

# Introduction to Ontologies

Rafael Gonçalves

Protégé Short Course  
9 – 11 October, 2017

# Searching for directions

Google  Microphone Search

All Maps News Videos Images More Settings Tools

About 34,800,000 results (0.70 seconds)

My location

IBM Corporation, 1001 E Hillsdale Blvd, Foster City, CA 94404

Closed now

Map data ©2017 Google

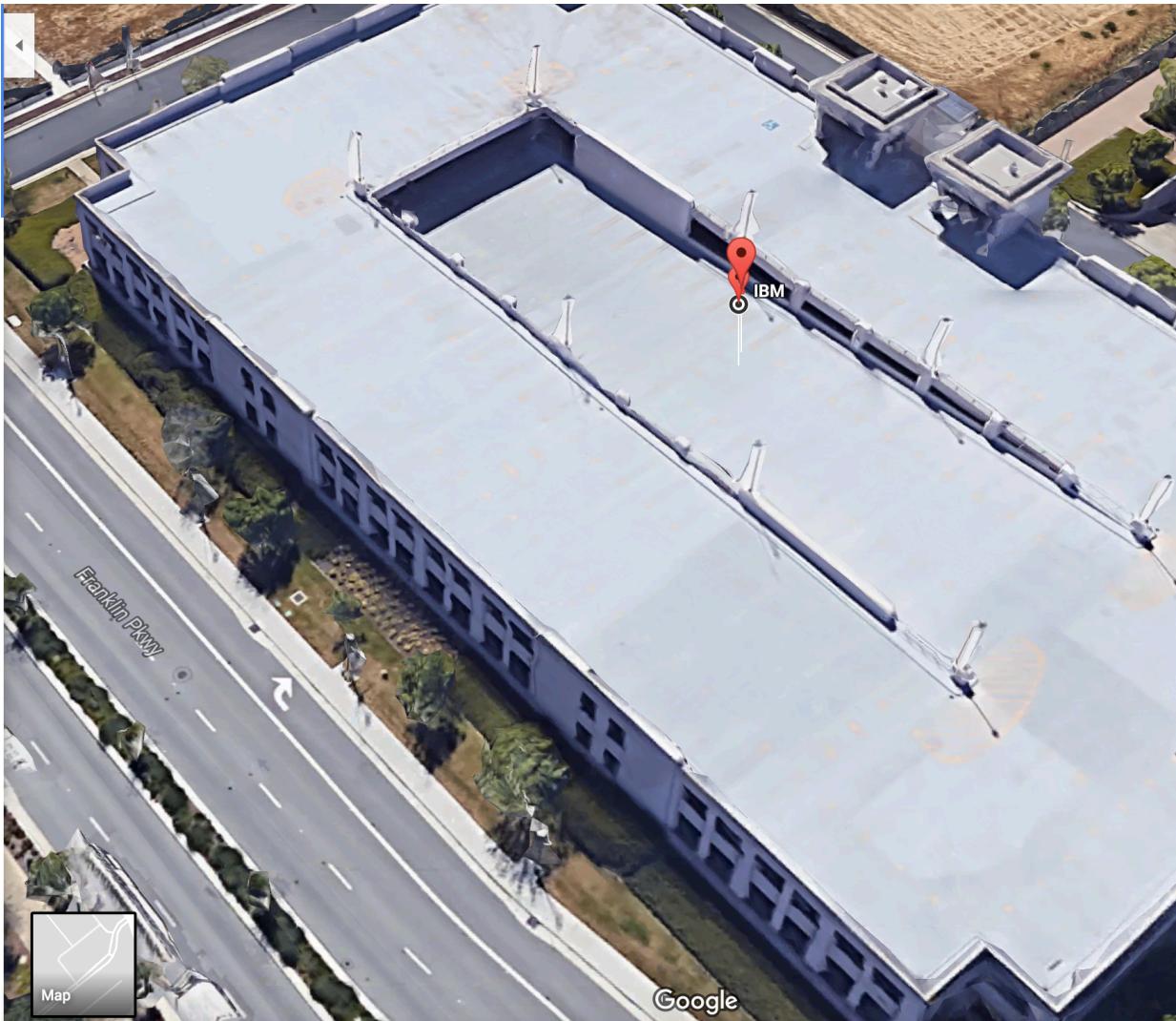
11 min (4.5 mi) via CA-92 E

12 min (5.1 mi) via S El Camino Real and CA-92 E

13 min (6.1 mi) via US-101 N

Directions

# Searching for directions



# Searching for directions

GOOGLE

how to get to ibm research center

All Maps Videos News Images More Settings Tools

About 3,010,000 results (1.18 seconds)

A map of the San Francisco Bay Area showing roads and landmarks. Three locations are marked: A (IBM Research Center in San Mateo), B (IBM Corporation in Foster City), and C (IBM Research Center in San Jose). The map also shows cities like San Mateo, Palo Alto, and San Jose, along with various highways like 1, 82, 92, 280, 84, 101, 680, 880, and 130.

**A IBM**  
0.8 mi · 1 Franklin Pkwy # 910, San Mateo, CA 94403

**B IBM Corporation**  
1.9 mi · 1001 E Hillsdale Blvd, Foster City, CA 94404  
Closed now

**C IBM Research**  
34.4 mi · 650 Harry Rd, San Jose, CA 95120  
Closed now

**IBM Thomas J. Watson Research Center**

[www.research.ibm.com/labs/watson/index.shtml](http://www.research.ibm.com/labs/watson/index.shtml) ▾

Learn more about the IBM Thomas J. Watson Research Center. ... As the newest IBM Research lab, researchers at Albany Nanotech site have been developing ...

Directions

Directions

Directions

4

# Searching for directions

Google how to get to ibm watson research center

