Controlled vocabularies

CMT209
Informatics

Cardiff School of **Computer Science & Informatics**

http://www.cs.cf.ac.uk

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Lecture

- effective communication
- language & semantics
- controlled vocabulary
  - basic elements: term, definition, related/ broader/narrower term, use, used for
  - relationships: equivalence, hierarchical, associative
  - different types: flat/hierarchal list, synonym ring, authority file, classification scheme, facet vocabulary, thesaurus, ontology
  - uses
- case study: Amazon

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Controlled vocabulary (CV)

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Controlled vocabulary

- What is a controlled vocabulary?
- the most effective communication occurs when all parties involved agree on the meaning of the terms being used
- everyone is using the same term to mean the same thing
- Any misunderstanding stories?

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Controlled vocabulary

- What is a controlled vocabulary?
- the most effective communication occurs when all parties involved agree on the meaning of the terms being used
- everyone is using the same term to mean the same thing
- Any misunderstanding stories?

In one anecdote, a young man, in the course of a passionate courtship, tells his American girlfriend, "I'll give you a ring tomorrow." All he meant was that he would call her by telephone. But she understood him to have offered betrothal, and the relationship didn't survive the misunderstanding.

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Controlled vocabulary

- What is a controlled vocabulary?
- the most effective communication occurs when all parties involved agree on the meaning of the terms being used
- everyone is using the same term to mean the same thing
- Any misunderstanding stories?

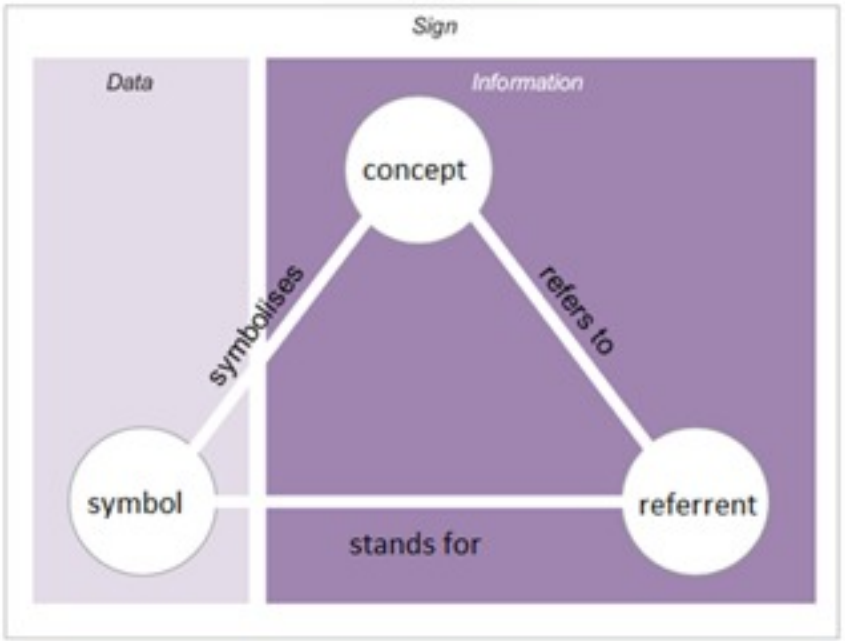
In one anecdote, a young man, in the course of a passionate courtship, tells his American girlfriend, "I'll give you a ring tomorrow." All he meant was that he would call her by telephone. But she understood him to have offered betrothal, and the relationship didn't survive the misunderstanding.

In my first week on a U.S. university campus, I asked an American where I could post a letter to my parents. "There's a bulletin board at the Student Center," he replied, "but are you sure you want to post something so personal?" I soon learned that I needed to "mail" letters, not "post" them (even though in the United States you mail them at the "post office").

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Communication

- when we converse, we speak in natural language, e.g. English
- words are used to convey meaning
- there is so much richness, variance & confusion in language
- problems with synonymy (variation) & homonymy (ambiguity)
- What's in a name?!?



rose

Romeo & Juliet by William Shakespeare

'Tis but thy name that is my enemy;
Thou art thyself, though not a Montague.
What's Montague? it is nor hand, nor foot,
Nor arm, nor face, nor any other part
Belonging to a man. O, be some other name!
What's in a name? that which we call a rose
By any other name would smell as sweet;
So Romeo would, were he not Romeo call'd,
Retain that dear perfection which he owes
Without that title. Romeo, doff thy name,
And for that name which is no part of thee
Take all myself.

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

What's in a name?

