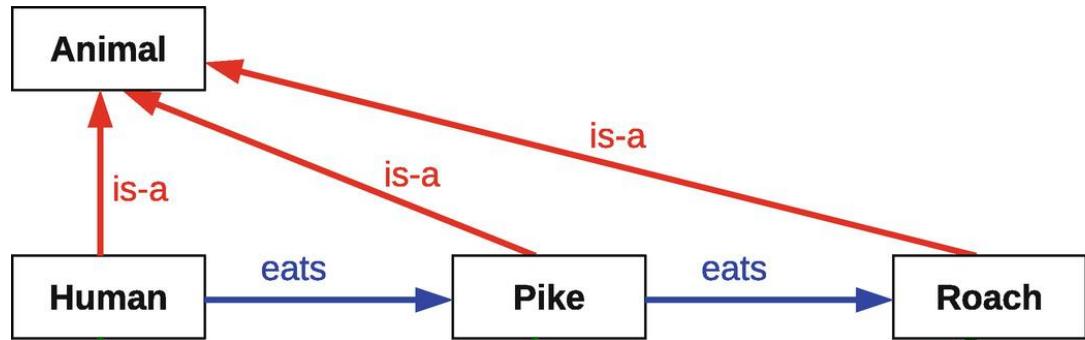
$$isa(a, b) \text{ AND } isa(b, c) \rightarrow isa(a, c)$$

In English:

A pike is a fish; a fish is an animal, therefore a pike is an animal.

# ONTOLOGY EXAMPLE

## RELATION LINKS ALLOW ADDITIONAL INFERENCES



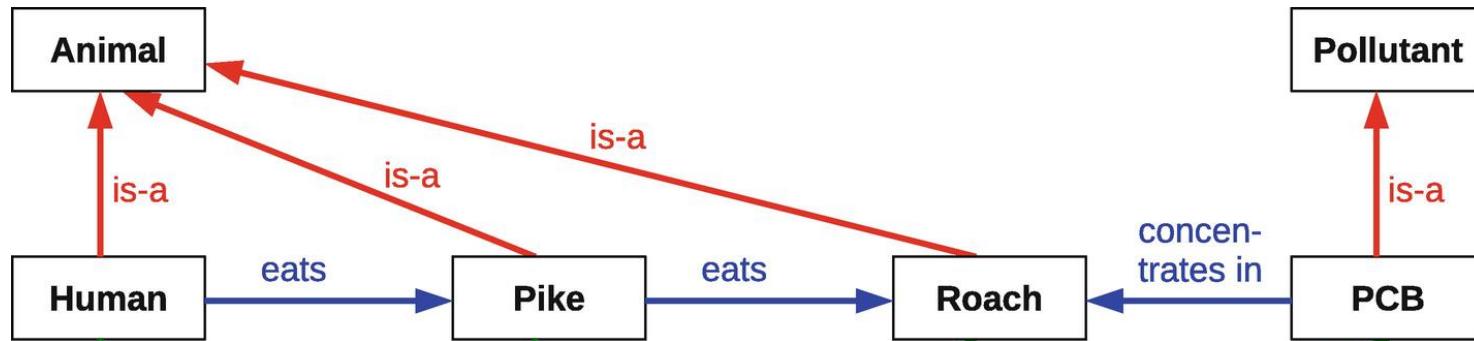
We've added eats() relation links; Now it is an ontology

We can now make new rule and inferences; create new knowledge

$$\text{eats}(a, b) \text{ AND } \text{eats}(b, c) \rightarrow \text{eats}(a, c)$$

Humans (indirectly) eat roaches

# ONTOLOGY EXAMPLE: ADDING MORE INFERENCE RULES



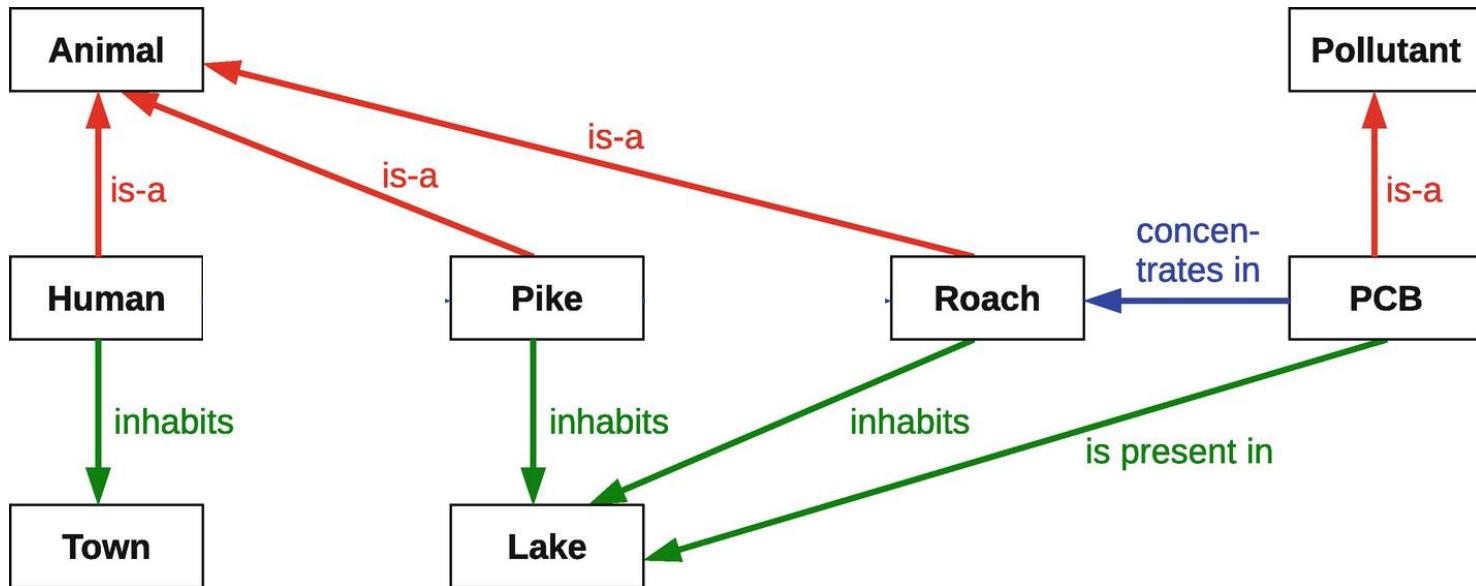
More rules, more inferences

$$\text{concentratedIn}(p, a) \text{ AND } \text{eat}(b, a) \rightarrow \text{poisonedBy}(b, p)$$
$$\text{eats}(a, b) \text{ AND } \text{eats}(b, c) \rightarrow \text{eats}(a, c)$$

Can conclude:

Pike poisonedBy PCB, therefore  
Human poisonedBy PCB

# ONTOLOGY EXAMPLE: ADD MORE RELATION TYPES



# ONTOLOGY EXAMPLE: IMDB



## Movies

- [Release Calendar](#)
- [DVD & Blu-ray Releases](#)
- [Top Rated Movies](#)
- [Most Popular Movies](#)
- [Browse Movies by Genre](#)
- [Top Box Office](#)
- [Showtimes & Tickets](#)
- [In Theaters](#)
- [Coming Soon](#)
- [Movie News](#)
- [India Movie Spotlight](#)

## TV Shows

- [What's on TV & Streaming](#)
- [Top Rated Shows](#)
- [Most Popular Shows](#)
- [Browse TV Shows by Genre](#)
- [TV News](#)
- [India TV Spotlight](#)

## Awards & Events

- [Oscars](#)
- [Best Picture Winners](#)
- [Golden Globes](#)
- [Emmys](#)
- [San Diego Comic-Con](#)
- [New York Comic-Con](#)
- [Sundance Film Festival](#)
- [Toronto Int'l Film Festival](#)
- [Awards Central](#)
- [Festival Central](#)
- [All Events](#)

## Watch

- [What to Watch](#)
- [Latest Trailers](#)
- [IMDb TV](#)
- [IMDb Originals](#)
- [IMDb Picks](#)
- [IMDb Podcasts](#)

## Celebs

- [Born Today](#)
- [Most Popular Celebs](#)
- [Celebrity News](#)

## Community

- [Help Center](#)
- [Contributor Zone](#)
- [Polls](#)



# Lauren Bacall (1924–2014)

Actress | Soundtrack



2:47 | Trailer

31 VIDEOS | 549 IMAGES

Lauren Bacall was born Betty Joan Perske on September 16, 1924, in New York City. She was the daughter of Natalie Weinstein-Bacal, a Romanian Jewish immigrant, and William Perske, who was born in New Jersey, to Polish Jewish parents. Her family was middle-class, with her father working as a salesman and her mother as a secretary. They divorced ... [See full bio »](#)

**Born:** September 16, 1924 in The Bronx, New York City, New York, USA

**Died:** August 12, 2014 (age 89) in New York City, New York, USA

**Nominated for 1 Oscar.** Another 21 wins & 13 nominations. [See more awards »](#)

## Photos



549 photos | 31 videos »

## Known For



## Quick Links

[Biography](#)

[Awards](#)

[Photo Gallery](#)

[Filmography \(by Job\)](#)

[Trailers and Videos](#)

[Explore More](#)

Top 10 Moments From the 2020 Emmy Awards



From presenters in hazmat suits to record-breaking wins, our editors name the best moments and biggest surprises from the 72nd Primetime Emmy Awards.