All Maps News Videos Images More Settings Tools

About 5,270,000 results (0.86 seconds)

My location

IBM Thomas J. Watson Research Center, 1101 Kitchawan Rd, Yorktown Height

Not likely!

Map data ©2017 Google, INEGI

! No route found

# Disambiguating search keywords

jaguar

Web Images News Videos Shopping More ▾ Search tools

About 201,000,000 results (0.40 seconds)

[Jaguar® Bay Area Site - bayareajaguar.com](#)  
Ad [www.bayareajaguar.com/](http://www.bayareajaguar.com/) ▾  
Take A Jaguar Out For A Spin When You Locate A Bay Area Jaguar Dealer  
📍 4040 Stevens Creek Blvd, San Jose, CA - (408) 246-7600  
Locate A Dealer      Trade-In Value  
\$1,200 Toward 1st Mo. Pmt      View Offers

[Jaguar: Luxury Cars & Sports Cars | Jaguar USA](#)  
[www.jaguarusa.com/](http://www.jaguarusa.com/) ▾ Jaguar Cars ▾  
The official home of Jaguar USA. Our luxury cars feature innovative designs along with legendary performance to deliver one of the top sports cars in the ...

Models  
XF - XJ - XE - XK - R Performance -

F-Type  
The all-new Jaguar F-TYPE is

Jaguar Cars  
Car manufacturer  
Jaguar Cars is a brand of Jaguar Land Rover, a British multinational car manufacturer headquartered in Whitley, Coventry, England, owned by the Indian company Tata Motors since 2008. [Wikipedia](#)

Founded: September 4, 1922, Blackpool, United Kingdom  
CEO: Ralf Speth  
Founders: William Walmsley, William Lyons

See results about  
[Jaguar \(Animal\)](#)  
Scientific name: *Panthera onca*  
Lifespan: 12 – 15 y (In Wild)



# Investigating medicine

Google

## Ibuprofen

Prescription drug, Over-the-counter drug

Consult a doctor if you have a medical concern.

Treats pain and fever. This medicine is an NSAID.