- words facilitate communication when we have mutual agreement on their meaning
- there is so much richness, variance & confusion in terminology
- we often need to impose some restrictions to facilitate agreement between the concepts and the vocabulary used to describe them
- technical communication improves by ensuring that everyone is using the same term to mean the same thing

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Controlled vocabulary

http://www.youtube.com/watch?v=sm8jlsI-vgA

Controlled vocabulary (CV)

- controlled vocabulary (CV) = an organised list of terms (i.e. words & phrases) that are acceptable values for completing certain metadata fields
- we have many different ways of describing concepts
- drawing all of them together under a single term eliminates guesswork
- precise & efficient: most of the relevant information is grouped together in one place, saving time having to search with other synonyms for the given term
- the terms are chosen & organised by trained professionals who possess expertise in the subject area

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Basic elements

- a well-developed vocabulary is usually called a thesaurus
- terms which constitute the thesaurus are listed in alphabetical order to facilitate finding them
- usually hierarchical in structure

Basic Elements of a Controlled Vocabulary

- **TERM:** An accepted element in the controlled vocabulary
- **[DEFINITION/QUALIFIER]:** An optional descriptive paragraph that serves to either distinguish terms that are homographs or precisely define terms that may have multiple meanings.
- **RELATED TERM(S):** An element or elements also in the controlled vocabulary that an indexer may prefer to the one listed.
- **BROADER TERMS**: Elements in the controlled vocabulary that are higher in the hierarchy than the element listed.
- **NARROWER TERMS**: Elements in the controlled vocabulary that are lower in the hierarchy than the element listed.
- **USE**: This follows an uncontrolled listing, and points the indexer to an accepted term in the controlled vocabulary
- **USED FOR**: This is the reciprocal of the USE reference and accompanies the term to which the USE reference refers.

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Controlled vocabulary (CV)

- can be organised with the following relationships:
  - equivalence
  - hierarchical
  - associative
- different CV types:
  1. flat term list
  2. hierarchal term list
  3. synonym ring
  4. authority file
  5. classification scheme
  6. facet controlled vocabulary
  7. thesaurus
  8. ontology

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

CV: flat term list

- a list of terms with no relationships implied
- good CV includes term definitions
- terms are easily differentiated (no overlap)
- terms are on the same scale, i.e. no term encompasses any other term

- **GOLDFISH** – a small usually golden yellow or orange cyprinid fish (Carassius auratus) often kept as an aquarium and pond fish
- **CATFISH** – any of an order (Siluriformes) of chiefly freshwater stout-bodied scaleless bony fishes having long tactile barbels
- **COW** – a domestic bovine animal regardless of sex or age
- **HORSE** – a large solid-hoofed herbivorous mammal (Equus caballus, family Equidae, the horse family)

CV: hierarchical term list

- a list of terms that are grouped to imply a certain order or method of organisation
- parent & child relationships between terms, usually kind-of or is-a
- all child terms across a group are generally on the same scale

- **FISH** – aquatic animals
  - **GOLDFISH**
  - **CATFISH**
- **MAMMALS** – any of a class (Mammalia) of warm-blooded higher vertebrates (as placentals, marsupials, or monotremes) that nourish their young with milk
  - **COW**
  - **HORSE**

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

CV: synonym ring

- extends term lists by providing additional equivalent terms
- challenge: deciding what constitutes a synonym
- synonyms can be:
  - words with the same or very similar meaning
  - acronyms, e.g. BBC & British Broadcasting Company
  - variant spellings, e.g. cancelled & canceled
  - scientific terms vs. popular use terms: acetylsalicylic acid & aspirin
- search queries can then be expanded to include all synonyms, e.g. BBC → BBC or "British Broadcasting Company"

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Query expansion

code.google.com/apis/searchappliance/documentation/50/help_gsa/serve_query_expansion.html

Serving > **Query Expansion**

Query expansion enables the search appliance to automatically add extra terms to a user's search query, in order to return additional relevant results. When query expansion is enabled, the appliance can expand two types of terms:

- Words that share the same word stem as the word given by the user. For example, if the user search query includes "engineer," the search appliance could add "engineers" to the query. Query expansion behavior is context sensitive. The search term "engineer" alone might not be expanded, but "software engineer" is expanded to include "engineers."
- Terms of one or more space-separated words that are synonymous or closely related to the words given by the user. For example, if a user searches for "FAQ," the appliance could add "frequently asked questions " to the query, or if a user enters "office building," the query could expand to include "office tower."

Query expansion is disabled when a query contains special query terms, such as `inurl:`, `allintitle:`, and so on.

This topic has the following sections:

About query expansion terms

Built-in word matching logic is provided, and you can specify your own list of word matches. Each front end has a policy that specifies whether it uses the search appliance built-in logic (the "standard" set of terms), your own list of synonyms (the "local" set), or both (the "full" set).

As you create a query expansion policy, you'll need to balance the positive effects of adding additional terms and producing additional results with the possibility of creating accidental expansions that are not useful. You'll need to monitor the quality of results to ensure that unwanted expansions do not occur.

Standard terms

CV: authority file

- synonym ring + one type of term relationship
- instead of all of the terms being equal, one term is identified as the preferred term & the others are considered variant terms

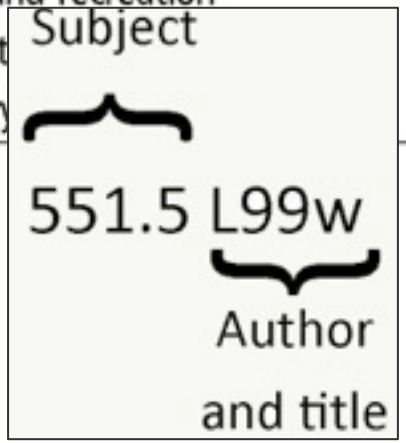
Preferred term	Other terms
meteorology	weather, atmospheric science
ultraviolet radiation	UV, ultra-violet radiation