[See the full list »](#)

Share this page:

## Related News

Memo to Oscars: Bring back Governors Awards even if they have to be virtual  
21 September 2020 | Gold Derby

John Wayne: Top 10 Best Westerns  
04 September 2020 | Gold Derby

The Best Celebrity Memoirs, from Carrie Fisher to Katharine Hepburn  
19 August 2020 | Indiewire

# IMDB Entry: Actor

## Many relation types!

Acted in  
Photos of  
Biography of  
Known for

Related News

All actors born on Sept 16

All actors born in 1924

All actors with same astrological sign

Nicknames

Quotes

# + To Have and Have Not (1944)

★ 7.8 /10  
30,365 | Rate This

Passed | 1h 40min | Adventure, Comedy, Film-Noir | 20 January 1945 (USA)



2:47 | Trailer

1 VIDEO | 62 IMAGES

During World War II, American expatriate Harry Morgan helps transport a French Resistance leader and his beautiful wife to Martinique while romancing a sensuous lounge singer.

Director: Howard Hawks

Writers: Ernest Hemingway (novel), Jules Furthman (screen play) | 1 more credit »

Stars: Humphrey Bogart, Lauren Bacall, Walter Brennan | See full cast & crew »

Watch on Prime Video ▾  
rent/buy from \$2.99

...

+ Add to Watchlist

Reviews

164 user | 93 critic



Popularity  
4,829 (• 4,191)

IMDbPro View production, box office, & company info ▾

1 win. See more awards »

# IMDB Entry: Performance

## Many relation types!

Directors

Writers

Actors

Screenwriters

Awards

Reviews

Box office statistics

Videos

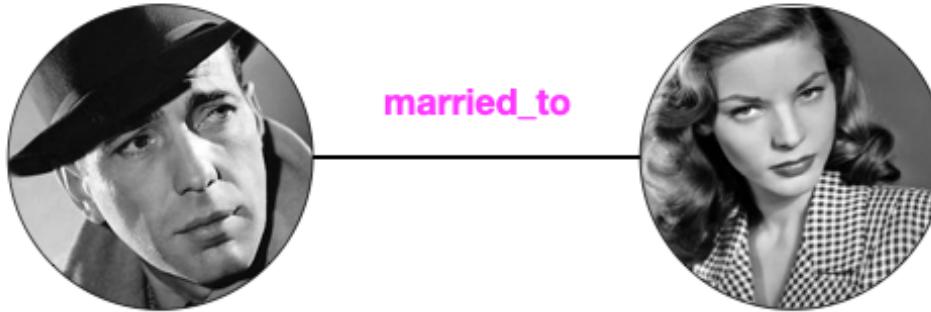
Photos

Plot

Quotes

Soundtracks

# IMDB



**married\_to(x,y) is a symmetric relation that holds between “resources”**

$$\text{married\_to}(x,y) = \text{married\_to}(y,x)$$

# IMDB



stars\_in

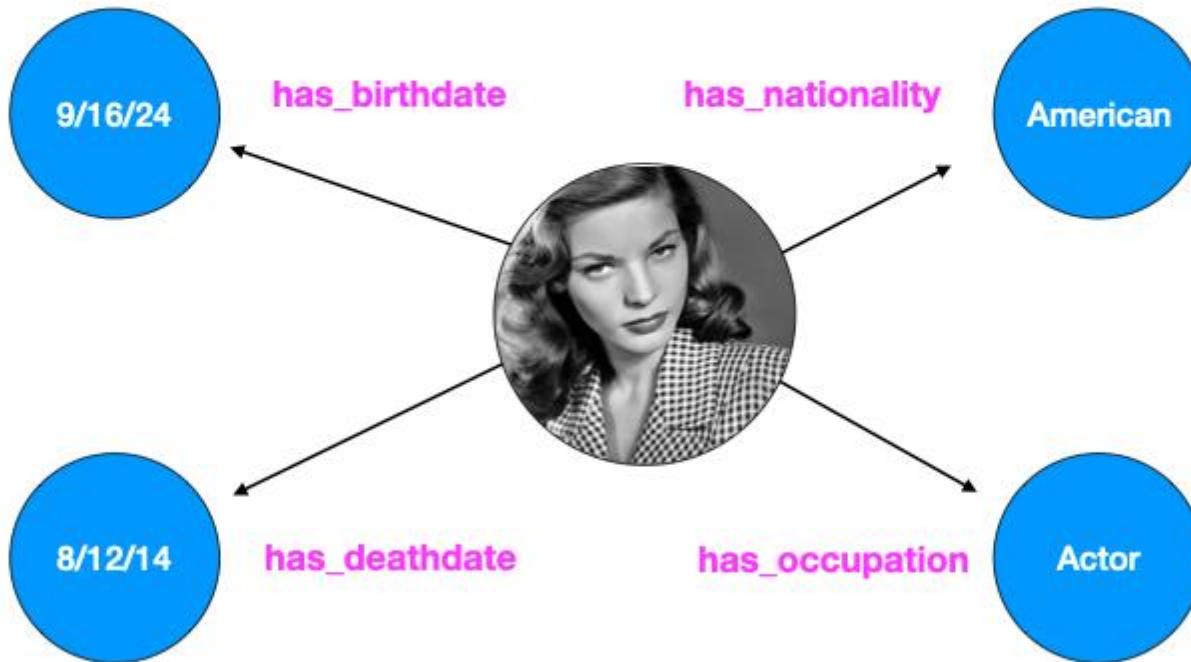


stars\_in



**stars\_in** is an  
asymmetric relation  
between x and y

# IMDB



# USES FOR DIGITAL ONTOLOGIES

- **Store and re-use knowledge**
  - *WordNet (will see next week)*
- **Link Knowledge Across Different Sources**
  - *Semantic Web*
  - *Knowledge Graphs*
- **Automated Reasoning**
  - *Formal logic, theorem provers, inference engines*
  - *Hermit OWL reasoner, Pellet reasoner*
  - *Graph-based Neural Net reasoning*

# UNDERSTANDING & NAVIGATING ONTOLOGIES

- Ontologies are difficult to show visually
- They are also difficult to convey how to navigate
- Most websites provide a simplified view
- Inference is not usually done where users can see it, but instead used behind the scenes to generate information
- Inference is usually much slower than what a relational database can do equivalently

## EXERCISE: LINKED-IN

- What are the relations?
- What inferences can be done with them?
- How does the interface express these?

# INFERRING ONTOLOGIES FROM DATA

- We can impose connections by following links
- The “6 degrees of Kevin Bacon game”
- Play it with Wikipedia links to find new knowledge

# SIX DEGREES OF WIKIPEDIA

Find the shortest paths from

Unobtainium

to

Andromeda Galaxy

Go!