By mouth: Side effects [\(1\)](#) - Warnings [\(1\)](#) - How to use [\(1\)](#)

Injection: Side effects [\(1\)](#) - Warnings [\(1\)](#) - How to use [\(1\)](#)

National Library of Medicine

**Brand names:** Advil, Midol, NeoProfen, Caldolor, Motrin, Ibu

**Pregnancy risk:** Category C (Risk cannot be ruled out)

**Possible side effects:** Headache, Nausea, Bleeding, Vomiting, Dizziness, Flatulence

**Drug classes:** Nonsteroidal anti-inflammatory drug, Analgesic

**Other drugs in same class:** Acetaminophen, Naproxen, Aspirin, More

**May treat:** Fever, Inflammation, Sinus infection, Gout, More

## Acetaminophen

Prescription drug, Over-the-counter drug

Consult a doctor if you have a medical concern.

Treats minor aches and pain and reduces fever.

By mouth: Side effects [\(1\)](#) - Warnings [\(1\)](#) - How to use [\(1\)](#)

Injection: Side effects [\(1\)](#) - Warnings [\(1\)](#) - How to use [\(1\)](#)

Rectal: Side effects [\(1\)](#) - Warnings [\(1\)](#) - How to use [\(1\)](#)

National Library of Medicine

**Brand names:** Tylenol, Panadol, Mapap, Tempra, Ofirmev, Feverall, More

**Formula:** C8H9NO2

**Pregnancy risk:** Category C (Risk cannot be ruled out)

**Drug class:** Analgesic

**Other drugs in same class:** Ibuprofen, Aspirin, Codeine, Tramadol, More

**May treat:** Pain, Fever, Sinus infection, Lyme disease, Sore throat, More

Acetaminophen > Other drugs in same class

Ibuprofen

Aspirin

Codeine

Tramadol

Hydrocodone/Acetaminophen

Oxycodone

Ketoprofen

Gabapentin

Morphine

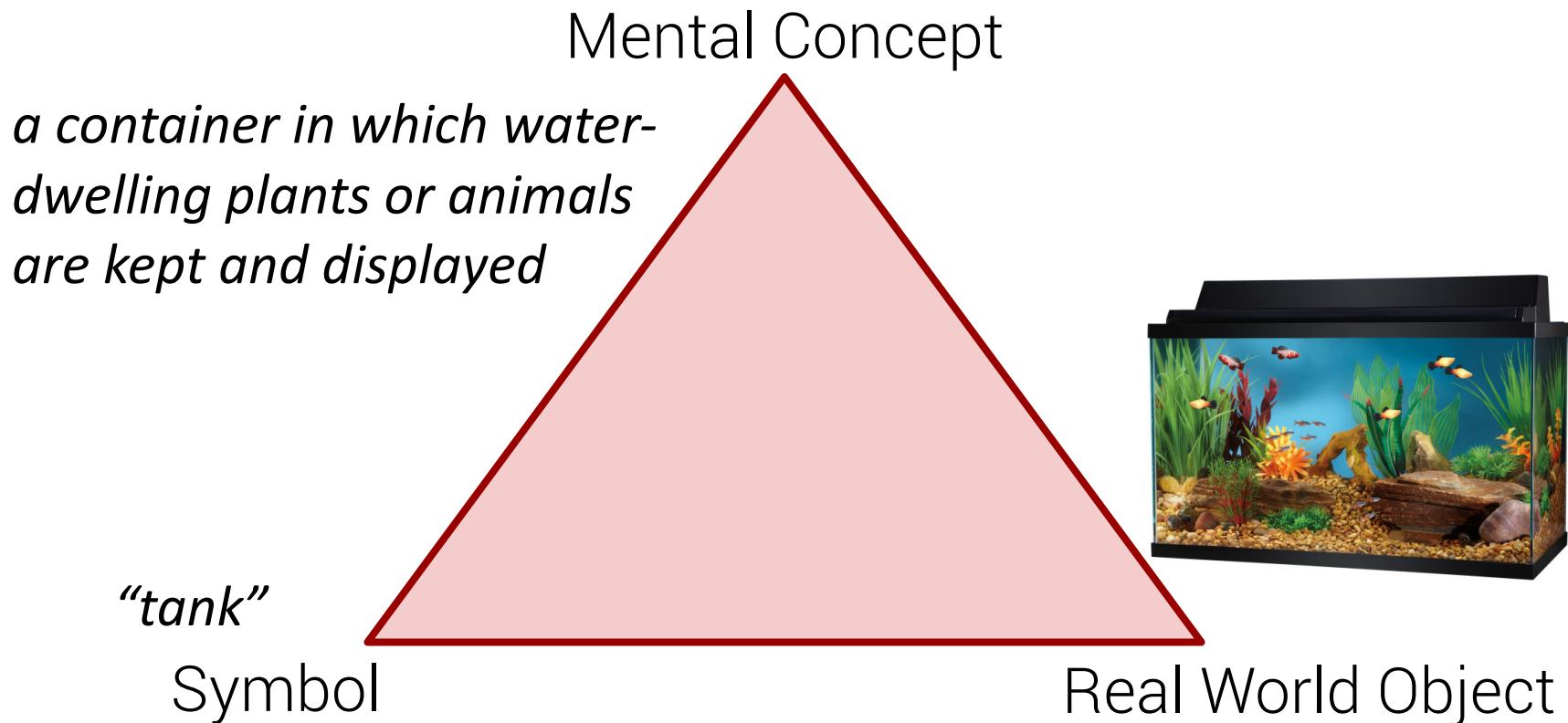
Propoxyphene

Hydroxyzine

Hydromorphone

What do these use cases have in common?

# The triangle of meaning



According to the Oxford Dictionary, the 500 most used words in the English language have at least 14,070 different definitions. This is an average of 28 meanings per word.

Source: <http://inmyownterms.com/mysmartterms/mysmartterms-5-the-semantic-triangle-words-dont-mean-people-mean/>

Merriam-Webster added more than 250 new terms and definitions to its online dictionary this month.

Source: Time, Oct 2, 2017, p16.