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

CV: classification scheme

- codes (letters and/or numbers) that represent CV terms
- e.g. Dewey Decimal Classification
- a proprietary system of library classification developed by Melvil Dewey in 1876
- classifies books on library shelves in an efficient, specific & repeatable order
- makes it easy to find any book and return it to its proper place on the library shelves

Class	DDC
000	Computers, information and general reference
100	Philosophy and psychology
200	Religion
300	Social sciences
400	Language
500	Science
600	Technology
700	Arts and recreation
800	Literat
900	Histor



CV: Faceted controlled vocabulary

- facet – clearly defined, mutually exclusive & collectively exhaustive aspects, properties or characteristics of a class
- facets provide different avenues into describing a resource
- e.g. facets for restaurants might be type & cost
  - type (steakhouse, Italian, Mexican, etc.)
  - cost (high, low, medium)
- terms allowed in each facet may come from flat term lists, hierarchies, synonym rings, authority files or classification schemes

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

CV: Thesaurus

legal service centres	
SN	Centres serving the public which may provide a variety of legal services including information, advice, referral, or help finding a lawyer.
BT	legal service providers
NT	clinics
NT	legal information centres
RT	legal services
UF	justice access centres

- CV networked together by relationships between terms:

1. **equivalence**: synonyms with the ability to suggest which term is the preferred term
2. **hierarchy**: a broad class (CL) for a term; broad term (BT) & narrow term (NT)
3. **associative**: relationships across hierarchies – related terms (RT)
4. **scope notes** (SN): defines term or breadth of term and its usage

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

CV: Ontology

- describe concepts & their arbitrary relationships in programmatic ways
- somewhat like a thesaurus, but:
  - often no preferred terms
  - concepts & relationships are described in a machine readable format
- supports semantic interoperability (exchange of data with unambiguous, shared meaning)
- often described by standard XML-based languages accepted by the W3C:
  - RDF (Resource Description Framework)
  - OWL (Web Ontology Language)

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

CV: Ontology

- ontologies enable child terms to inherit all properties of their parents
- ontologies support multiple inheritance, so that compound concepts can be created
- more then just a terminology – knowledge representation!
- ontologies support formal reasoning:
  - class inferences
  - instance inferences

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Class inferences

- e.g. bus drivers are drivers
- OWL:

```
Class(a:bus_driver complete intersectionOf(a:person
  restriction(a:drives someValuesFrom (a:bus))))

Class(a:driver complete intersectionOf(a:person
  restriction(a:drives someValuesFrom (a:vehicle))))

Class(a:bus partial a:vehicle)
```

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Class inferences

- e.g. bus drivers are drivers
- OWL:

```
Class(a:bus_driver complete intersectionOf(a:person
  restriction(a:drives someValuesFrom (a:bus))))

Class(a:driver complete intersectionOf(a:person
  restriction(a:drives someValuesFrom (a:vehicle))))

Class(a:bus partial a:vehicle)
```

1. A bus driver is a person who drives a bus.

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Class inferences

- e.g. bus drivers are drivers
- OWL:

```
Class(a:bus_driver complete intersectionOf(a:person
  restriction(a:drives someValuesFrom (a:bus))))

Class(a:driver complete intersectionOf(a:person
  restriction(a:drives someValuesFrom (a:vehicle))))

Class(a:bus partial a:vehicle)
```

1. A bus driver is a person who drives a bus.
2. A driver is a person who drives a vehicle.

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Class inferences

- e.g. bus drivers are drivers
- OWL:

```
Class(a:bus_driver complete intersectionOf(a:person
  restriction(a:drives someValuesFrom (a:bus))))

Class(a:driver complete intersectionOf(a:person
  restriction(a:drives someValuesFrom (a:vehicle))))

Class(a:bus partial a:vehicle)
```

1. A bus driver is a person who drives a bus.
2. A driver is a person who drives a vehicle.
3. A bus is a vehicle.

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Class inferences

- e.g. bus drivers are drivers
- OWL:

```
Class(a:bus_driver complete intersectionOf(a:person
  restriction(a:drives someValuesFrom (a:bus))))

Class(a:driver complete intersectionOf(a:person
  restriction(a:drives someValuesFrom (a:vehicle))))

Class(a:bus partial a:vehicle)
```

1. A bus driver is a person who drives a bus.
2. A driver is a person who drives a vehicle.
3. A bus is a vehicle.

   -> a bus driver is a driver

Instance inferences

- e.g. Pete is a Person, Spike is an Animal
- OWL:

```
Individual(a:Spike type(owl:Thing) value(a:is_pet_of a:Pete))

Individual(a:Pete type(owl:Thing))

ObjectProperty(a:has_pet domain(a:person) range(a:animal))

ObjectProperty(a:is_pet_of inverseOf(a:has_pet))
```

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Instance inferences

- e.g. Pete is a Person, Spike is an Animal
- OWL:

```
Individual(a:Spike type(owl:Thing) value(a:is_pet_of a:Pete))

Individual(a:Pete type(owl:Thing))

ObjectProperty(a:has_pet domain(a:person) range(a:animal))

ObjectProperty(a:is_pet_of inverseOf(a:has_pet))
```