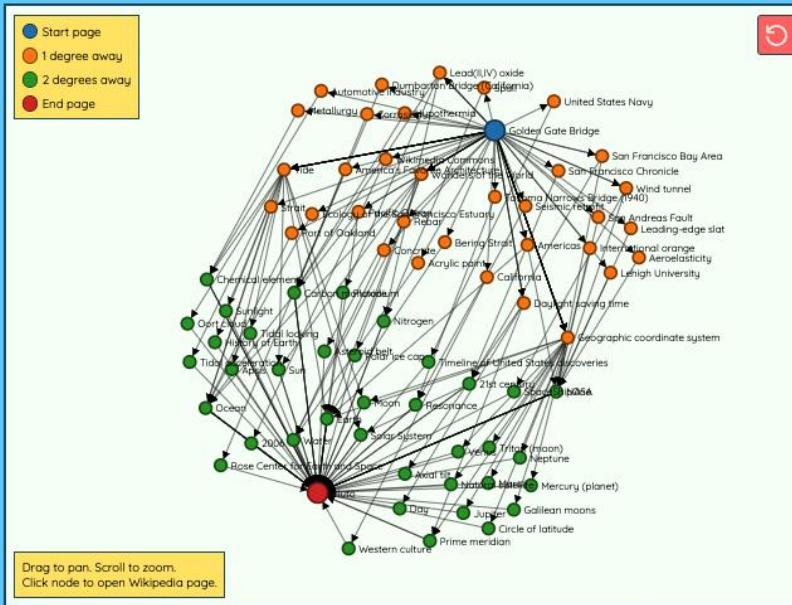
Find the shortest paths from

## golden gate bridge

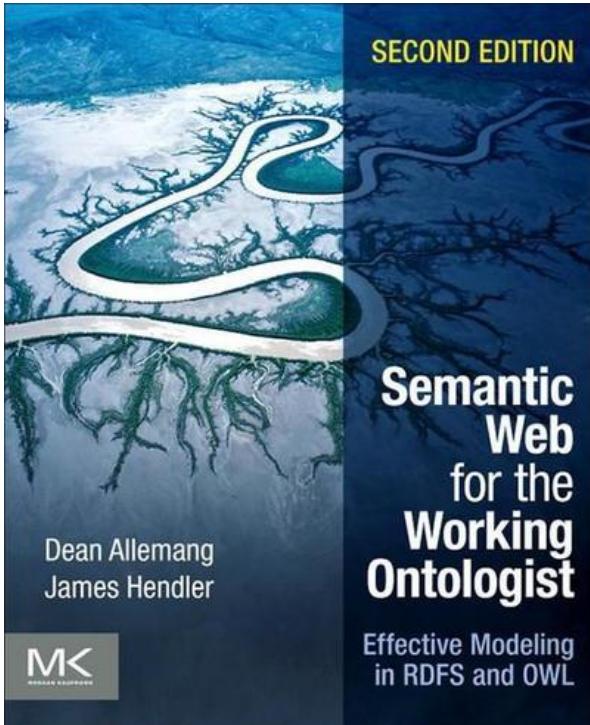
pluto

Go!

Found **69** paths with **3 degrees** of separation from  
**Golden Gate Bridge** to **Pluto** in **3.53 seconds!**



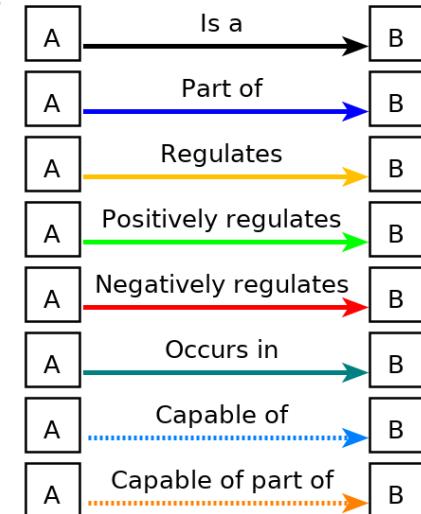
# THE SEMANTIC WEB



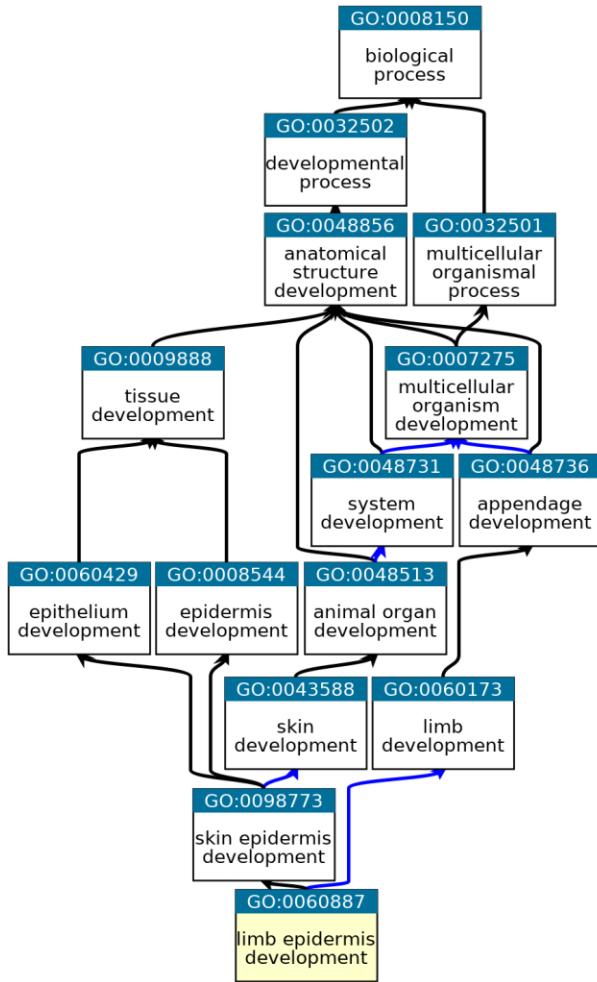
- Goal: support a distributed Web at the level of the data rather than at the level of the presentation.
- A data item can point to another, using global references called Uniform Resource Identifiers (URIs).
- Information about an entity can be distributed over the entire web.
- A way to achieve inter-operability
- Lots of people work on it, but it doesn't seem to ever really take off in practical applications

# GENE ONTOLOGY (GO)

- Annotates gene function: molecular function, cellular components, biological processes
- Under active development for > 25 years!
- 40,000 terms
- 134,000 relations
- >9M annotated gene products



# Go Ontology Example



## Taxon Constraints

The use of this term should conform to the following taxon constraints:

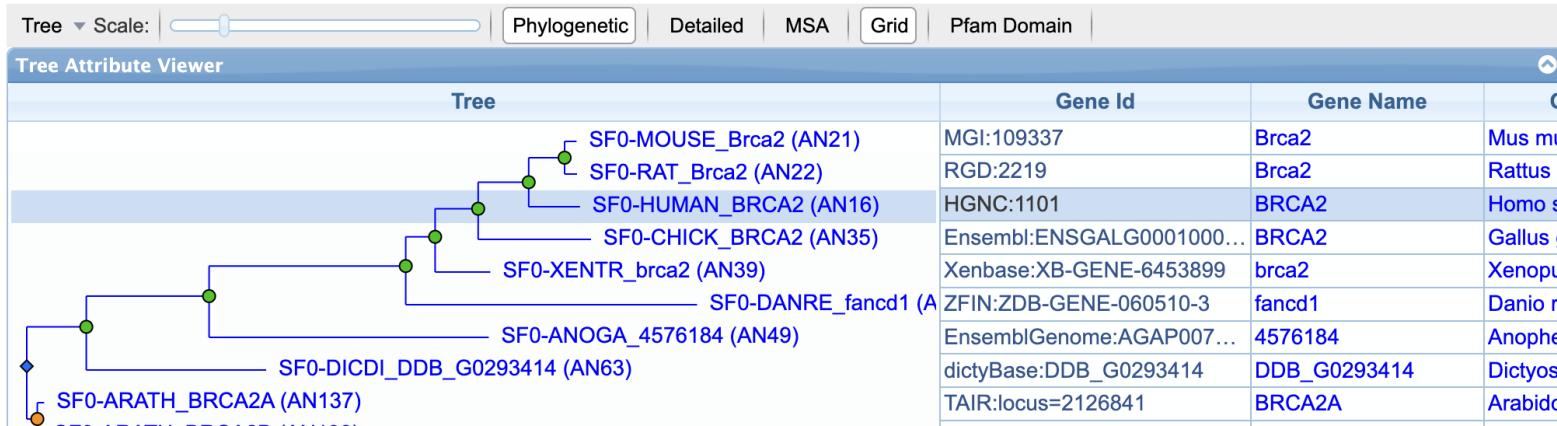
Ancestor GO ID	Ancestor GO Term Name	Relationship	Taxon ID	Taxon
GO:0008544	epidermis development	Never in Taxon	33090	Viridiplantae
GO:0009888	tissue development	Never in Taxon	147554	Schizosaccharomycetes
GO:0009888	tissue development	Never in Taxon	33630	Alveolata
GO:0009888	tissue development	Never in Taxon	33682	Euglenozoa
GO:0009888	tissue development	Never in Taxon	38254	Glaucocystophyceae
GO:0009888	tissue development	Never in Taxon	4891	Saccharomycetes

# BRCA1 Gene Example

## PANTHER TREE VIEWER

Family: BREAST CANCER TYPE 2 SUSCEPTIBILITY PROTEIN BRCA2 (PTHR11289)

Reduced Tree View



# BRCA1 Gene Example

PANTHER TREE VIEWER 



Family: BREAST CANCER TYPE 2 SUSCEPTIBILITY PROTEIN BRCA2 (PTHR11289)