# What is an ontology?

**Origin of the word:** A branch of philosophy concerned with "*that which exists*"; that is, a description of the things in the world

**Computer scientists:** *a shared understanding or specification of the concepts of interest in a domain of information that can be used by both computer and humans to describe and process that information.* The goal with a computer science ontology is to make knowledge of a domain computationally usable

# Ontology: Definitions

Most famous one (Tom Gruber, 1992):

*"An ontology is a specification of a conceptualization"*

An updated version (Tom Gruber, 2009):

*"An ontology defines a set of representational primitives with which to model a domain of knowledge or discourse"*

# An ontology...

Contains explicit definitions of terms

Provides unambiguous interpretation of the meaning of terms

Intended to be shared among people and software systems

Developed for use with automated reasoning systems used to infer new facts from existing definitions

# Ontology about “acetaminophen”

Acetaminophen also goes by the name “paracetamol”

It is a mild analgesic drug (so it's a kind of drug)

Has several trade/brand names: Panadol, Tylenol, Feverall, ...

It is used in the treatment of: Pain, fever, sinus infections, sore throats, ...

Has methods of consumption: Oral, intravenous, rectal

It has side effects (depending on method of consumption):  
Dark urine, nausea, loss of appetite, ...

# Why develop an ontology?

Share common understanding of the structure of information among people, and among software systems

To make domain assumptions explicit:  
easier to change domain assumptions  
easier to understand and update legacy data

# How are ontologies used?

As shared, "standard" vocabularies for a domain (e.g., NCI Thesaurus, SNOMED CT)

As a way to integrate heterogeneous data sources

For annotating data with standardized terms (e.g., electronic health records, experiment datasets, ...)

Faceted browsing

Decision support

# SNOMED and NHS

"Better patient safety through electronic communication between clinicians has been secured by the adoption of a common clinical language across all healthcare settings and organisations. Nurses, doctors, physiotherapists are now using SNOMED Clinical Terms (CT), the most comprehensive, multilingual clinical healthcare terminology in the world, meaning that information is exchanged accurately and safely across England."

Source: <https://www.gov.uk/government/news/common-clinical-language-for-nhs-will-help-improve-patient-care-and-safety>

# SNOMED and Kaiser Permanente

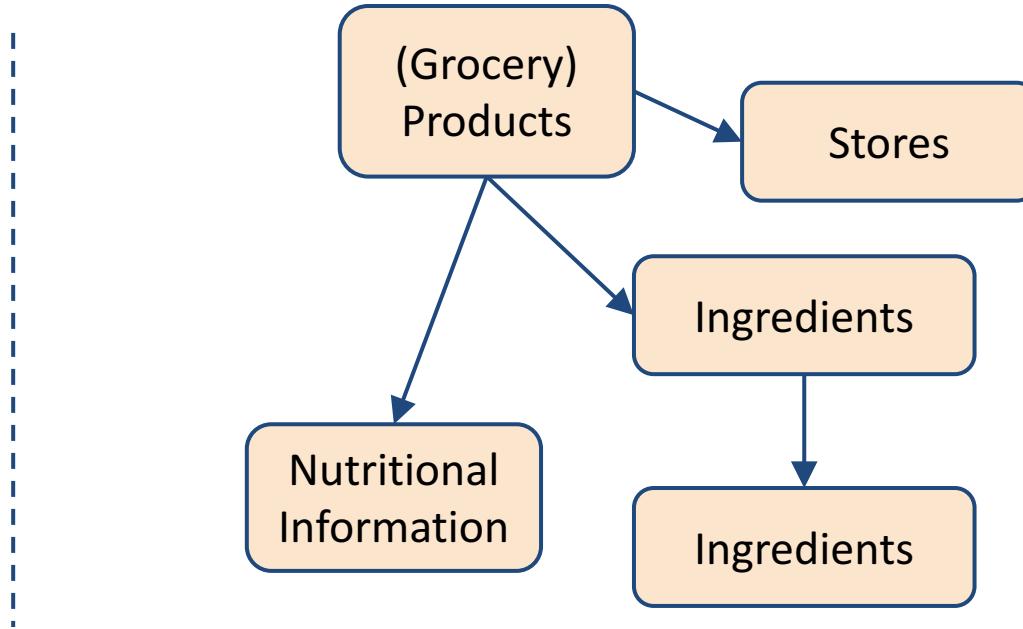
"SNOMED CT is known to be used for Clinical Decision Support in a number of organizations including Kaiser Permanente, Duke University Medical Hospital in the US..."

"Kaiser Permanente has a central terminology service which develops and maintains clinical decision support rules and prompts. These can be quickly distributed for implementation throughout the Organisation. In addition to using this type of approach for clinical decision support, Kaiser Permanente also use it to identify potential cohorts for clinical trials."

Source: SNOMED CT Guide:

[http://ihtsdo.org/fileadmin/user\\_upload/doc/download/doc\\_StarterGuide\\_Current-en-US\\_INT\\_20141202.pdf?ok](http://ihtsdo.org/fileadmin/user_upload/doc/download/doc_StarterGuide_Current-en-US_INT_20141202.pdf?ok)

# Course running example: An online shop



Design an intelligent Web shop that allows customers to search and filter the products according to criteria such as their ingredients

E.g., what cookies contain gluten?

# Google Shopping Express

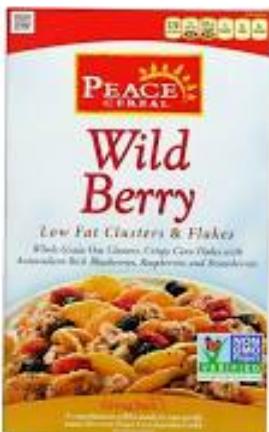
The screenshot shows the Google Shopping Express website interface. At the top, there's a navigation bar with a back arrow, a lock icon, the URL <https://www.google.com/shopping/express/>, a dropdown menu, and a search bar with the placeholder "Search for products". Below the navigation is a header with the Google logo, a "Google Express" button with a delivery truck icon, a "Stores" dropdown, and a "Grocery" link which is highlighted in red. Other links in the header include "Health & Beauty", "Home", "Baby", "Toys", "Apparel", and "More". A banner at the top left says "Get free delivery on eligible orders". The main content area has a blue header with the word "Grocery" and a "View all" link. To the right of the header is a small graphic of various grocery items like bread, milk, and juice. The main content is organized into several categories:

- GRAINS, RICE & CEREAL** (with arrows): Cereal, Oats & Grits, Rice, Quinoa.
- BEVERAGES** (with arrows): Coffee, Juice, Water, Soda.
- HOUSEHOLD SUPPLIES** (with arrows): Laundry, Cleaning, Paper Products, Trash Bags.
- CONDIMENTS** (with arrows): Ketchup, Hot Sauce, Pasta Sauce, Salad Dressing.
- SNACK FOODS** (with arrows): Crackers, Chips, Fruit Snacks, Cereal & Granola Bars.
- BAKING** (with arrows): Baking Mixes, Flour, Sugar, Cooking Oil.

On the right side, there's a "Featured Stores" section with three boxes: "WHOLE FOODS", "NOB HILL FOODS", and "Costco WHOLESALE". In the bottom right corner of the main content area, there's a large cardboard box with a "Google Express" logo and a colorful graphic of a pie chart and a red ribbon.

# Google Shopping Express

Q: Cereals with low salt



*Nutritional  
information*



Peace All Natural Clusters &  
Flakes Cereal, Low Fat, Wild  
Berry - 10 oz box

**\$3.99**

3 reviews

NOB HILL FOODS  
[More stores](#)

Serving Size	1 cup
Servings Per Container	5
Total Fat	3 g (5 %)
Saturated Fat	0 g (0 %)
Cholesterol	0 mg (0 %)
<b>Sodium</b>	<b>330 mg (14 %)</b>
Total Carbohydrate	44 g (15 %)
Dietary Fiber	3 g (12 %)
Sugars	7 g
Protein	5 g
Vitamin A	2 %
Vitamin C	2 %
Calcium	2 %
Iron	6 %

How does one go about developing ontologies?

# Domain and scope of ontologies

What is the domain that the ontology will cover?

For what are we going to use the ontology?

What questions should the ontology be able to answer?

Who will use and maintain the ontology?

# Defining the scope of an ontology

"How task-dependent are ontologies? Presumably, the kinds of things that actually exist do not depend on our goals. In that sense, ontologies are not task-dependent.

For example, in the domain of fruits, we would focus on particular aspects of reality if we were developing the ontology for the selection of pesticides; we would focus on other aspects of reality if we were developing an ontology to help chefs select fruits for cooking"

Source: Chandrasekaran, 1999

<http://www.csee.umbc.edu/courses/771/papers/chandrasekaranetal99.pdf>

# Hands-on: Competency questions for the shopping domain

Imagine you are the customer of a grocery Web shop

Think about the different types/categories of groceries

What kind of questions would you like to ask this smart Web shop?

Formulate queries and filter grocery products based on some criteria

Come up with competency questions for the grocery domain and write each one on a post-it note

# Hands-on: Competency questions for the shopping domain

Discuss competency questions

Organize the questions



# Domain and scope: Competency questions

Competency questions are questions that the ontology should be able to answer

Helpful for:

defining the scope of the ontology

defining the vocabulary (main terms) of the ontology

testing the ontology

Examples:

*What are the cereals that are low in salt/sodium?*

*How many calories are there per serving of Cheerios?*

*What children's multi-vitamins don't contain glucose syrup?*

# Some competency questions

What beverages are sugar-free?

What are the ingredients of a product?

What baked goods (such as cakes) do not contain canola oil?

Does this product contain soy?

Which supplements are vegetarian?

What types of gluten free breads are there?

Who is the producer of a certain product?

Is this product high in salt?

How many calories does a serving of this product contain?

What kind of gluten-free biscuits do you have?

Does this product contain nuts (or traces of nuts)?

Where can I buy this product from?

Is this product sold from WholeFoods in Palo Alto?

# What makes good competency questions?

Ones that set out precise requirements for the ontology

Ensure that the ontology covers (some aspect of) the domain

Test the usability of the ontology for querying data  
how hard is it to formulate a query to answer a question?

# Reusing ontologies

Save time and effort

Avoid defining terms that are well-modeled elsewhere

Use ontologies that have been validated through use in applications

Increase interoperability

# Ontology repositories

## Biomedical ontologies

NCBO Bioportal (<http://bioportal.bioontology.org>)

Ontology Lookup Service (OLS) (<http://www.ebi.ac.uk/ols>)

## Ontologies in various representation languages

Ontohub (<https://ontohub.org>)

OntoBee (<http://www.ontobee.org>)

Linked Open Vocabularies (<http://lov.okfn.org/dataset/lov>)

# Types of reuse

Reuse the entire ontology (via *imports*)

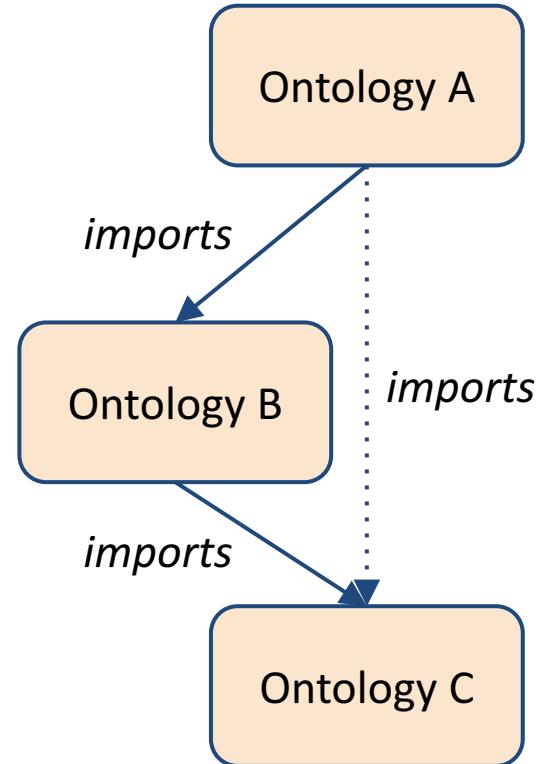
Reuse of specific parts of an ontology:  
individual terms  
a module

# Reuse an entire ontology

OWL provides a mechanism for importing an ontology into another one, through the *owl:imports* statement

The import is by URI (the location of the ontology to import)

The imports statement is transitive



# Reusing terms of an ontology

Reuse single terms via their IRI (identifier)

E.g., the term *Gene* in *NCI Thesaurus* is identified by:

<http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C16612>

Use ontology repositories to search for terms

MIREOT – Minimum Information to Reference an External Ontology Term

Tools to search for terms

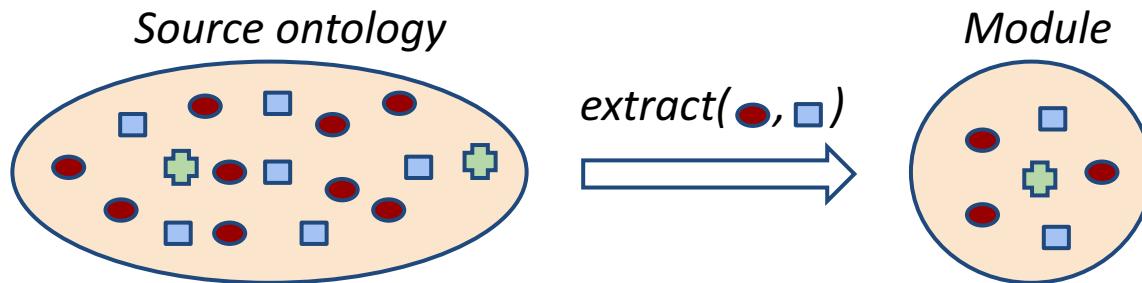
E.g., the ProtegeLOV plugin:

<http://boris.villazon.terrazas.name/projects/prolov/index.html>

# Reusing subsets of an ontology

Module extractors allow users to extract modules that represent the knowledge one wants to reuse

Specify a set of terms to extract a module for – the output is a subset that represents all knowledge about these terms from the original ontology



# Identifying terms in the domain

What are the terms we need to talk about?

What are the properties of these terms?

What do we want to say about the terms?

# Identifying terms in the domain

Remember the competency questions:

What are the ingredients of a grocery item?

What baked goods (such as cookies) do not contain canola oil?

Does this product contain soy?

Is this product sold from WholeFoods in Palo Alto?

We use the terms from the competency questions as the domain vocabulary of our ontology

# Terms we talk about

Products, groceries, ingredients, stores, places  
contains, has ingredient, sells  
soy, canola oil, cake  
Whole Foods, Palo Alto

...

# Hands-one exercise: Classifying ingredients

## Work in groups of ~5

You will get a stack of cards with ingredient names on them  
E.g., baking soda, milk, canola oil, corn starch, cream of tartar, ...

Group these cards into categories that make sense to your group

Once you identify the groups, try to give a name to that group, and write down the criteria used for creating the group

You may use Wikipedia or other online resources to figure out some of the ingredients

# Testing your ontology

Check if the competency questions can be answered by your ontology (good check for coverage)

Create some unit tests (such as predefined queries) that should always return the expected values

# Approaches to modeling

## Build a **generic ontology**

conceptual model

more reusable

a more accurate reflection of the domain

may not be always practical

may be harder to use in applications

## Build an **application ontology**

conceptual model highly adapted to a use case

less reusable

easy to use in your specific application

not necessarily an accurate reflection of the domain

# Ontology design patterns

Some modeling patterns (e.g., part-of) are very common and needed in a lot of ontologies

A design pattern is a small set of reusable modeling constructs that solve a particular modeling issue independently of the domain

Online repositories with the most common ontology design patterns:

<http://ontologydesignpatterns.org>

<http://www.gong.manchester.ac.uk/odp/html/>

# Further resources

Ontology 101: A guide to creating your first ontology (Natasha Noy):  
<http://protegewiki.stanford.edu/wiki/Ontology101>

BD2K Ontology Guide (Nicole Vasilevsky)  
<https://github.com/OHSUBD2K/BDK14-Ontologies-101>

Semantic Web Best Practices and Deployment Working Group's Semantic Web Tutorials page: <http://www.w3.org/2001/sw/BestPractices/Tutorials>

Protégé OWL Tutorial: A step-by-step guide to modeling in OWL using the popular Protégé OWL tools (Matthew Horridge, *et al*)  
<http://owl.cs.manchester.ac.uk/publications/talks-and-tutorials/protg-owl-tutorial/>

OBO Tutorial (James Overton)  
<https://github.com/jamesaoverton/obo-tutorial>

OBO Ontology Starter Kit (Chris Mungall)  
<https://github.com/INCATools/ontology-starter-kit>

# Further resources

A Description Logic Primer: <http://arxiv.org/abs/1201.4089>

Brief introduction to ontologies, by Robert Stevens:  
<http://www.cs.man.ac.uk/~stevensr/menupages/background.php>

User documentation on the Protégé wiki:  
<http://protegewiki.stanford.edu/wiki/Protege4UserDocs>

ACE Plugin, a natural language interface to OWL:  
<http://attempto.ifi.uzh.ch/aceview/>

# Module extraction resources

Information about modularity:

<http://owl.cs.manchester.ac.uk/research/modularity/>

A Web-based module extractor:

<http://owl.cs.manchester.ac.uk/modularity/>

OntoFox: <http://ontofox.hegroup.org/>

ROBOT, a command-line tool: <https://github.com/ontodev/robot>