1. Spike is_the_pet_of Pete.

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Instance inferences

- e.g. Pete is a Person, Spike is an Animal
- OWL:

```
Individual(a:Spike type(owl:Thing) value(a:is_pet_of a:Pete))

Individual(a:Pete type(owl:Thing))

ObjectProperty(a:has_pet domain(a:person) range(a:animal))

ObjectProperty(a:is_pet_of inverseOf(a:has_pet))
```

1. Spike is_the_pet_of Pete.
2. So Pete has_pet Spike.

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Instance inferences

- e.g. Pete is a Person, Spike is an Animal
- OWL:

```
Individual(a:Spike type(owl:Thing) value(a:is_pet_of a:Pete))

Individual(a:Pete type(owl:Thing))

ObjectProperty(a:has_pet domain(a:person) range(a:animal))

ObjectProperty(a:is_pet_of inverseOf(a:has_pet))
```

1. Spike is_the_pet_of Pete.
2. So Pete has_pet Spike.
3. Pete must be a Person.
   Spike must be an Animal.

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Ontology inference tools

protege.stanford.edu

protégé

Search:

HOME | OVERVIEW | DOCUMENTATION | DOWNLOADS | SUPPORT | COMMUNITY | WIKI | ABOUT US

welcome to protégé

Protégé is a free, open source ontology editor and knowledge-base framework.

The Protégé platform supports two main ways of modeling ontologies via the Protégé-Frames and Protégé-OWL editors. Protégé ontologies can be exported into a variety of formats including RDF(S), OWL, and XML Schema. (more)

Protégé is based on Java, is extensible, and provides a plug-and-play environment that makes it a flexible base for rapid prototyping and application development. (more)

Protégé is supported by a strong community of developers and academic, government and corporate users, who are using Protégé for knowledge solutions in areas as diverse as biomedicine, intelligence gathering, and corporate modeling.

community	
Registered Users	176,656
protege-users list members	17,307
protege-discussion list members	2,646
protege-owl list members	2,295