Reduced Tree View

Tree	Gene Id	Gene Name	Organism	Definition	Subf
Eukaryota					BREAS
Unikonta					BREAS
Bilateria					BREAS
Euteleostomi					BREAS
Tetrapoda					BREAS
Amniota					BREAS
Euarchontoglires					BREAS
Murinae					BREAS
<i>Mus musculus</i>	MGI:109337	Brca2	<i>Mus musculus</i>	Breast cancer type 2...	BREAS
<i>Rattus norvegicus</i>	RGD:2219	Brca2	<i>Rattus norvegicus</i>	Breast cancer type 2...	BREAS
<i>Homo sapiens</i>	HGNC:1101	BRCA2	<i>Homo sapiens</i>	Breast cancer type 2...	BREAS
<i>Gallus gallus</i>	Ensembl:ENSGALG0001000...	BRCA2	<i>Gallus gallus</i>	BRCA2, DNA repair ...	BREAS
<i>Xenopus tropicalis</i>	Xenbase:XB-GENE-6453899	brca2	<i>Xenopus tropicalis</i>	BRCA2, DNA repair...	BREAS
<i>Danio rerio</i>	ZFIN:ZDB-GENE-060510-3	fancd1	<i>Danio rerio</i>	BRCA2 DNA repair...	BREAS
<i>Anopheles gambiae</i>	EnsemblGenome:AGAP007...	4576184	<i>Anopheles gambiae</i>	AGAP007032-PA	BREAS
<i>Dictyostelium discoideum</i>	dictyBase:DDB_G0293414	DDB_G0293414	<i>Dictyostelium discoid...</i>	Breast cancer type 2...	BREAS
AN136					BREAS
<i>Arabidopsis thaliana</i>	TAIR:locus=2126841	BRCA2A	<i>Arabidopsis thaliana</i>	Protein BREAST CA...	BREAS
<i>Arabidopsis thaliana</i>	TAIR:locus=2149800	BRCA2B	<i>Arabidopsis thaliana</i>	Protein BREAST CA...	BREAS

# BRCA1 Gene Example: Disease Associations

## Disease Associations

Compare Ortholog Genes

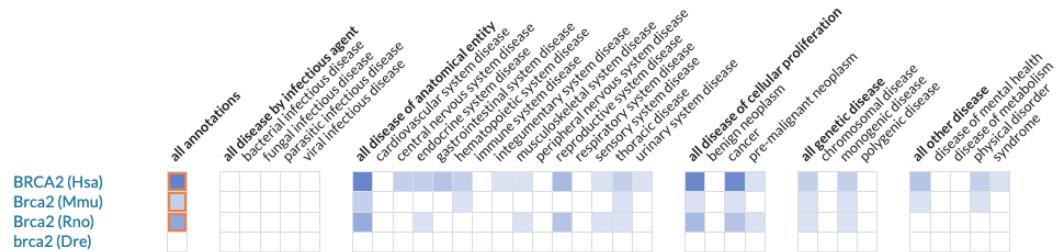
Stringency: Stringent ▾

Species ▾

?

Include Negative Annotations

Cases where the expected disease association was NOT found

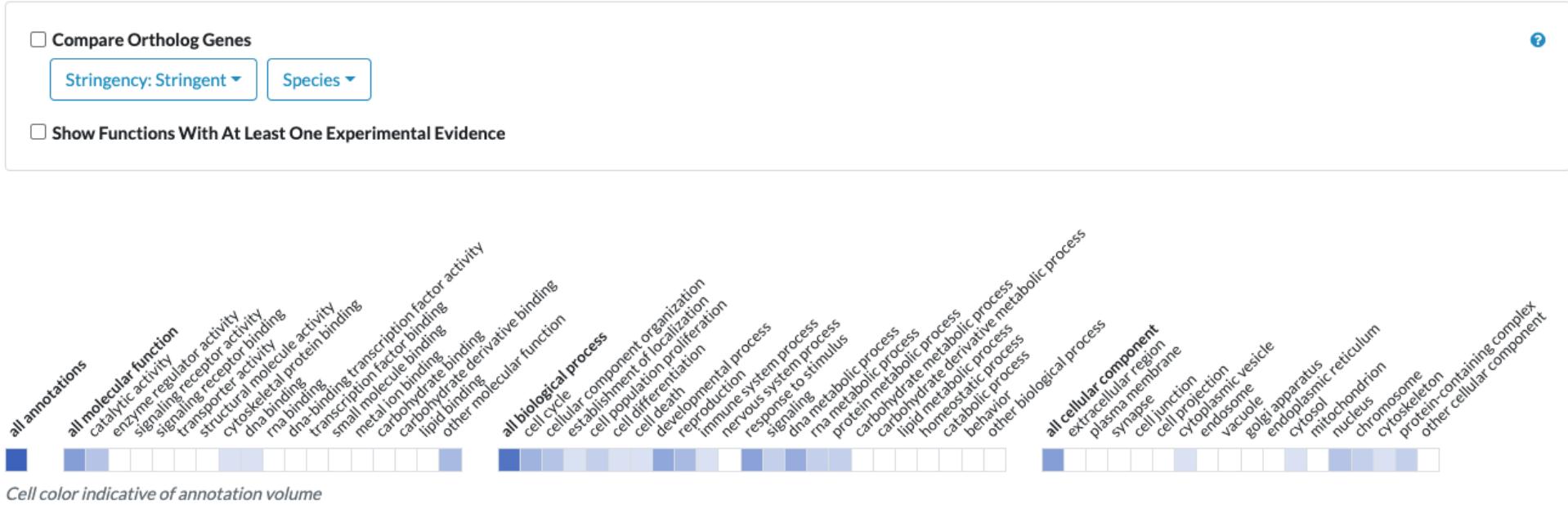


Cell color indicative of annotation volume

Species ▾	Gene ▾	Association ▾	Disease ▾	Evidence ▾	Source ▾	Based On ▾	References ▾
<i>Homo sapiens</i>	<b>BRCA2</b>	is implicated in	breast cancer	IAGP	OMIM  via RGD		RGD:7240710
<i>Homo sapiens</i>	<b>BRCA2</b>	is marker for	colorectal cancer	IEP	RGD		PMID:16533773
<i>Homo sapiens</i>	<b>BRCA2</b>	is marker for	colorectal carcinoma	IEP	RGD		PMID:11497291
<i>Homo sapiens</i>	<b>BRCA2</b>	is implicated in	ductal carcinoma in situ	IAGP	RGD		PMID:16280055
<i>Homo sapiens</i>	<b>BRCA2</b>	is implicated in	endometrial cancer	IAGP	RGD		PMID:10451700
<i>Homo sapiens</i>	<b>BRCA2</b>	is implicated in	endometrial serous adenocarcinoma	IAGP	RGD		PMID:16650962
<i>Homo sapiens</i>	<b>BRCA2</b>	is implicated in	esophagus squamous cell carcinoma	IAGP	RGD		PMID:21279724

# BRCA1 Example

## Function - GO Annotations ?



# BRCA1 Example

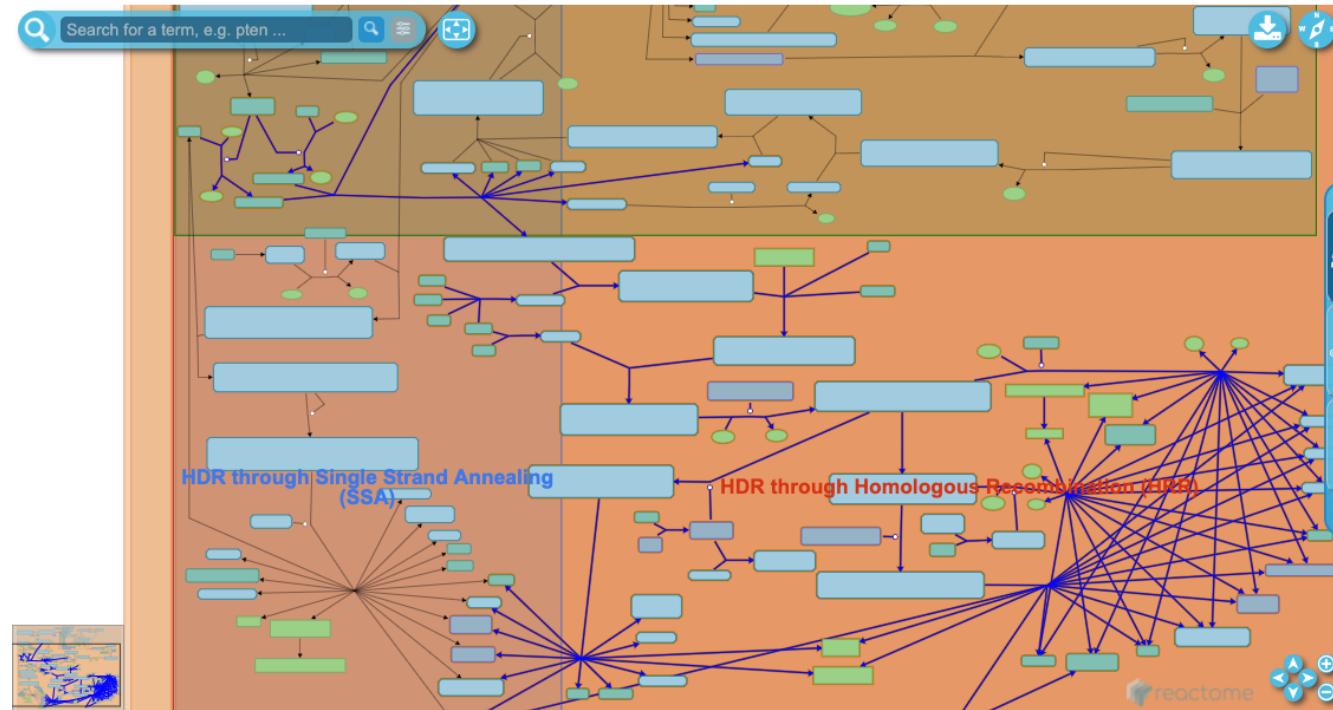
## Pathways ?