news

12 Jan 2012
Protege 3.4.8 released!
(read more) | (download)

28-30 March 2012
Protégé-OWL Short Course
Stanford, California

Find us on Facebook

twitter FOLLOW ME

632

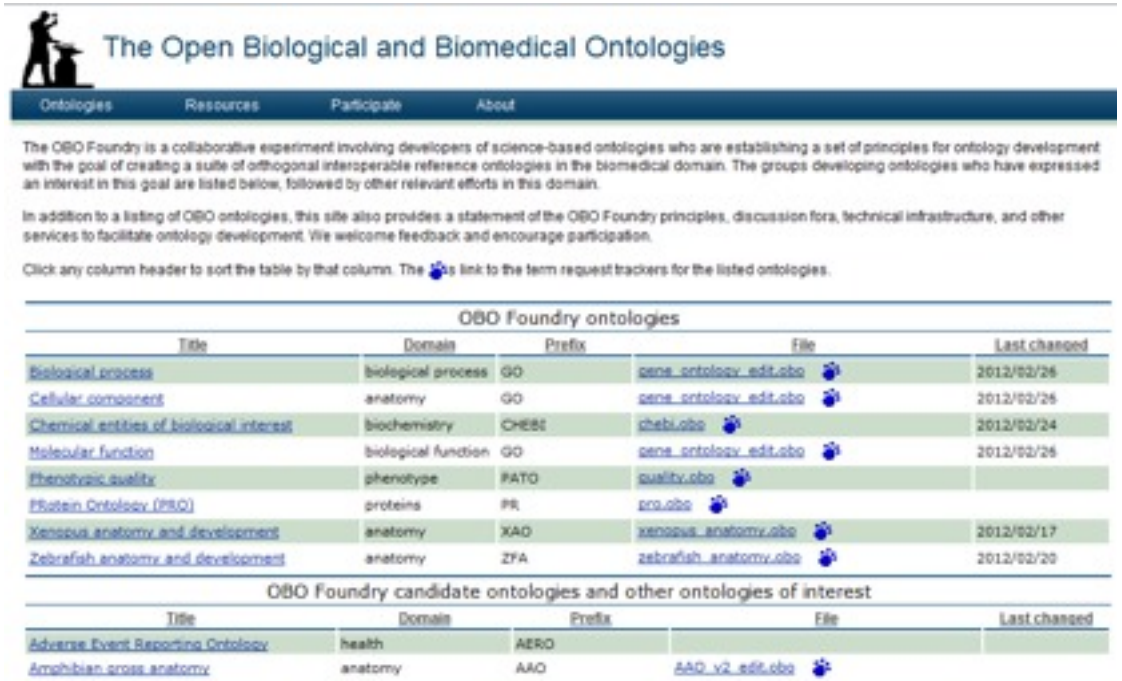
Like

go to protégé-owl

go to protégé-frames

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

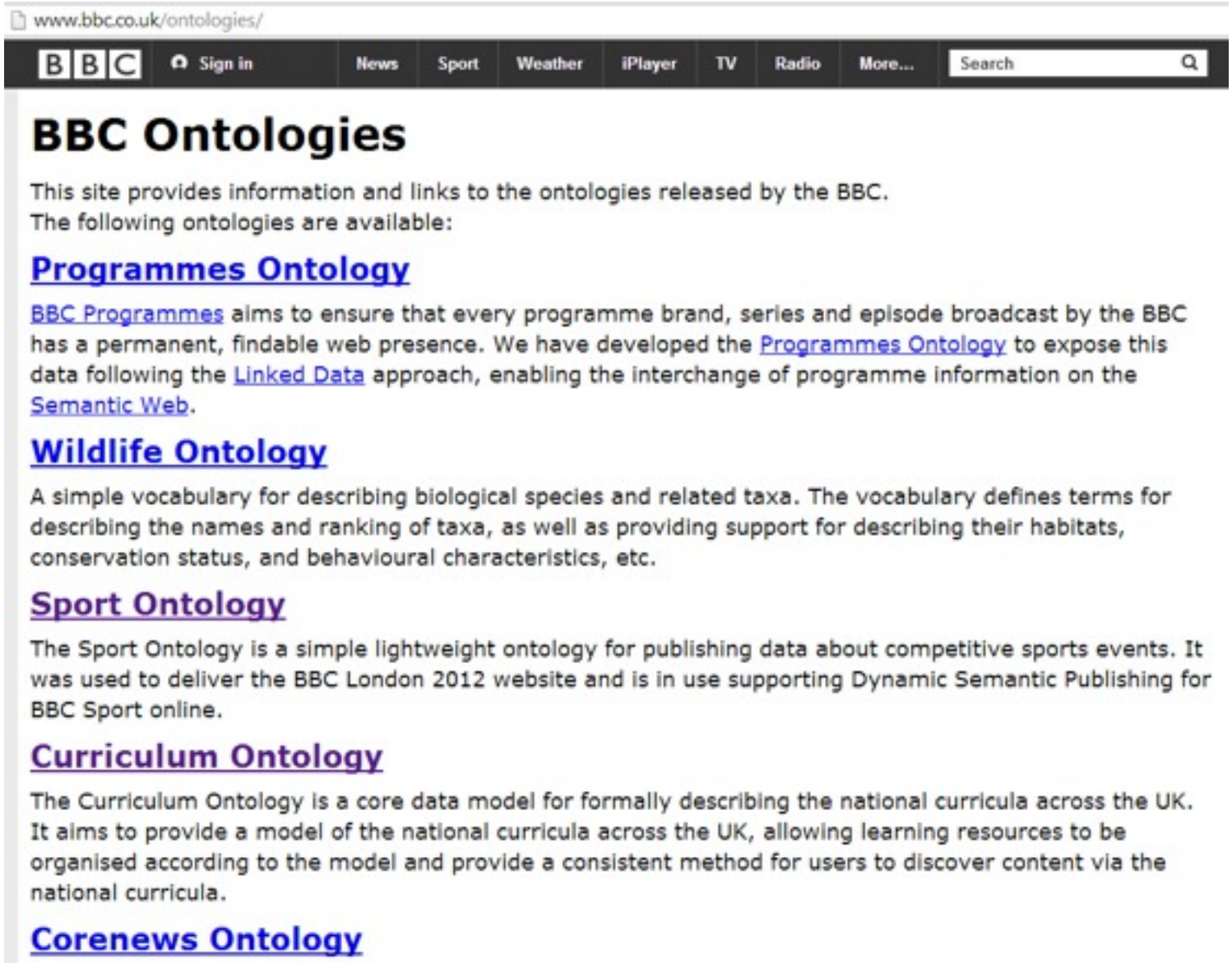
obofoundry.org



- Open Biological and Biomedical Ontologies
- a collaborative experiment involving developers of science-based ontologies
- Exercise: Pick an ontology in a domain that you can understand and analyse it, e.g. Units of measurement http://obofoundry.org/ontology/uo.html

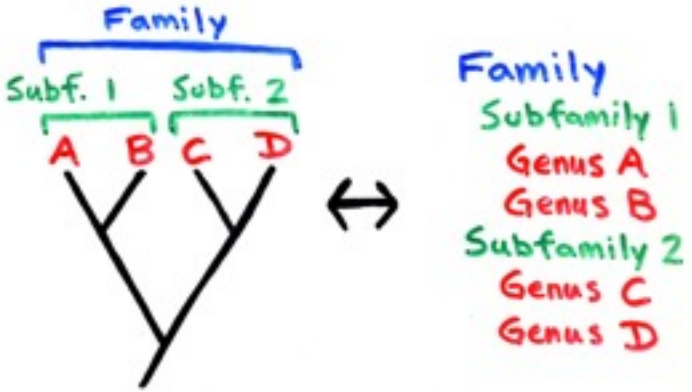
CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Ontologies in practice

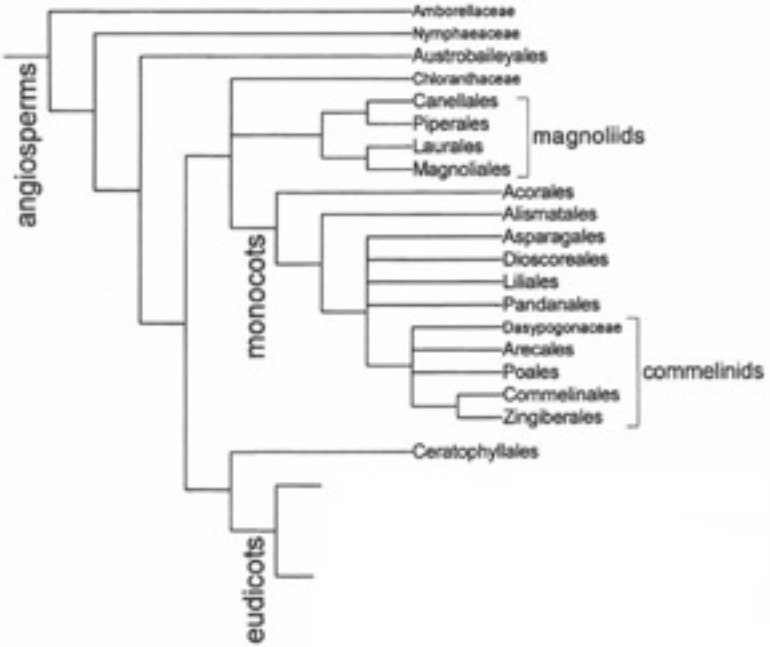


CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

What is a taxonomy then?



- strictly speaking...
- taxon (plural: taxa) = a group
- **taxonomy** = the science of identifying & naming species, and arranging them into a classification
- a resulting **taxonomy** is a hierarchically arranged classification scheme
- e.g. Linnaean taxonomy



CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Non-biological taxonomies

- almost anything may be classified according to some taxonomic scheme
- generic categories (types or classes) are arranged into a hierarchy
- human mind naturally organises its knowledge into such systems
- **taxonomy** = any hierarchically arranged classification scheme

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Uses

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Effective communication

- the most effective communication occurs when all parties involved agree on the meaning of the terms being used
- finding the right words to communicate the message
- we speak in natural language
- so much richness, variance and confusion in terminology
- we often need to impose some order to enforce agreement between the concepts & the vocabulary of people using it
- this order can come through a controlled vocabulary

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Uses

- synonyms: two words with the same meaning, e.g. car & automobile
- homonyms: words that sound the same, but have different meanings, e.g. bank the financial institution & bank the side of a river
- common misspellings
- changes in content, e.g. countries that change their name or have multiple spellings
- identifying 'best bets' or the most popular pages associated with a certain term
- woman's married & maiden name
- abbreviations & full terms, e.g. NY & New York

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Findability

- the most basic form or application of a CV is a consistent labelling system → increases findability
- calling the same thing by the same name every time
- users start developing a mental model of the information they can find
- CV inserts an interpretive layer of semantics between the term entered by the user and the underlying information resource (e.g. database)

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Uncontrolled vocabulary

- What happens when you do not use a CV?
- consider searching a site about chemistry with Ctrl + F
- iron ... or .... Fe .... or ... ferrum ?
- there is a good chance the user would miss some documents
- few users will enter ≥1 terms
- many users will be reviewing their results thinking they are seeing all relevant documents, thus possibly missing important information

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Equivalence relationship

- certain things might have multiple names
- e.g. if you said automobile on your homepage and car on the next page, users might get confused
- users will start to wonder if there is a difference between the two terms
- better to choose one, e.g. automobile and do not use car at all
- we call automobile the preferred term, and car is a variant term – a different word representing the same concept

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Example

- Gap's web site: http://www.gap.com
- excellent branding: we already know what they sell
- most of their content is generally referred to by the same terms as used in our general culture, e.g. jeans vs. denim
- furthermore, Gap does not sell hundreds of pairs of jeans that must somehow be distinguished from one another
- very simple organisation system: all concepts are consistently labelled using language familiar to their users
- works so well that they **used to not even offer search!**
- very unusual for an e-commerce site

Adding a search option

- need to translate the natural language of search into the controlled language of the web site
- a more advanced CV needs to take the concepts of users' expressed in natural language & match them to the concepts expressed in the language of a web site (CV)
- developers of the site need to create a way to tell the system "when someone searches for denim give them the results for jeans"
- jeans becomes the preferred term & denim is a variant term, and they have an equivalence relationship
- a powerful tool for increasing findability

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Case study: Amazon

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

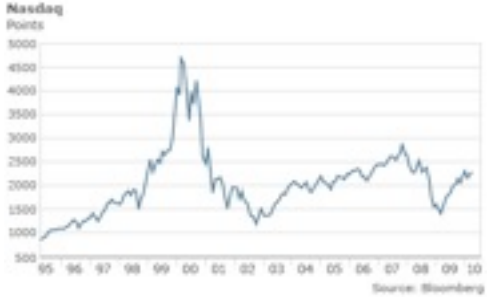
Amazon.com

- an American multinational electronic commerce company
- the world's largest online retailer
- founded by Jeff Bezos in 1994
- went online in 1995
- started as an online bookstore, but soon diversified
- Bezos wanted a name for his company to begin with A so that it would appear early in alphabetic order
- looked through the dictionary and settled on Amazon, as it was a place that was 'exotic & different' and one of the largest rivers in the world