Reactome Pathway (9)

Reactome Reactions (20)

GO-CAMs (0)

Available pathways: HDR through Homologous Recombination (HRR)



# BRCA1 Example

## Alleles and Variants

---



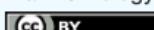
# MAINTAINING THE GO KNOWLEDGE BASE

- Experts read papers and enter information
- Other experts can report on problems
- The standards used have evolved over time, reflecting improvements in knowledge representation
- Algorithms analyze the data to help researchers analyze results and formulate hypotheses
- Visualizations have been developed to simplify the complexity

# EXERCISE: EXPLORE THE GENE ONTOLOGY

- Search for quickgo, click on the link
  - <https://www.ebi.ac.uk/QuickGO/annotations>
- Explore one of the genes there
- See what kinds of relations are supported
- Follow links, look at the graph structures, the Ontology tab

Download table as: [ [YAML](#) | [JSON-LD](#) | [RDF/Turtle](#) ]

bfo	Basic Formal Ontology 	The upper level ontology upon which OBO Foundry ontologies are built. <a href="#">Detail</a>	       
chebi	Chemical Entities of Biological Interest 	A structured classification of molecular entities of biological interest focusing on 'small' chemical compounds. <a href="#">Detail</a>	       
doid	Human Disease Ontology 	An ontology for describing the classification of human diseases organized by etiology. <a href="#">Detail</a>	       
go	Gene Ontology 	An ontology for describing the function of genes and gene products <a href="#">Detail</a>	       
obi	Ontology for Biomedical Investigations 	An integrated ontology for the description of life-science and clinical investigations <a href="#">Detail</a>	       
pato	Phenotype And Trait Ontology 	An ontology of phenotypic qualities (properties, attributes or characteristics) <a href="#">Detail</a>	       
po	Plant Ontology 	The Plant Ontology is a structured database resource that links plant and growth and development to p	  
pr	PRotein Ontology (PRO) 	An ontological representation of p <a href="#">Detail</a>	  

A Collection of Ontologies



# Principles: Overview

- [Overview](#)
- [Open \(principle 1\)](#)
- [Common Format \(principle 2\)](#)
- [URI/Identifier Space \(principle 3\)](#)
- [Versioning \(principle 4\)](#)
- [Scope \(principle 5\)](#)
- [Textual Definitions \(principle 6\)](#)
- [Relations \(principle 7\)](#)
- [Documentation \(principle 8\)](#)
- [Documented Plurality of Users \(principle 9\)](#)
- [Commitment To Collaboration \(principle 10\)](#)
- [Locus of Authority \(principle 11\)](#)
- [Naming Conventions \(principle 12\)](#)
- [Maintenance \(principle 16\)](#)

[View](#)[Edit](#)

This page is generated via  
[\\_layouts/principle.html](#). See [edit guide](#)

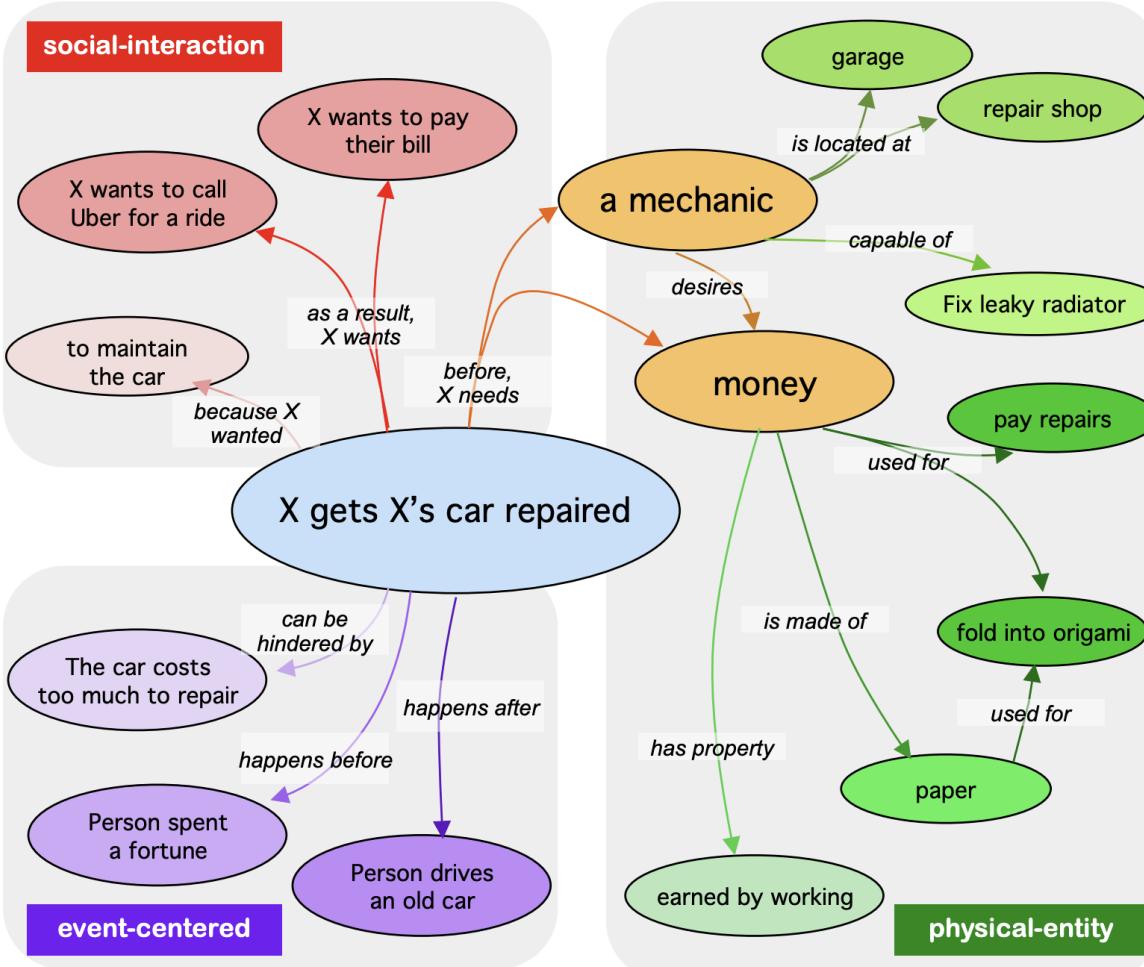
These principles are intended as normative for OBO Foundry ontologies, and ontologies submitted for review will be evaluated according to them. We consider these to be generally good practice, and recommend they be considered even if there are no plans to submit an ontology for review by the Foundry. Where we use capitalized words such as "MUST", and "SHOULD", they will be interpreted according to [RFC 2119: Key words for use in RFCs to Indicate Requirement Levels](#) when the principles are applied during reviews of ontologies for inclusion in the Foundry.

There is currently an ongoing process to clarify the wording of the principles and expand on their purpose, implementation, and criteria to be used to evaluate ontologies for compliance with each principle. We have gone through this process for the following principles: open, common format, users, collaboration, locus of authority. The new wording has been posted for these and we are continuing to work on the others. Please use the [issue tracker](#) to let us know if there are further clarifications that you would like to see addressed for any of the principles.