Amazon.com

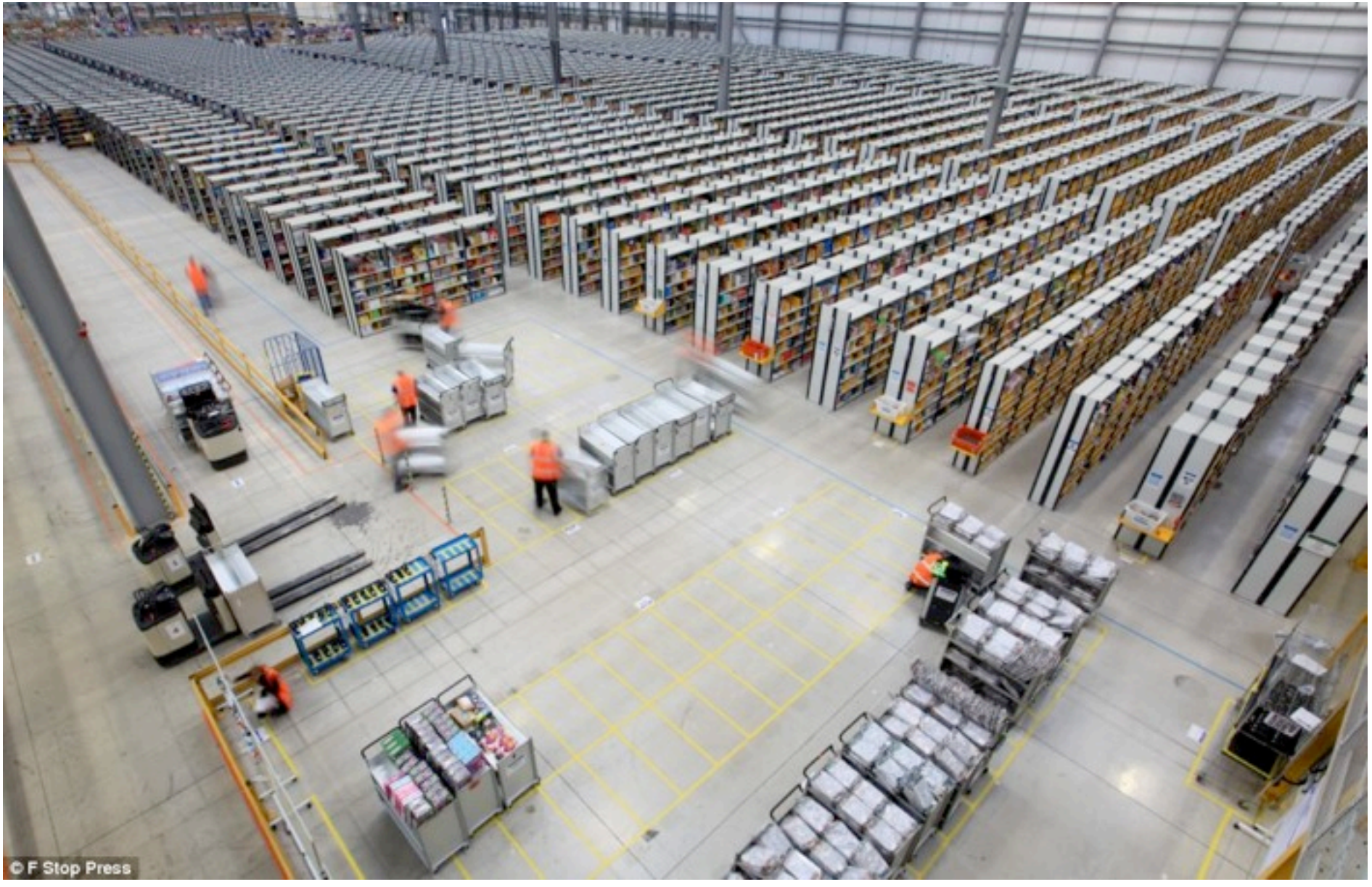
amazon.co.uk

- since 2000, Amazon's logotype is an arrow leading from A to Z
  - represents customer satisfaction
  - a goal was to have every product in the alphabet
- unusual business plan: did **not** expect a profit for 4–5 years!
- criticised by stockholders for slow growth, but when the dot-com bubble burst Amazon persevered
- turned its first profit in 2001
- Time magazine named Bezos the Person of the Year 1999



CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

amazon.co.uk



Optimizing for search & browse

- the power of the Amazon.com web site is the discoverability of products
- when buyers can find what they want, that influences conversion from looking to buying
- the technology & philosophy behind that power is the use of the single Product Detail Page combined with strong search & browse capabilities
- search & browse are the two primary methods for finding products on the Amazon.com
- sellers are encouraged to optimise their listings for both

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

- search is the primary method buyers use to locate products
- users search by entering terms in the search box
- typically words that they think of as related to the product they want to find
- the Amazon system matches relevant products based on those words and returns a page of search results
- when listing products for sale, well-chosen search terms can increase the visibility & selling potential
- Amazon searches against the words in the title, part number & standard product ID → **no** need to repeat those
- think of all the different words a buyer might use to find your product