## Principles for Ontology Creation

# AI-BASED COMMON-SENSE REASONING

- Pre-LLM work in natural language processing
- Makes use of large datasets and deep neural nets
- Getting some interesting results now





# Mosaic Knowledge Graphs

Model

People &amp; Events (COMeT 2020)

Entities (COMeT 2020)

Images (VisualCOMET)

Knowledge Graph

ATOMIC 2020

About

## Commonsense Inferences about Entities (COMmonsensE Transformers on Atomic2020)

A commonsense knowledge graph with 1.33M everyday inferential knowledge tuples about entities and events. It represents a large-scale common sense repository of textual descriptions that encode both the social and the physical aspects of common human everyday experiences. We also provide COMET-ATOMIC-2020 model which can generate representative knowledge for new, unseen entities and events. | [read the paper](#) / [download the data](#)

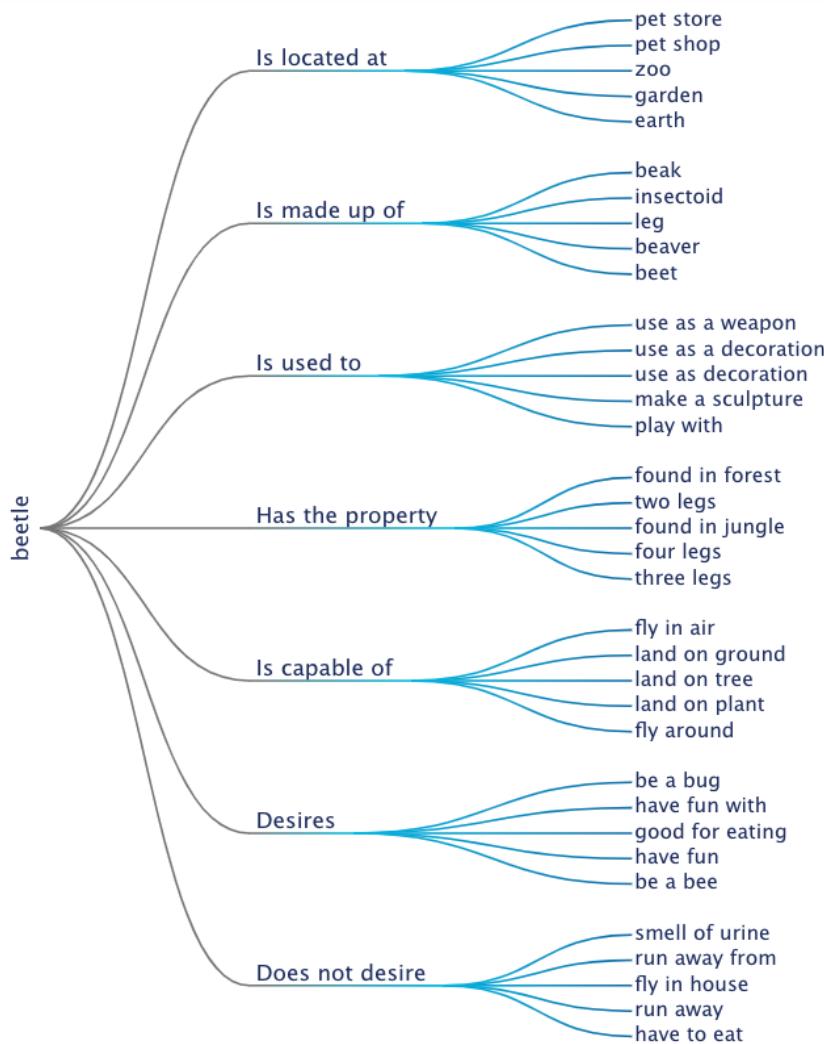
## Explore

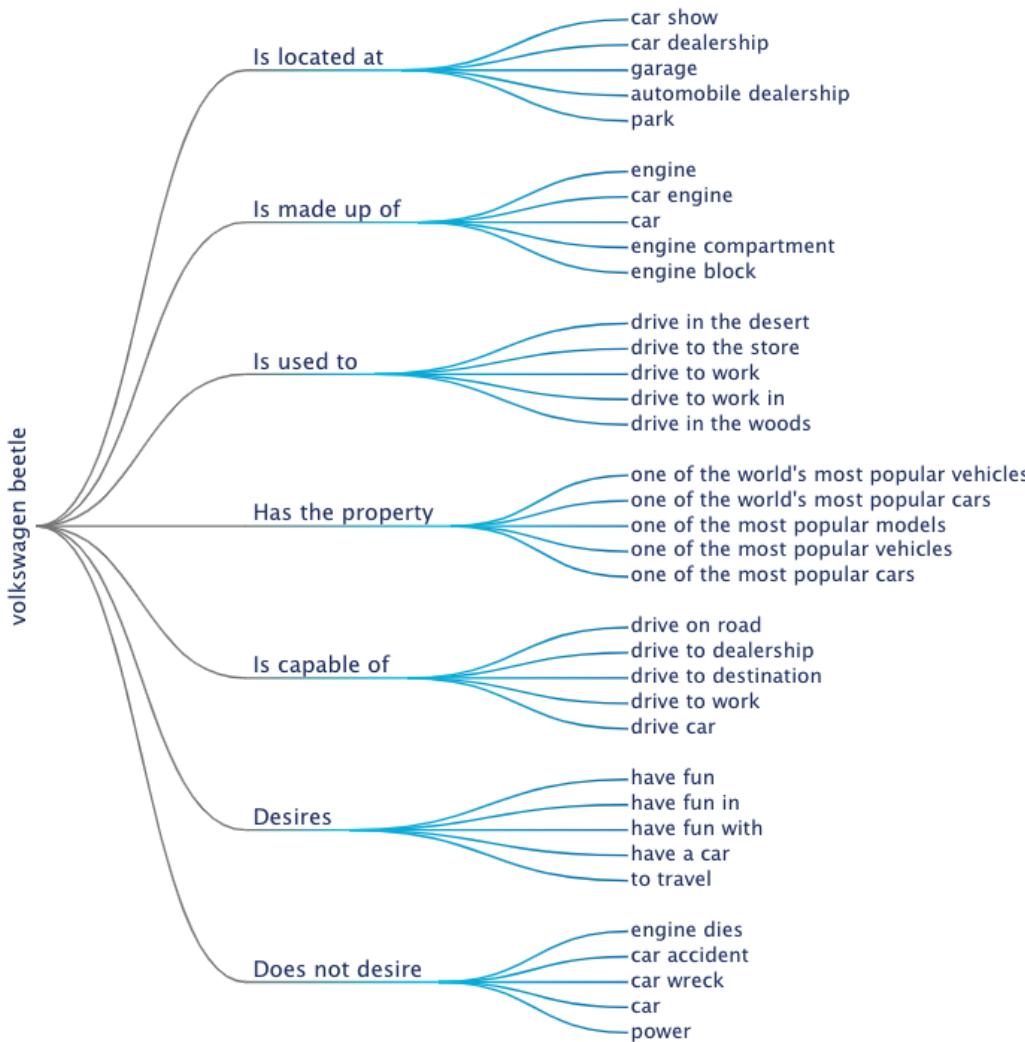
Predict

Try: [car](#), [student](#), [brain](#), [artificial intelligence](#)

Project link no longer active

[https://mosaickg.apps.allenai.org/model\\_comet2020\\_entities](https://mosaickg.apps.allenai.org/model_comet2020_entities)





# Commonsense Inferences about People and Events

(COMmonsensE Transformers on  
Atomic2020)

A commonsense knowledge graph with 1.33M everyday inferential knowledge tuples about entities and events. It represents a large-scale common sense repository of textual descriptions that encode both the social and the physical aspects of common human everyday experiences. We also provide COMET-ATOMIC-2020 model which can generate representative knowledge for new, unseen entities ... [Expand](#) | [read the paper](#) / [download the data](#)

## Explore

Predict

Try: PersonX acts quickly, PersonX is a big deal, My boss is very good

---

Select one or more relations.

Select all

PersonX is seen as

Before, PersonX needed

As a result, PersonX wants

As a result, others want

Is hindered by

PersonX then

As a result, PersonX feels

Others then

Happens before

Causes

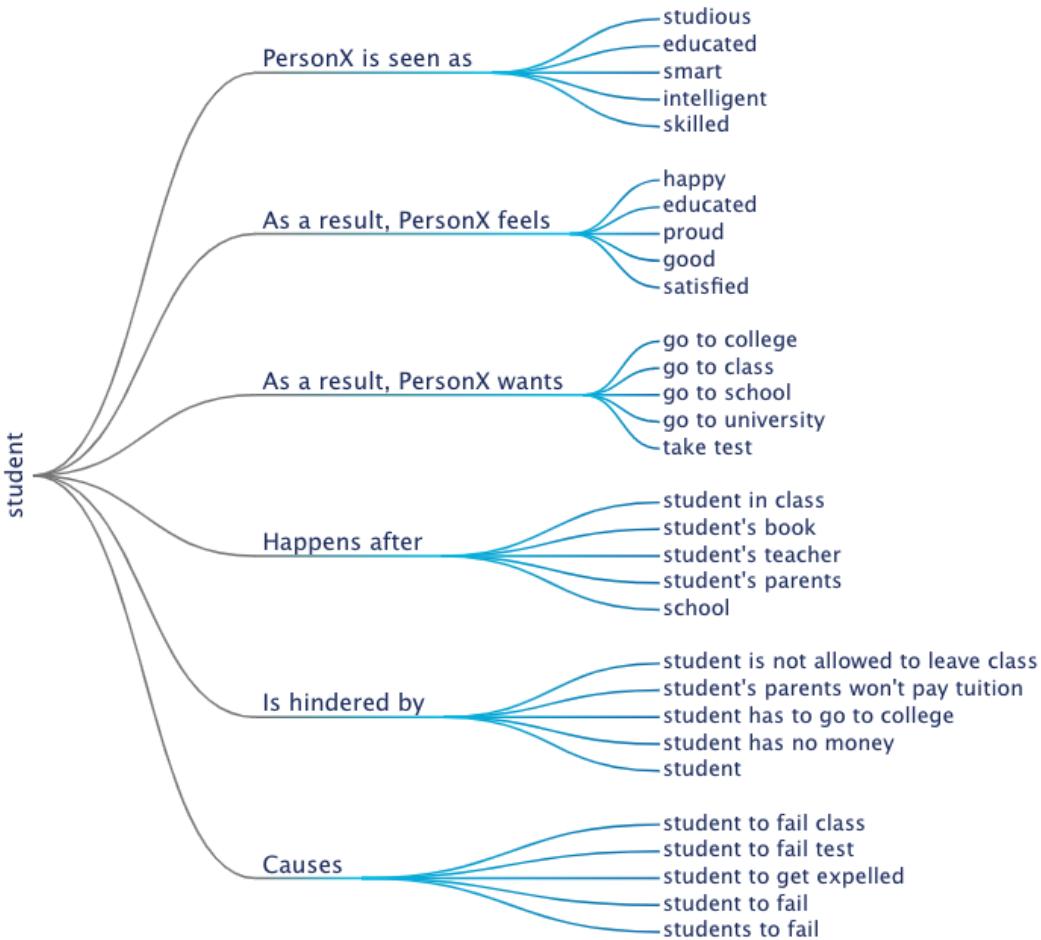
Because PersonX wanted

As a result, PersonX reasons

As a result, others feel

Happens after

The model has predicted these relationships for 'student'



# SUMMARY: ONTOLOGIES

- Goes beyond taxonomies and faceted metadata to express relationships, constraints, and rules
- Useful (maybe) for computer-assisted analysis
- Complex to build and maintain
- Difficult to show in an interface for everyday use
- AI methods are making progress in automating creation

# TAGS VS FACETED CATEGORIES

# Faceted Categories vs Tags

## Faceted categories

- Multiple facets per item
- Hierarchical (often)
- Intentionally Constructed Organization
- Controlled vocabulary
- Collection owner

## Tags

- Multiple tags per item
- Flat categories (usually)
- Any term is ok
- User-driven

Home

PUBLIC

Stack Overflow

Tags

Users

FIND A JOB

Jobs

Companies

TEAMS What's this?

 Free 30 Day Trial

# Tags

A tag is a keyword or label that categorizes your question with other, similar questions. Using the right tags makes it easier for others to find and answer your question.

[Show all tag synonyms](#) Filter by tag name[Popular](#) [Name](#) [New](#)**javascript**

For questions regarding programming in ECMAScript (JavaScript/JS) and its various dialects/implementations (excluding ActionScript). This tag is...

2086025 questions 605 asked today, 5388 this week

**java**

a popular high-level programming language. Use this tag when you're having problems using or understanding the language itself. Th...

1711975 questions 324 asked today, 3104 this week

**python**

a multi-paradigm, dynamically typed, multipurpose programming language. It is designed to be quick to learn, understand, and use, and enforce a...

1534354 questions 842 asked today, 6957 this week

**c#**

a high level, statically typed, multi-paradigm programming language developed by Microsoft. C# code usually targets Microsoft's .NET famil...

1435109 questions 193 asked today, 2177 this week

**php**

a widely used, high-level, dynamic, object-oriented and interpreted scripting language primarily designed for server-side web development. Us...

1370376 questions 210 asked today, 1849 this week

**android**

Google's mobile operating system, used for programming or developing digital devices (Smartphones, Tablets, Automobiles, TVs, Wear, Glass, IoT)....

1298309 questions 275 asked today, 2169 this week

**html**

the main markup language for creating web pages and other information to be displayed in a web browser. Questions regarding HTML should include a...

1021249 questions 243 asked today, 2143 this week

**jquery**

a JavaScript library, consider also adding the JavaScript tag. jQuery is a popular cross-browser JavaScript library that facilitates Document Obj...

996899 questions 83 asked today, 830 this week

**c++****css****ios****mysql**

Home

PUBLIC

Stack Overflow

Tags

Users

FIND A JOB

Jobs

Companies

TEAMS

What's this?

Free 30 Day Trial

# Top Questions

[Ask Question](#)[Interesting](#)

374

[Bountied](#)[Hot](#)[Week](#)[Month](#)

0  
votes      0  
answers      2  
views

[Sqlalchemy one to many relationship in asyncio](#)[python](#) [sqlalchemy](#) [python-asyncio](#) [one-to-many](#) asked 23 secs ago [Dilshod Bakhtiyorov](#) 1

0  
votes      0  
answers      7  
views

[How to slowly extend the expiry time of a cache?](#)[php](#) [wordpress](#) [caching](#) modified 46 secs ago [Ben Turner](#) 143

0  
votes      0  
answers      2  
views

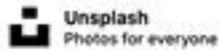
[OpenCV: Find co-ordinates of trapezium with rounded corner](#)[python](#) [opencv](#) [opencv-python](#) asked 1 min ago [sameep baraiya](#) 65

0  
votes      3  
answers      30  
views

[why does it not allow me to print "player 1 is x"](#)[python](#) [if-statement](#) [input](#) [while-loop](#) [boolean](#) answered 1 min ago [chanakya kasper](#) 1

0  
votes      0  
answers      2  
views

[How and where to store data once it is clustered by k-mean](#)[cluster-analysis](#) [k-means](#) [storing-data](#) [storing-information](#) asked 1 min ago [user2782345](#) 7



Search free high-resolution photos

Topics Explore +

Submit a photo

Editorial FEATURED Work From Home FEATURED Technology Wallpapers Travel Nature Film Textures & Patterns Current Events People Business & Work COVID-19 Animals In

# Unsplash

The internet's source of freely-usable images.

Powered by creators everywhere.

Search free high-resolution photos

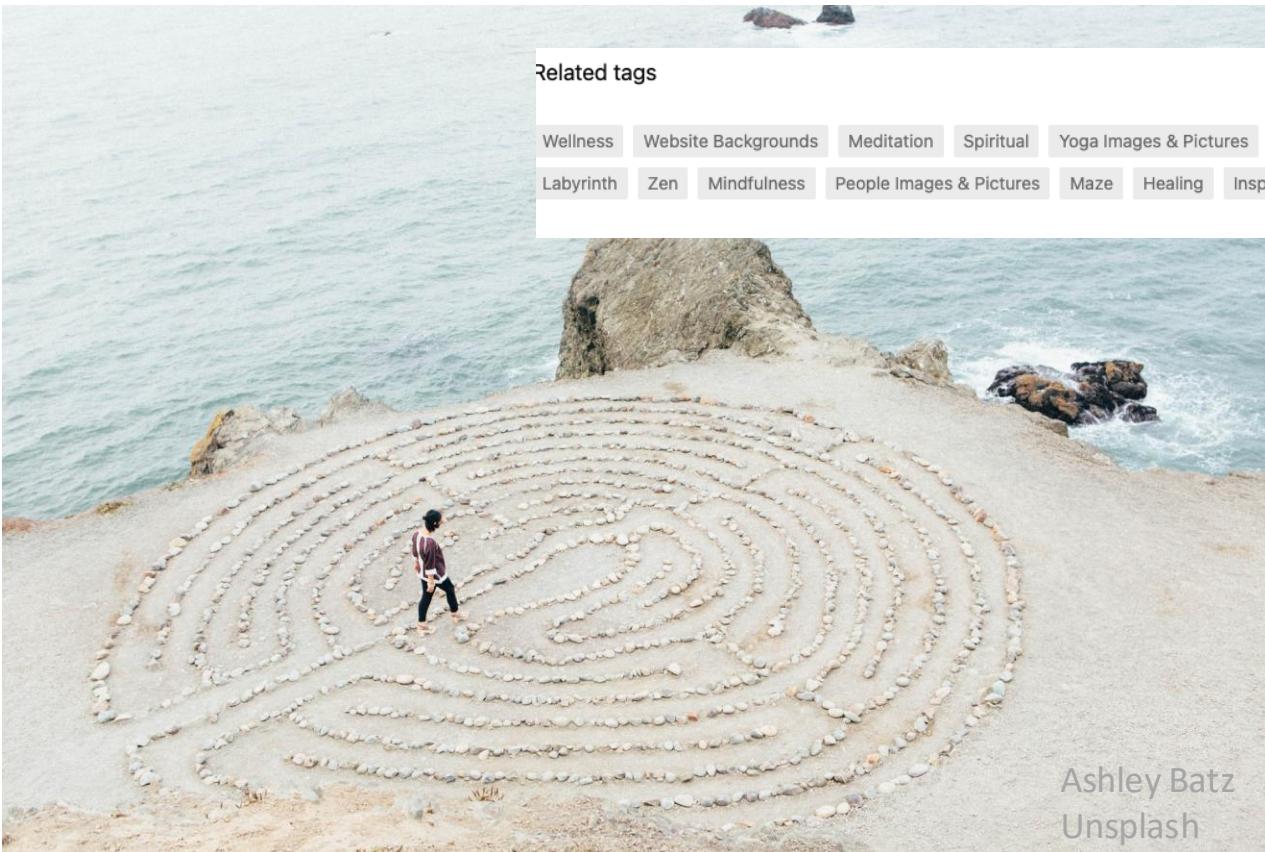
Trending: flower, wallpapers, backgrounds, happy, love

Photo of the Day by Hello I'm Nik

Read more about the Unsplash License

All yo

# Tags for an Unsplash Image



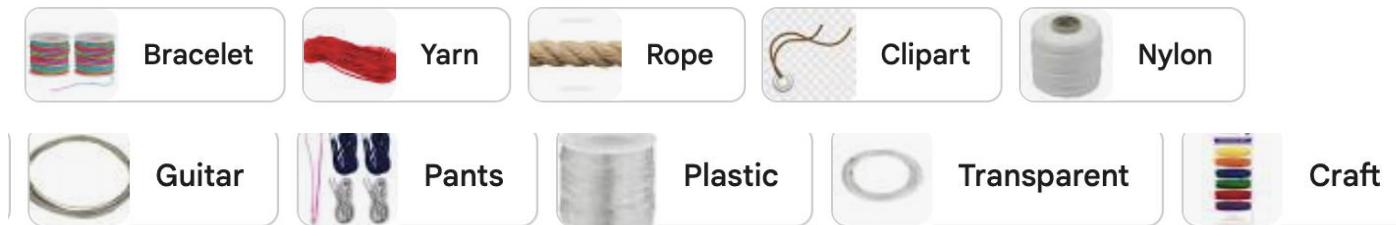
Related tags

Wellness Website Backgrounds Meditation Spiritual Yoga Images & Pictures Balance Nature Images Blog Ritual Coaching Spirit  
Labyrinth Zen Mindfulness People Images & Pictures Maze Healing Inspiration Health Images Sand Backgrounds

Ashley Batz  
Unsplash

# Google image groupings vs facets

Example: query on “string”



If this were organized, we'd have categories like:

- material (nylon, plastic)
- appearance (transparent, colors)
- uses (jewelry > bracelets, instruments > guitar, etc)

# Faceted Categories vs Automated Tags (google images, etc)

## Faceted categories

- Multiple facets per item
- Hierarchical (often)
- Controlled vocabulary
- Collection owner
- Category labels intentionally selected from an organized topic or field

## Tags

- Multiple tags per item
- Flat categories (usually)
- Any term is ok
- Algorithmically and data driven
- Categories data-driven, and uneven in content coverage.

# ASSIGNMENT: CREATE A FACETED IMAGE SEARCH AND NAVIGATION INTERFACE

1. Select 25 images for a collection
2. Create the faceted hierarchical category structure
3. Assign the categories to the images
4. Put everything into the Algolia system
5. Write up your assignment

# FACETED IMAGE SEARCH INTERFACE

Renad Abdallah and Shreshta Bhat Alevooru

VACATION TYPE

- Adventure
- Educational
- Art
- Museum
- Performance

Getaways

DURATION

- < 1 Week
- 1 Week
- > 1 Week



Italian Renaissance Tour	London Museums Tour	Scandinavian Art Exhibit	Museum Tour of the American West
Matteo Maretto	Grant Ritchie	Dev Benjamin	Michau0142 Parzuchowski
2019	2018	2017	2017
5K	500	500	500

Show more

VACATION TYPE

- Adventure
- Aquatic
- Sailing
- Scuba Diving

Educational

Getaways

DURATION

- < 1 Week
- 1 Week



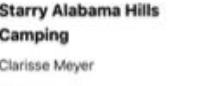
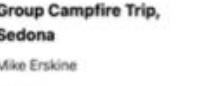
Under the sea in Cabo Pulmo	Dive Buddies Tour, Gulf of Mexico	Scuba diving in Redang Island
Pascal van de Vendel	Laya Clode	Nazarizal Mohammad
2018	2019	2020
5K	2.5K	2.5K

Show more

Example from previous class

VACATION TYPE					
Adventure	6				
Educational	2				
Getaways	1				

PRICE					
\$500	\$5000	2017 15K	2021 4.5K	2017 1.2K	2016 1.5K

LOCATION					
City	11				
Mountains	9				
Ocean	5				
Beach	4				
Country-Side	4				
Forest	4				
Desert	1				
Mountain	3	Bali Mountain Experience Aldino Hartan Putra 2018 3K	Peruvian Mountain Trekking Michau0142 Parzuchowski 2016	Exploring Cuisine of The West Indies Juan Jose 2018	Machu Picchu, Peru Tour Aaron Thomas 2019 2K

Example from previous class

## HIERARCHICAL FACETS

### FACET 1: TYPE OF TRIPS

#### Category 1: Getaways

- a. Health & Wellness
  - i. Spiritual
  - ii. Spa
- b. Retreats
  - i. Eco Tours
  - ii. Lodging & Treehouses

#### Category 2: Educational

- c. Culture
  - i. Archeological Site-Seeing
  - ii. Culinary Tours
- d. Art
  - i. Performance
  - ii. Museums

#### Category 3: Adventure

- e. Mountain
  - i. Trekking
  - ii. Skiing & Snowboarding
- f. Aquatic
  - i. Sailing
  - ii. Scuba Diving

### CET 2: LOCATION

#### Category 1: International:

- 1. Continent
  - a. Country

#### Category 2: Domestic:

- 1. State
  - a. City

### 3: TYPE OF BOOKINGS

#### Category 1: Solo Bookings

- a. Meet Other Travelers
  - i. Women Only
  - ii. Co-Ed
- b. Just you
  - i. Escorted
  - ii. Self-Guided

#### Category 2: Group Bookings

- c. Travel with Loved Ones
  - i. Families
  - ii. Couples
- d. Field Trip Options
  - i. Students
  - ii. Team Bonding

## FLAT FACETS

### Environment:

- City
- Beach
- Forest
- Mountains
- Country-side
- Desert
- Savanna

### Duration

- < 1 Week
- 1 week
- > 1 Week

### Cost:

Range from 500-5,000 (USD)

Example from previous class

## ASSIGNMENT DETAILS

- Assigned one partner to work with
- Need to create a JSON file
  - *describes both the categories and the images*
  - *Be sure to use a code editor to create the JSON*

# Algolia JSON Format

```
{  
    "name": "bold t-shirt",  
    "desc": "Be bold, wear a t-shirt with only one color",  
    "color": "white",  
    "sleeves": "short"  
,  
{  
    "name": "Hippie Vest with Fringes",  
    "desc": "Be hip: get back to where you belong, peace + love",  
    "color": ["red", "orange", "yellow", "green", "blue"],  
    "sleeves": "long"  
}
```

These are flat facets

Notice that you can have a list of attributes for a given item (colors in this example)

# Algolia JSON Format

```
{  
  "name": "Ursula K. Le Guin",  
  "categories": [  
    "level0": "Books",  
    "level1": ["Books > Science Fiction", "Books > Literature & Fiction"],  
    "level2": ["Books > Science Fiction > Time Travel", "Books > Literature & Fiction > Literary"]  
  ]  
}
```

This is a hierarchical facet

Notice that you can have a list of facet attributes for a given item (two kinds of level1 and two kinds of level2 in this example)

Thus, you can assign Le Guin to be an author of both Scifi and Lit,  
And within scifi, to Time travel and within lit to Literary.

# JSON format example

## including hierarchical facet with multiple assignments



```
,  
{  
    "title": "Hawaiian Sunset sailimg",  
    "artist": "Joel Moysuh",  
    "year": 2021,  
    "description": "Sunset sail cruise off the coast of Kapolei, Hawaii.",  
    "image": "https://images.unsplash.com/photo-1620261952751-b294980bcd8c?  
q=80&w=387&auto=format&fit=crop&ixlib=rb-  
4.1.0&ixid=M3wxMjA3fDB8MHxwaG90by1wYWdlfHx8fGVufDB8fHx8fA%3D%3D",  
    "activity_type": {  
        "at_level0": "Water Sport",  
        "at_level1": "Water Sport > Boating",  
        "at_level2": "Water Sport > Boating > Sailing"  
    },  
    "location": {  
        "loc_level0": "Country",  
        "loc_level1": "Country > U.S.A.",  
        "loc_level2": "Country > U.S.A. > Hawaii",  
        "loc_level3": "Country > U.S.A. > Hawaii > Waikiki"  
    },  
    "natural_wonder": {  
        "nw_level0": ["Mountains", "Oceans"],  
        "nw_level1": ["Mountains > Ko olau Range", "Oceans > Pacific"],  
        "nw_level2": ["Mountains > Ko olau| Range > Diamond Head (Le ahi)", "Oceans > Pacific >  
        Mamala Bay"]  
    }  
}
```