Search terms

- if a buyer searches for a black backpack
  - you've included black in your search terms for the blue version
  - a blue backpack surfaces in their search
  - poor customer experience
- if a buyer searches for a blue backpack
  - many people think of your blue version as aquamarine
  - if aquamarine is not the true title ...
  - ... then aquamarine is a relevant search term

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Browsing

- browse is another way buyers find products
- where Amazon.com really adds value is in the browse structure presented to the buyer
- thousands of merchants sell millions of products
- to help customers find products easily, Amazon developed a detailed product hierarchy or browse-tree structure
- customers refine by category & subcategory links until they reach the most specific product type

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Browsing

- in order for products to appear when customers refine their category options, they must be classified correctly to the deepest level
- each step along a customer's browse path is called a browse node
- Amazon assigns products to ≥1 browse nodes based on how a seller classifies a product using the terms in the Item Classification Guides (ICGs)
- ICG – a category-specific document that provides proper Item Type Keywords & attribute values for setting up products

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Item Classification Guide for Grocery

How to Use this Document

This document is known as the Item Classification Guide. It contains a list of the valid terms you can use in your text file product template to populate the Item Type, Used For, Target Audience, and Other Item Attributes fields. When adding terms to these specific fields, you are giving Amazon.com information that can be used to properly classify your products for inclusion in Search and Browse. When terms are correctly applied to these fields, you are telling Amazon.com useful information about your product:

- **ItemType**: What is your product?
- **TargetAudience**: Who uses your product?
- **UsedFor**: What is the context for which your product was designed?
- **OtherItemAttributes**: What generic properties does your product have?
- **SubjectContent**: What is your product about? What image does your product depict?

Failure to use these terms will likely prevent your products from appearing on Amazon.com.
At the very least, your products will be difficult to find.

What does Amazon.com do with these terms? We use them to create search results. We also use them to assign products to predefined browse nodes. Unless your product data includes the terms assigned to a particular browse node, your products will not be included in the site's browse structure. You can refer to the Seller Central help documentation or your Account Manager for a list of the browse nodes for your category and their associated criteria.

In order to improve your product placement on Amazon.com, observe the following pointers:

- **Use as many appropriate terms as possible**: The more descriptive you are, the more likely it is that your products will surface in multiple browse cases and search results.
- **Use specific terms as opposed to general ones**: The list below is thematically organized to make it easier to find the right term. Note that not all terms map to a browse node. As a general rule, you should try to use terms that do not have subterms.
- **Make your best effort with the terms that are available**: The list is long, but it is far from exhaustive. If you cannot find the perfect term, use the next best thing. If you feel like your products don't even come close to being represented by the list below, you may be in the wrong category. Try exploring other Item Classification Guides to see if a more suitable alternative exists. If not, do

Categories for Grocery

Baking Supplies

- ItemType
- TargetAudience
- UsedFor
- OtherItemAttributes
- SubjectContent
- Genre

Terms that can go in the **ItemType** field

- baking-carob
- baking-chocolate
  - baking-cocoa
  - chocolate-chips
    - milk-chocolate-chips
    - semi-sweet-chips
  - chocolate-syrup
  - dutch-chocolate
  - mexican-chocolate
- baking-mixes
  - biscotti-mixes
  - biscuit-mixes
  - blinis-mixes
  - blintz-mixes
  - bread-mixes
    - cornbread-mixes
  - brownie-mixes
  - cake-mixes
  - cookie-mixes
  - crepe-mixes
  - custard-mixes
  - french-toast-mixes
- food-coloring
- gelatin-dessert-mixes
  > USE FOR: jell-o, jello
- graham-cracker-crumbs
- honey
  - alfalfa-honey
  - basswood-honey
  - buckwheat-honey
  - clover-honey
  - eucalyptus-honey
  - orange-blossom-honey
  - tulip-poplar-honey
  - tupelo-honey
  - wildflower-honey
- leaveners-and-yeasts
  - bakers-ammonia
  - baking-powders
  - baking-sodas
  - potash-baking-ingredient
    > USE FOR: potassium-bicarbonate
  - yeasts
    - active-dry-yeasts
    - bakers-yeasts
    - bread-machine-yeasts
    - yeast-starters
- marzipan-paste
- natural-flavoring-extracts
  > NOTE: Highly concentrated oils usually suspended in alcohol to make them easier to combine them with other foods.
  - vanilla-extracts
- non-stick-cooking-sprays
- pastry-decorations
  - pastry-decorating-cake-toppers
  - pastry-decorating-confetti
  - pastry-decorating-dragees
  - pastry-decorating-dusting-powders
  - pastry-decorating-glitter

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Classification

- seller classifies a product by describing it instead of assigning it to a particular browse node!!
- no need to reclassify a product if the browse structure changes!!
- this is an important difference between Amazon.com and other online marketplaces
- the more accurate & thorough classification, the better Amazon.com can place a product in relevant browse nodes across the site
- exercise: compare to eBay

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Lecture

- effective communication
- language & semantics
- controlled vocabulary
  - basic elements: term, definition, related/ broader/narrower term, use, used for
  - relationships: equivalence, hierarchical, associative
  - different types: flat/hierarchal list, synonym ring, authority file, classification scheme, facet vocabulary, thesaurus, ontology
  - uses
- case study: Amazon

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD