

INFO/CS 4302
Web Information Systems

FT 2012

Week 9 : Linked Data Technologies (SKOS)

- Bernhard Haslhofer -

Plan for today...

- Recap
- Discussion Homework 7
- SKOS
- Questions, Housekeeping, ...

RECAP

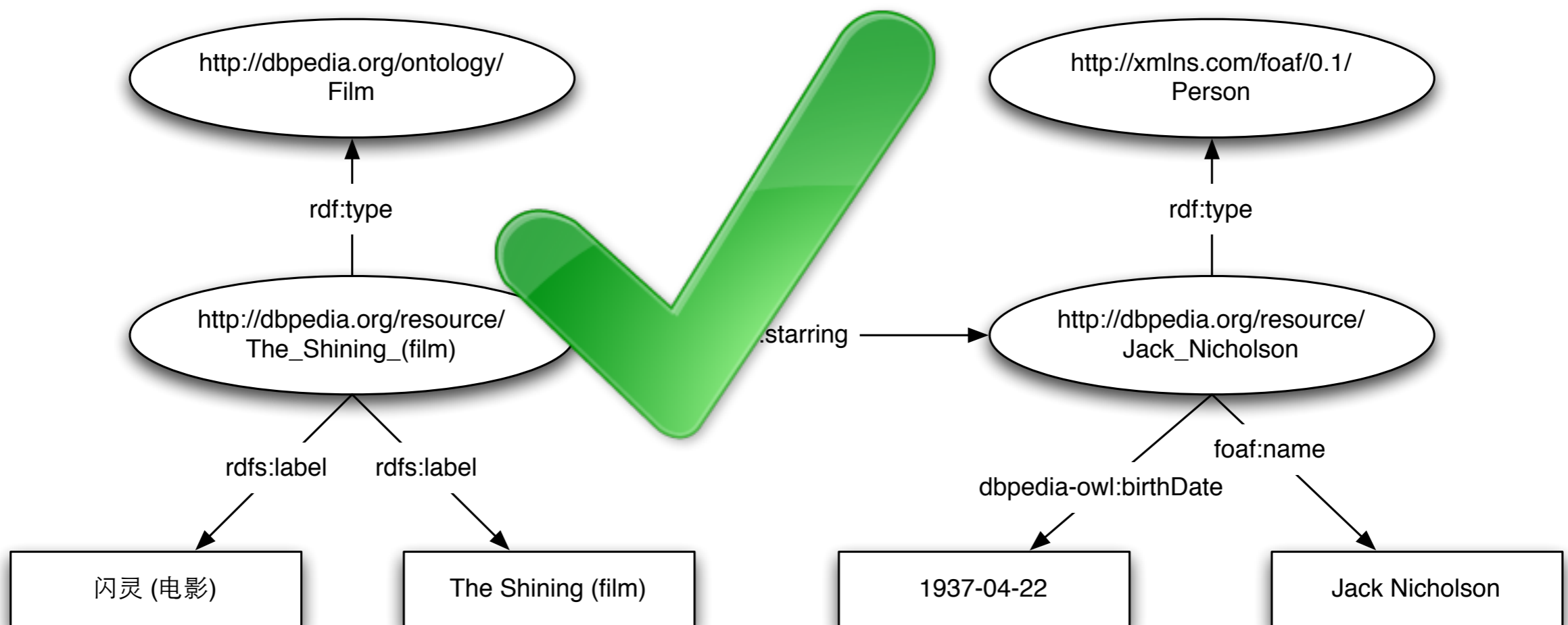
URI

- **Name** and **identify** things (resources)
- **Dereferencable** HTTP URIs



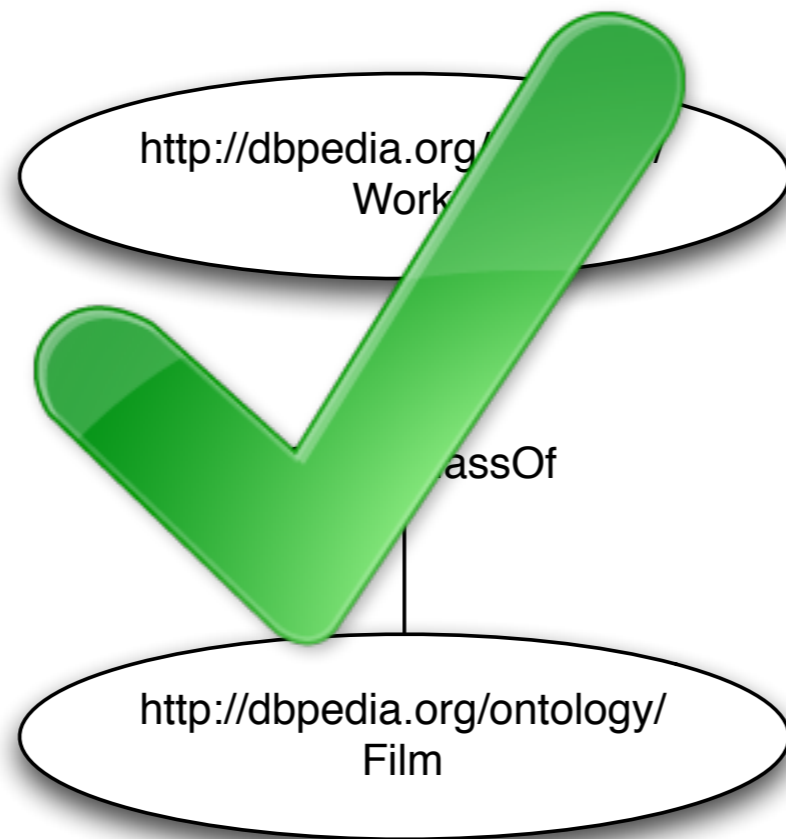
RDF

- A **data model** for representing data on the **Web**
- Several **statements** (triples) form a **graph**



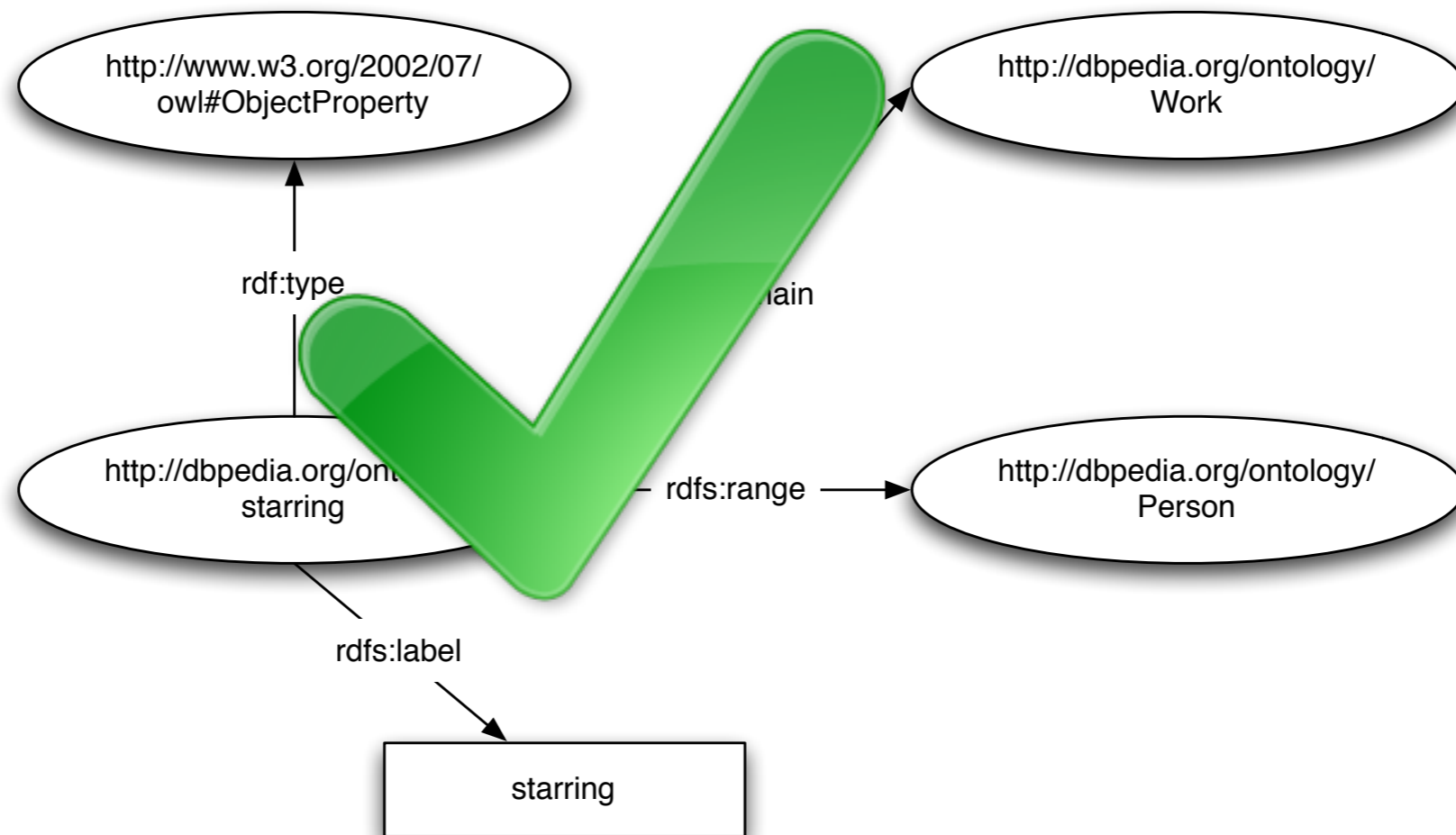
RDFS

- A **language** for describing the syntax and semantics of **vocabularies** in a machine-understandable way



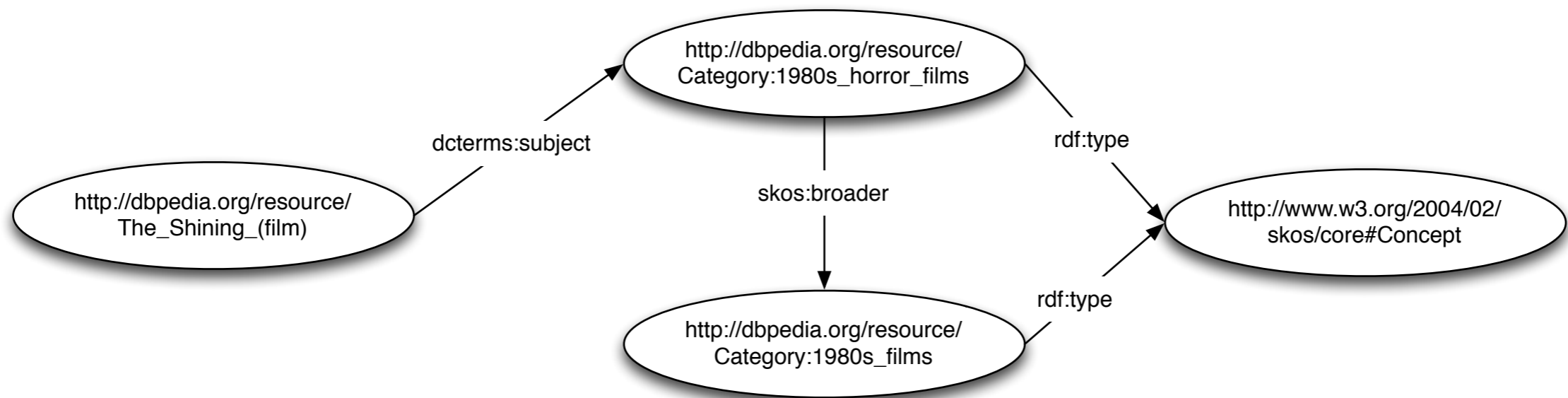
OWL

- A more expressive (formal) **language** for defining the syntax and semantics of **vocabularies**
- Solves RDFS shortcomings but introduces quite some complexity



SKOS

- A **language** for describing **controlled vocabularies** (taxonomies, thesauri, classification schemes)



SPARQL

- A **query language** and **protocol** for accessing RDF data on the Web

```
SELECT DISTINCT ?x
WHERE {
    ?x dct:subject
    <http://dbpedia.org/resource/Category:1980s_horror_films> .
}
```

RDF APIs

- Java
 - Jena Semantic Web Framework (<http://jena.apache.org/>)
 - Sesame RDF API (<http://www.openrdf.org>)
- PHP
 - ARC (<http://arc.semsol.org/>)
- Ruby
 - RDF.rb: Linked Data for Ruby (<http://rdf.rubyforge.org/>)
- Python
 - RDFLib (<https://github.com/RDFLib>)
- C
 - Redland RDF Libraries (<http://librdf.org/>)

Linked Data debugging

using cURL:

```
curl -iH "Accept: application/rdf+xml"  
http://dbpedia.org/resource/The\_Shining\_\(film\)
```

```
curl -LH "Accept: application/rdf+xml"  
http://dbpedia.org/resource/The\_Shining\_\(film\)
```

```
curl -iH "Accept: text/n3"  
http://dbpedia.org/resource/The\_Shining\_\(film\)
```

Linked Data debugging

using raptor (<http://librdf.org/raptor/>):

```
raper -o rdfxml  
http://dbpedia.org/resource/The\_Shining\_\(film\)
```

```
raper http://dbpedia.org/resource/The\_Shining\_\(film\)  
> ~/Desktop/the_shining.nt
```

DISCUSSION HW7



HW7: Basic RDF Construction and Programming (Due: 11/01 11:59pm)

The purpose of this homework is to give you experience with using the Resource Description Framework (RDF) and existing RDF libraries (APIs). You will also extend the RESTful Movie Service from the previous assignment and bring it one step further to a Linked Data service.

You can download and use the [provided code skeleton](#) for this homework.

Task 1: RDF APIs

Model the movie data set from the previous assignments as RDF statements. In as many cases as possible use elements (classes and properties) from existing vocabularies. Hint: although most facts can be expressed in these vocabularies, there are a few that can't, in which case you should use a single new namespace of your choosing.

Write a program (e.g., `convert.py`) that takes the three movie data CSV files as input, converts them into an in-memory RDF graph, and serialize that RDF graph into at least two different RDF formats (e.g., e.g., RDF/XML, Turtle) on completion. Use an existing RDF library (e.g., [RDFlib](#)) for handling RDF models in your program.

Make sure you provide instructions on how to run your program README.TXT.

Task 2: Enhance movie service to support RDF

Extend your Movie Service from the previous assignment so that it also supports at least two RDF serialization formats (e.g., RDF/XML, Turtle) for use cases UC3 (Retrieve a specific actor) and UC4 (Retrieve a specific movie). You don't need to implement resource creation / update functions based on RDF.

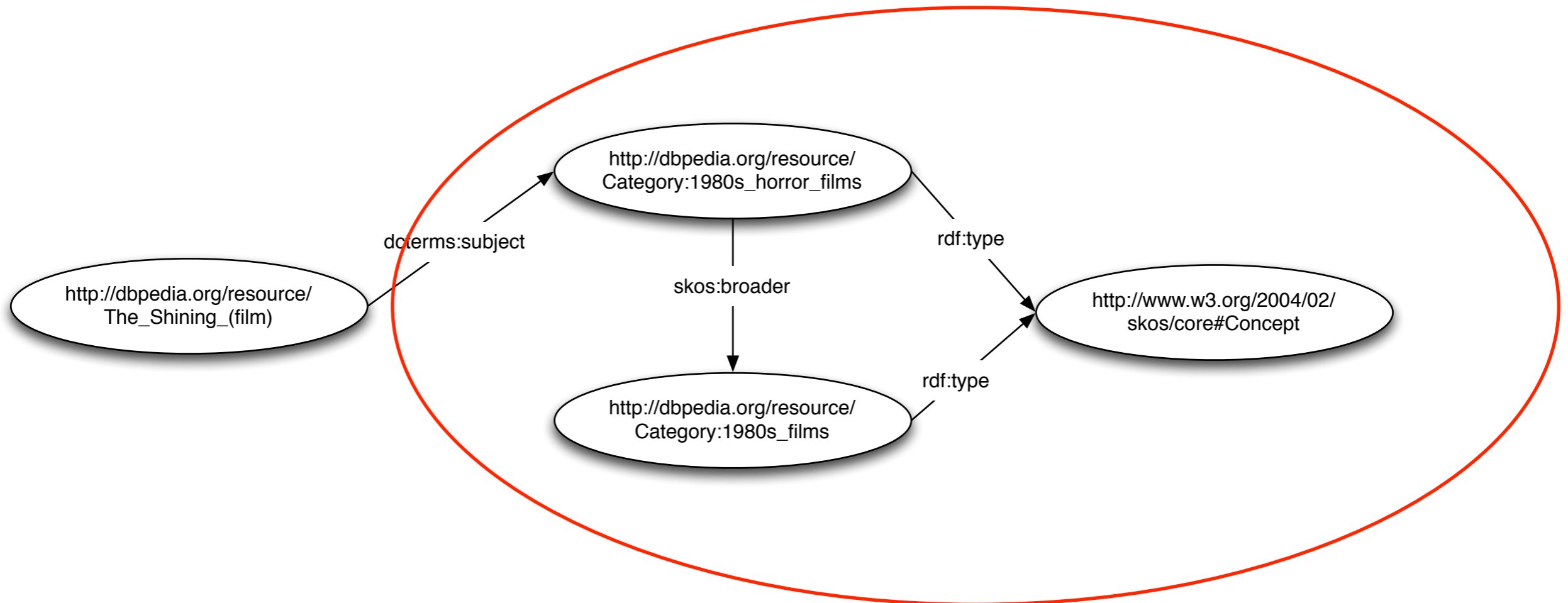
You can either use an RDF API or template mechanisms such as those provided by Tornado.

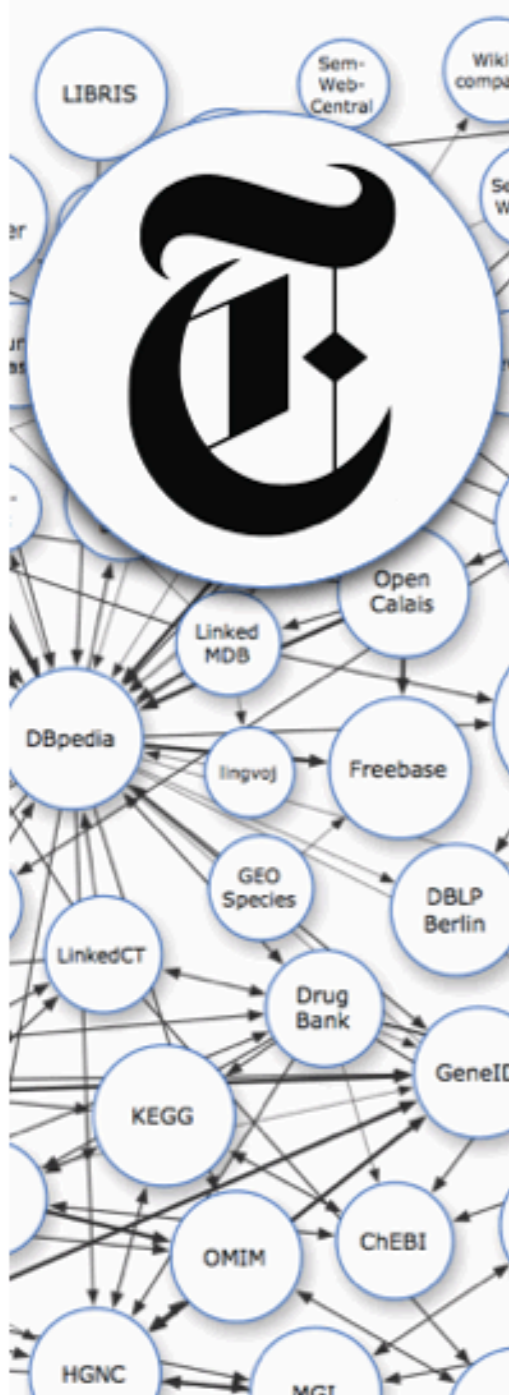
Test your implementation using [cURL](#) and/or [raptor](#) and provide example commands in README.TXT.

SKOS

Simple Knowledge Organization System (SKOS)

- A **language** for describing **knowledge organization systems** (taxonomies, thesauri, classification schemes)





data.nytimes.com

For the last 150 years, The New York Times has maintained one of the most authoritative news vocabularies ever developed. In 2009, we began to publish this vocabulary as linked open data.

The Data

As of 13 January 2010, The New York Times has published approximately 10,000 subject headings as [linked open data](#) under a [CC BY](#) license. We provide both [RDF](#) documents and a human-friendly [HTML](#) versions. The table below gives a breakdown of the various tag types and mapping strategies on data.nytimes.com.

Type	Manually Mapped Tags	Automatically Mapped Tags	Total
People	4,978	0	4,978
Organizations	1,489	1,592	3,081
Locations	1,910	0	1,910
Descriptors	498	0	498
			10,467

Browse individual data records:

[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)

SKOS Files

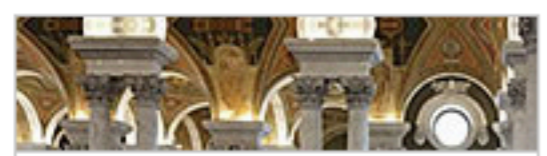
Download all of the data records as [SKOS](#) Files.

· [People](#)



AUTHORITIES & VOCABULARIES

Authorities and Vocabularies



- ➔ [Authorities & Vocabularies](#)
- ➔ [About](#)
- ➔ [Vocabulary Descriptions](#)
- ➔ [Search](#)
- ➔ [Download](#)
- ➔ [Technical Center](#)
- ➔ [Contact Us](#)



About

The Library of Congress Authorities and Vocabularies service enables both humans and machines to programmatically access authority data at the Library of Congress via URIs.

[Read more »](#)

Search Authorities & Vocabularies

Hold CONTROL key for 2 or more



Food and Agriculture
Organization of the
United Nations

for a world without hunger



AIMS

Agricultural
Information
Management
Standards

Username: *

Password: *

LOG IN >

Log in using OpenID

Home » Vocabularies

AGROVOC

[Search Terms](#)
[Browse Hierarchies](#)
[Concept Scheme](#)
[Maintenance](#)
[Copyright & Download](#)
[Linked Data](#)

The AGROVOC thesaurus contains more than 40 000 concepts in up to 22 languages covering topics related to food, nutrition, agriculture, fisheries, forestry, environment and other related domains.

A global [community of editors](#) consisting of librarians, terminologists, information managers and software developers, maintain AGROVOC using [VocBench](#), an open-source multilingual, web-based vocabulary editor and workflow management tool that allows simultaneous, distributed editing. AGROVOC is expressed in [Simple Knowledge Organization System \(SKOS\)](#) and published as [Linked Data](#).

- Explore [search terms](#) or [browse hierarchies](#)
- [Download](#) or access via [web services](#)
- [Suggest new terms](#) not present yet

LATEST NEWS

["Design and Implementation of the ZH/EN Bilingual Retrieval System Based on the CAT/AGROVOC Mapping"](#)

13 Sep 2012

[AIMS at Agricultural Ontology Service \(AOS\) 2012](#)

05 Sep 2012

[Dutch FSA to complete project "Apollon" for detection of Emerging Risks in Food Chains](#)

06 Aug 2012

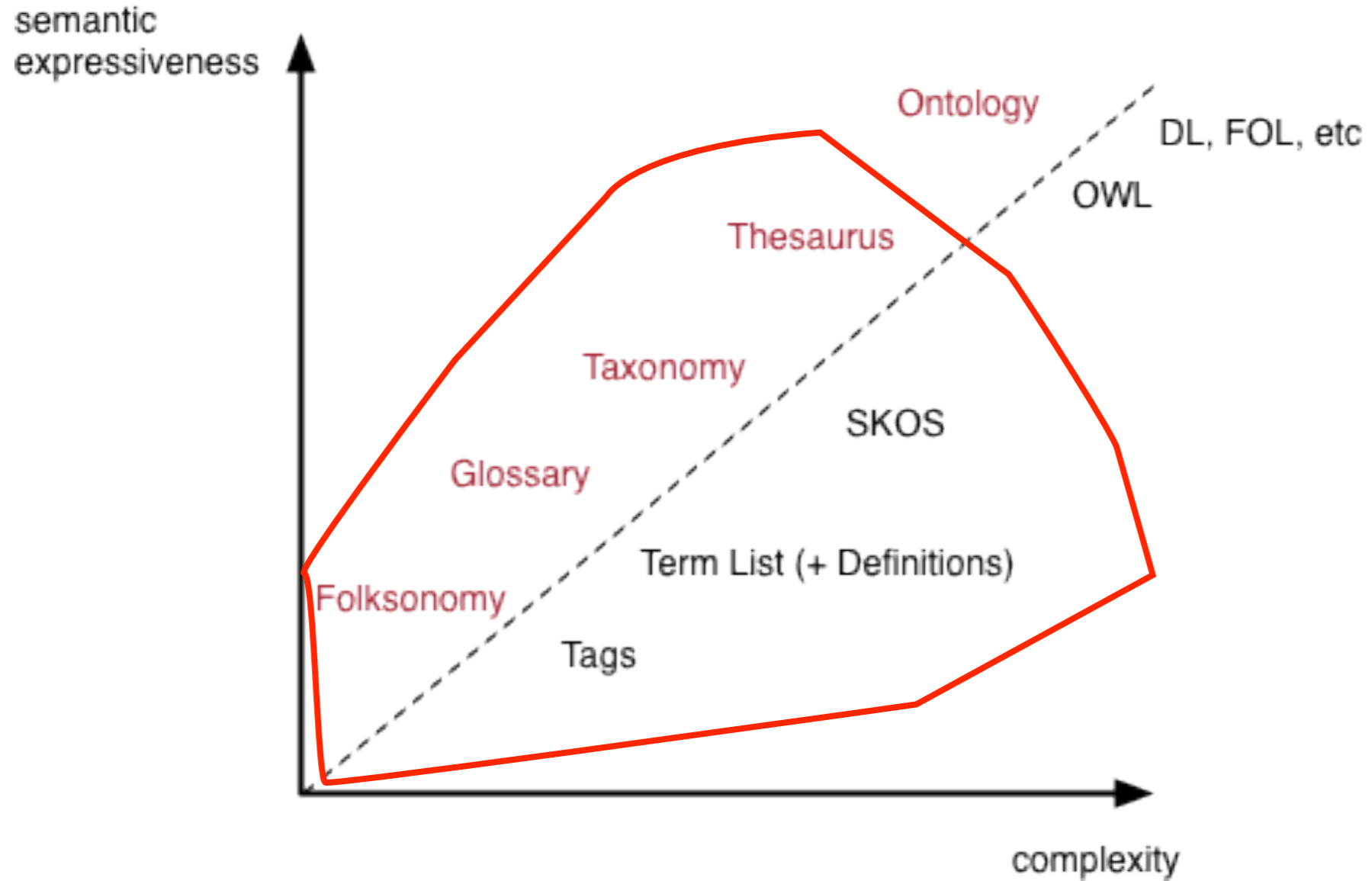
[AGROVOC 2012 edition released](#)

17 Jul 2012

[International Conference on Information and Knowledge Dynamics, Bangalore](#)

15 Jul 2012

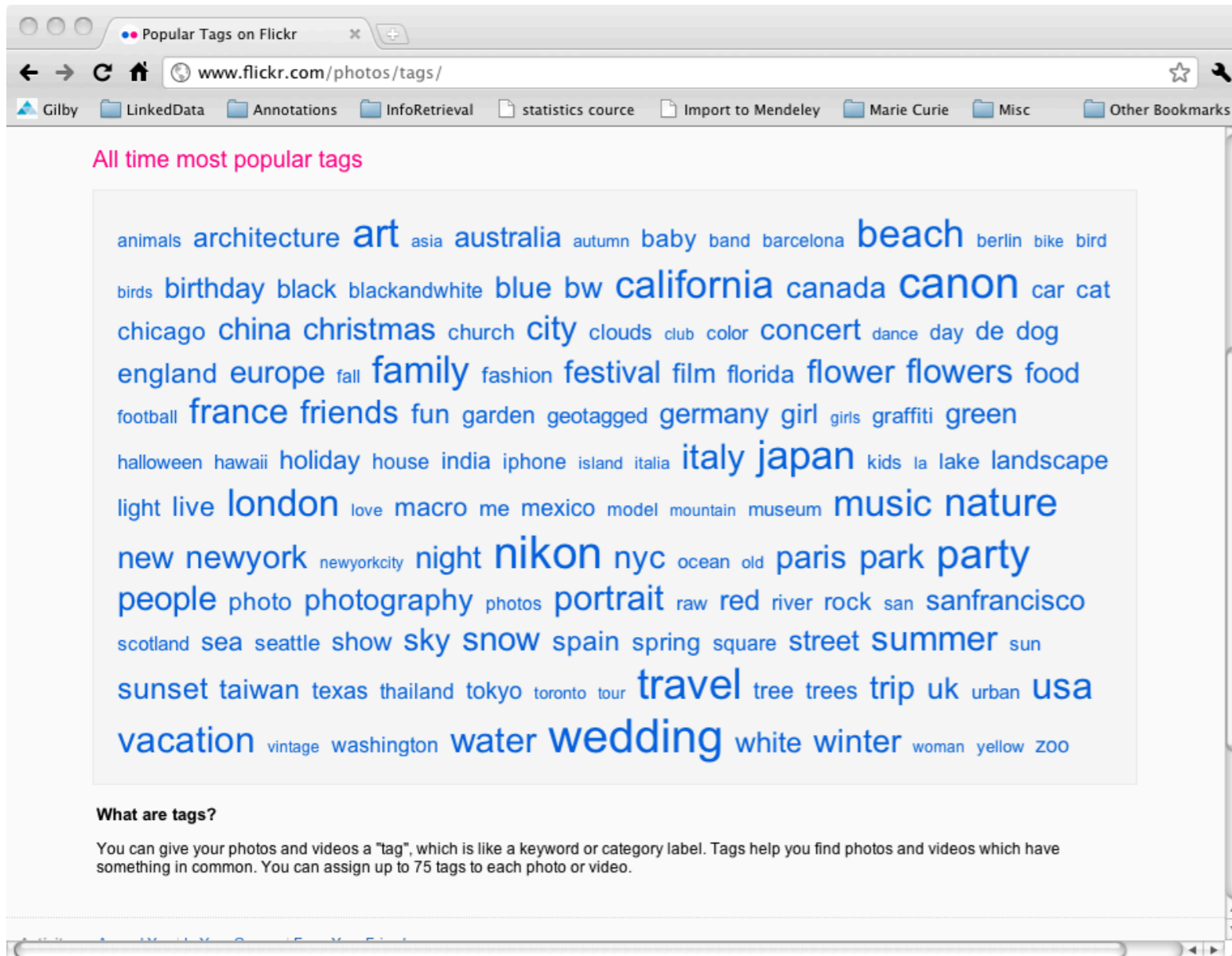
Knowledge Organization Systems (KOS)



Folksonomy

- **Not controlled** by some central authority
- Usually no explicit definition of meanings
- Used in
 - collaborative tagging systems
 - social media

Folksonomy Examples



twitter 

#info4302

Controlled Vocabulary

- A simple list of terms, definitions and naming conventions
- Some process “**controls**” adding and removing term definitions to ensure **consistency**
- Terms are often defined in **relationship** to each other
- **Glossaries, Taxonomies, and Thesauri** are example types of controlled vocabularies

Taxonomy

- Classification of entities in a **hierarchy**
- A taxonomy **node** stands for **a real-world entity type**

North American Product Classification System

You are here: [Census.gov](#) > [Business & Industry](#) > North American Product Classification System

[Main](#)[NAPCS Discussion Papers](#)[NAPCS Product Lists](#)[Related Documents](#)[NAICS](#)[FAQs](#)

Site Resources

[What's New](#)

Contact Us

[Email Us](#)

Introduction

In February 1999, the statistical agencies of Canada, Mexico, and the United States launched a joint multi-phase initiative to develop a comprehensive demand-oriented product classification, known as the North American Product Classification System (NAPCS). Work to date has focused on the products produced by service industries in 12 NAICS sectors 48-49 through 81. With that work nearing completion, this web page provides an overview of and progress report on the NAPCS initiative and presents the final versions of the product lists developed so far for the service industries included in those 12 sectors.

[Overview and Progress Report](#)

Final NAPCS Product Lists

- [Description of product lists](#)
- [View PDF files of individual product lists](#)
- [Download sequential Excel file of all product lists](#)

Thesaurus

- Associates the meaning of a term with the meaning of another term
- hypernym, hyponym, synonym, antonym

WordNet Search - 3.0 - [WordNet home page](#) - [Glossary](#) - [Help](#)Word to search for: Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Noun

- **S: (n) cat, true cat** (feline mammal usually having thick soft fur and no ability to roar: domestic cats; wildcats)
- **S: (n) guy, cat, hombre, bozo** (an informal term for a youth or man) "*a nice guy*"; "*the guy's only doing it for some doll*"
- **S: (n) cat** (a spiteful woman gossip) "*what a cat she is!*"
- **S: (n) kat, khat, qat, quat, cat, Arabian tea, African tea** (the leaves of the shrub *Catha edulis* which are chewed like tobacco or used to make tea; has the effect of a euphoric stimulant) "*in Yemen kat is used daily by 85% of adults*"
- **S: (n) cat-o'-nine-tails, cat** (a whip with nine knotted cords) "*British sailors feared the cat*"
- **S: (n) Caterpillar, cat** (a large tracked vehicle that is propelled by two endless metal belts; frequently used for moving earth in construction and farm work)
- **S: (n) big cat, cat** (any of several large cats typically able to roar and living in the wild)
- **S: (n) computerized tomography, computed tomography, CT, computerized axial tomography, computed axial tomography, CAT** (a method of examining body organs by scanning them with X rays and using a computer to construct a series of cross-sectional scans along a single axis)

Verb

- **S: (v) cat** (beat with a cat-o'-nine-tails)
- **S: (v) vomit, vomit up, purge, cast, sick, cat, be sick, disgorge, regorge, retch, puke, barf, spew, spue, chuck, upchuck, honk, regurgitate, throw up** (eject the contents of the stomach through the mouth) "*After drinking too much, the students vomited*"; "*He purged continuously*"; "*The patient regurgitated the food we gave him last night*"

[WordNet home page](#)

National Library of Medicine - Medical Subject Headings

2010 MeSH

MeSH Descriptor Data

[Return to Entry Page](#)

Standard View. [Go to Concept View](#); [Go to Expanded Concept View](#)

MeSH Heading	Heart
Tree Number	A07.541
Annotation	heart as pump, not tissue (MYOCARDIUM); qualif permitted: / anat / drug eff / embryol (FETAL HEART also exists) / growth / innerv (HEART CONDUCTION SYSTEM also exists) / microbiol / parasitol / physiol (see also ATRIAL FUNCTION & VENTRICULAR FUNCTION & their specifics; see also MYOCARDIAL CONTRACTION & its specifics) / physiopathol / rad eff / radiogr (do not confuse with ANGIOCARDIOGRAPHY) / radionuclide / virol ; / blood supply = CORONARY VESSELS but consider also CORONARY CIRCULATION ; / drug eff : consider also MYOCARDIAL DEPRESSANTS see ANTI-ARRHYTHMIA AGENTS & CARDIOTONIC AGENTS ; / transpl = HEART TRANSPLANTATION : do not coord with TRANSPLANTATION, HOMOLOGOUS unless particularly discussed; / ultrasonogr = ECHOCARDIOGRAPHY or ECHOCARDIOGRAPHY, DOPPLER ; inotropism: INOTROPISM, CARDIAC see MYOCARDIAL CONTRACTION & INOTROPIC AGENTS, POSITIVE CARDIAC are available but see MYOCARDIAL CONTRACTION note; chronotropism: see HEART RATE note; DIASTOLE & SYSTOLE are also available; heart-lung prep: index HEART / physiol (probably NIM); Manual 21.48+ ; carditis = MYOCARDITIS ; mural thrombosis: index HEART DISEASE (IM) + THROMBOSIS (IM)
Scope Note	The hollow, muscular organ that maintains the circulation of the blood.

Why do we need “knowledge organization” in the age of full-text search engines?

craigslist

post to classifieds
my account
help, faq, abuse, legal

search craigslist

for sale

event calendar

S	M	T	W	T	F	S
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31	1	2	3	4	5
6	7	8	9	10	11	12

avoid scams & fraud
personal safety tips
craigslist blog
craigslist factsheet
best-of-craigslist
craigslist TV
craigslist movie & dvd
craigslist foundation
craigconnects
system status
terms of use about
privacy help

ithaca ^w

community

activities lost+found
artists musicians
childcare local news
general politics
groups rideshare
pets volunteers
events classes

personals

strictly platonic
women seek women
women seeking men
men seeking women
men seeking men
misc romance
casual encounters
missed connections
rants and raves

discussion forums

1099	gifts	pets
apple	haiku	philos
arts	health	politic
atheist	help	psych
autos	history	queer
beauty	housing	recover
bikes	jobs	religion
celebs	jokes	rofo
comp	kink	science
crafts	l.t.r.	shop
diet	legal	spirit
divorce	linux	sports
dying	loc pol	t.v.
eco	m4m	tax
educ	money	testing
etiquet	motocy	transg
feedbk	music	travel
film	npo	vegan
fitness	open	w4w

housing

apts / housing
rooms / shared
sublets / temporary
housing wanted
housing swap
vacation rentals
parking / storage
office / commercial
real estate for sale

for sale

appliances	arts+crafts
antiques	auto parts
barter	baby+kids
bikes	beauty+hlth
boats	cars+trucks
books	cds/dvd/vhs
business	cell phones
computer	clothes+acc
free	collectibles
furniture	electronics
general	farm+garden
jewelry	garage sale
materials	household
rvs	motorcycles
sporting	music instr
tickets	photo+video
tools	toys+games
wanted	video gaming

services

beauty	automotive
creative	farm+garden
computer	household
cycle	labor/move
event	skill'd trade
financial	real estate
legal	sm biz ads

jobs

accounting+finance
admin / office
arch / engineering
art / media / design
biotech / science
business / mgmt
customer service
education
food / bev / hosp
general labor
government
human resources
internet engineers
legal / paralegal
manufacturing
marketing / pr / ad
medical / health
nonprofit sector
real estate
retail / wholesale
sales / biz dev
salon / spa / fitness
security
skilled trade / craft
software / qa / dba
systems / network
technical support
transport
tv / film / video
web / info design
writing / editing
[ETC]
[part-time]

gigs

crew	computer
event	creative
labor	domestic
talent	writing

nearby cl

albany
allentown
belleville
binghamton
buffalo
catskills
elmira
finger lakes
ithaca
kingston
oneonta
poconos
rochester
scranton
state college
syracuse
twin tiers
utica
watertown
williamsport

us cities

us states

canada

cl worldwide



Your Amazon.com | Today's Deals | Gift Cards | Help

The All-New kindle fire HD



Shop by Department

Search

All

Go

Hello. Sign in Your Account

Cart

Wish List

- Unlimited Instant Videos
- MP3s & Cloud Player
20 million songs, play anywhere
- Amazon Cloud Drive
5 GB of free storage
- Kindle
- Appstore for Android
Get Call of Atlantis free today
- Digital Games & Software
- Audible Audiobooks

- Books
- Movies, Music & Games
- Electronics & Computers**
- Home, Garden & Tools
- Grocery, Health & Beauty
- Toys, Kids & Baby
- Clothing, Shoes & Jewelry
- Sports & Outdoors
- Automotive & Industrial

Full Store Directory

Electronics

- TV & Video
- Home Audio & Theater
- Camera, Photo & Video
- Cell Phones & Accessories
- Video Games
- MP3 Players & Accessories
- Car Electronics & GPS
- Appliances
- Musical Instruments

Computers

- Laptops, Tablets & Netbooks
- Desktops & Servers
- Computer Accessories & Peripherals
External drives, mice, networking & more
- Computer Parts & Components
- Software
- PC Games
- Printers & Ink
- Office & School Supplies

Car Electronics Resource Center

Learn more



Get Charged with Powerful and Versatile Equipment

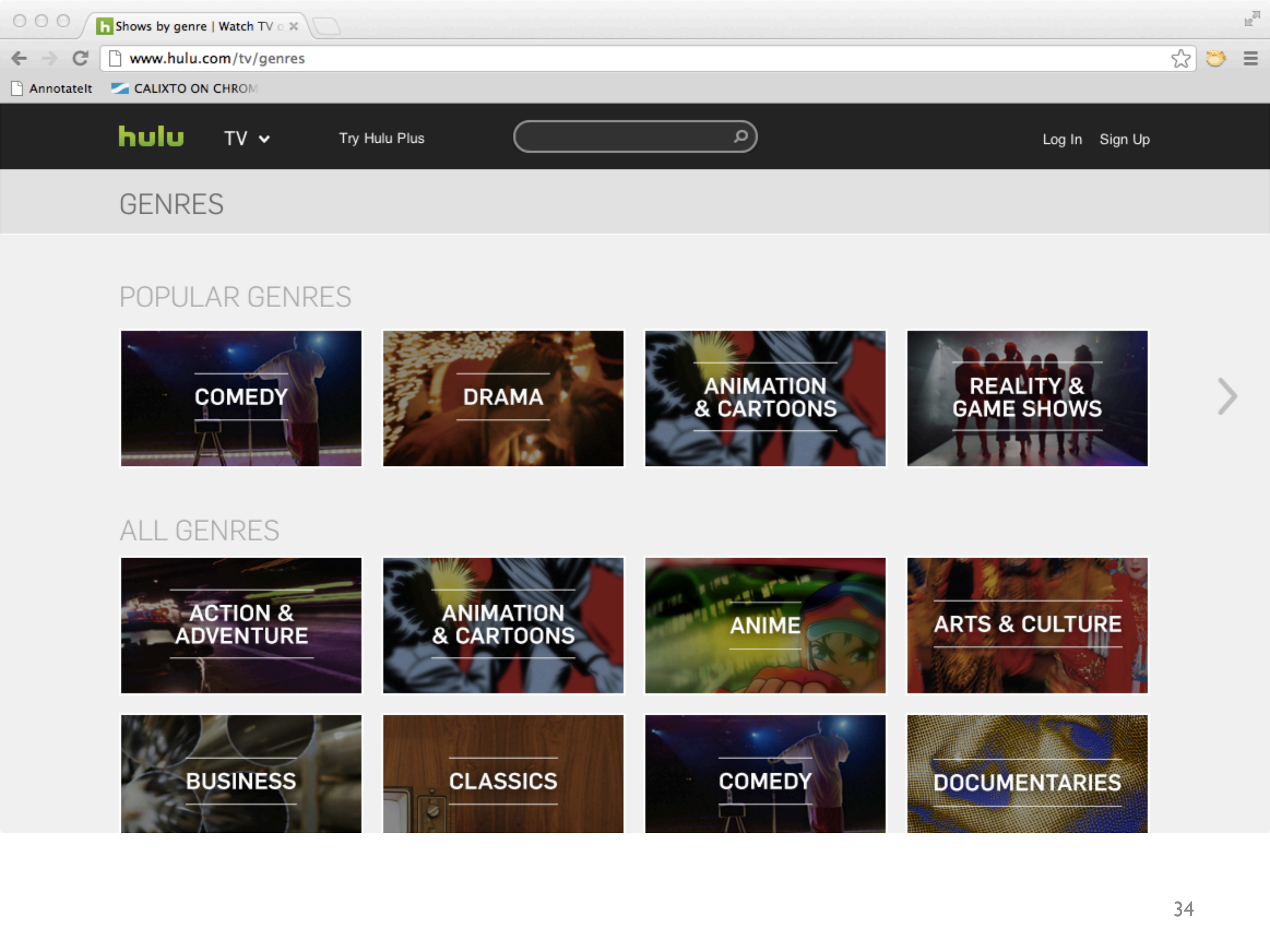
amazon Get the Free Amazon Mobile App
Search & buy millions of products on the go
[Learn more](#)

GERARD BUTLER
CHASING MAVERICKS
IN THEATERS OCTOBER 26TH
WATCH TRAILER [LIVELIKEJAY.COM](#)

Advertisement

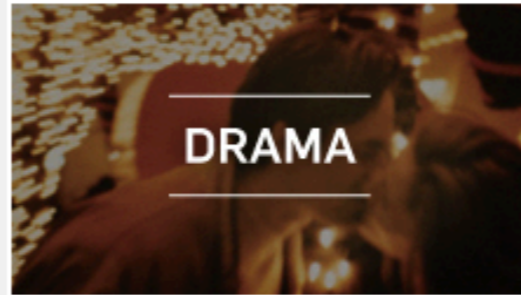
Introducing Nest
The Learning Thermostat
[Learn more](#)

amazon.com \$20 OFF

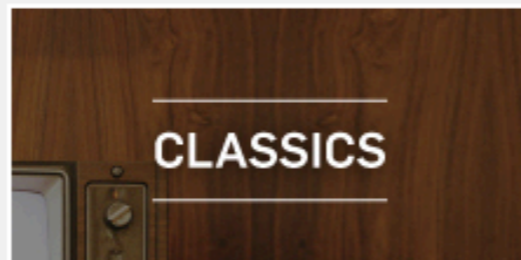
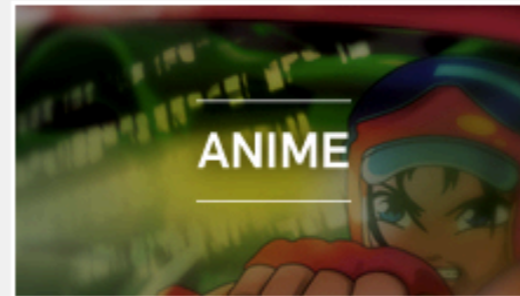


GENRES

POPULAR GENRES



ALL GENRES



What is SKOS?

- A model for expressing the basic structure and content of concept schemes such as **thesauri, classification schemes, taxonomies, folksonomies**, and other similar types of **controlled vocabularies**
- Allows concepts to be composed and published as **Linked Data** on the Web
- Hides the complexity of OWL - **easy to use**

SKOS Concepts are...

- ... identified by **URIs**
- ... **labeled** with 1..* natural language strings
- ... documented with various types of **notes**
- ... semantically **linked** to each other
- ... aggregated into **concept schemes**

Example SKOS Concept

The screenshot shows a web browser window with the address bar containing `id.loc.gov/authorities/sh2001000475`. The page header includes the Library of Congress logo and navigation links: [ASK A LIBRARIAN](#), [DIGITAL COLLECTIONS](#), and [LIBRARY CATALOGS](#). A search bar with a **GO** button and an [Options](#) link is also present. The breadcrumb trail reads: [The Library of Congress](#) > [Authorities & Vocabularies](#) > [LC Subject Headings](#) > [Parody films](#).

Parody films

From Library of Congress Subject Headings

Details | Visualization | Suggest Terminology

Parody films

This heading is used as a topical heading for works about films that comically imitate another work or group of works of a more serious nature. When used as a topical heading it is subdivided by the appropriate geographic, topical, and/or form subdivisions.

General works about the use of parody in motion pictures are entered under Parody in motion pictures.

URI
<<http://id.loc.gov/authorities/sh2001000475#concept>>

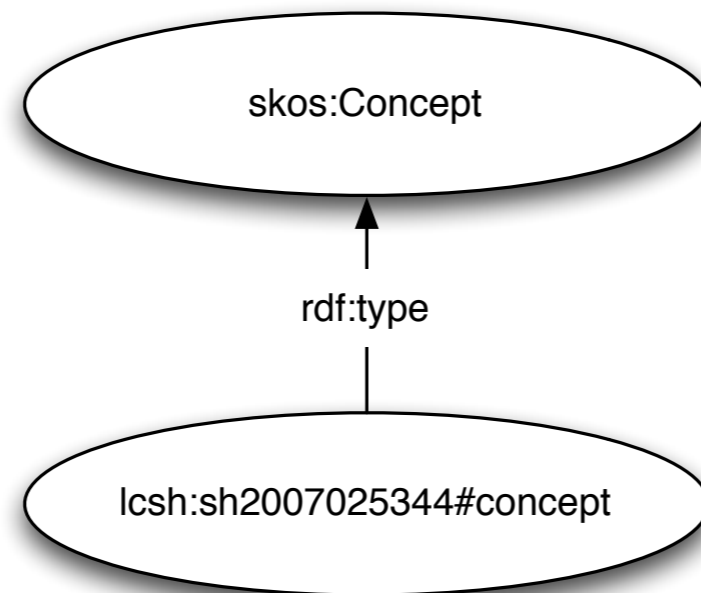
Type
Topical Term

Alternate Labels

- > Film genre parodies
- > Film parodies
- > Genre parodies (Motion pictures)
- > Genre parody films
- > Motion picture parodies

skos:Concept

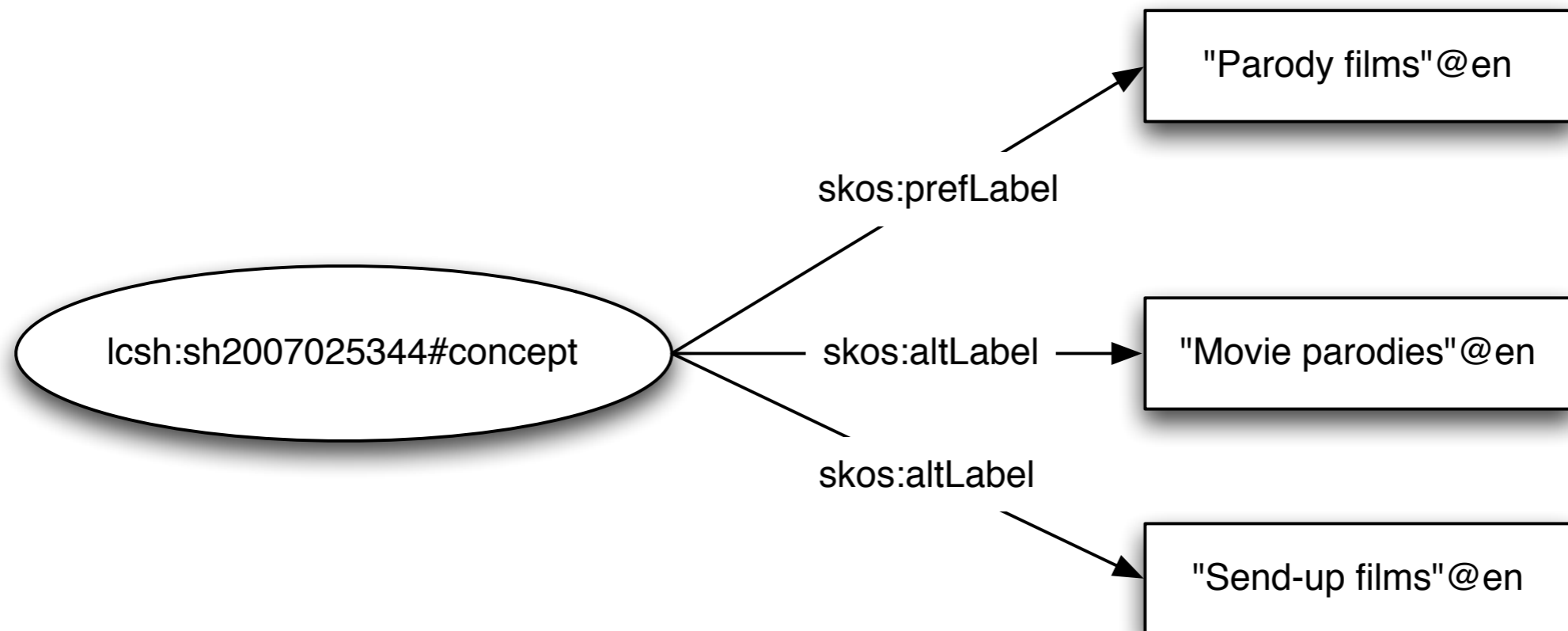
- Concepts are
 - the **units of thought**: ideas, meanings, categories of objects, etc.
 - **abstract entities** which are independent of the terms used to label them



```
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .  
@prefix lcsh: <http://id.loc.gov/authorities/> .
```

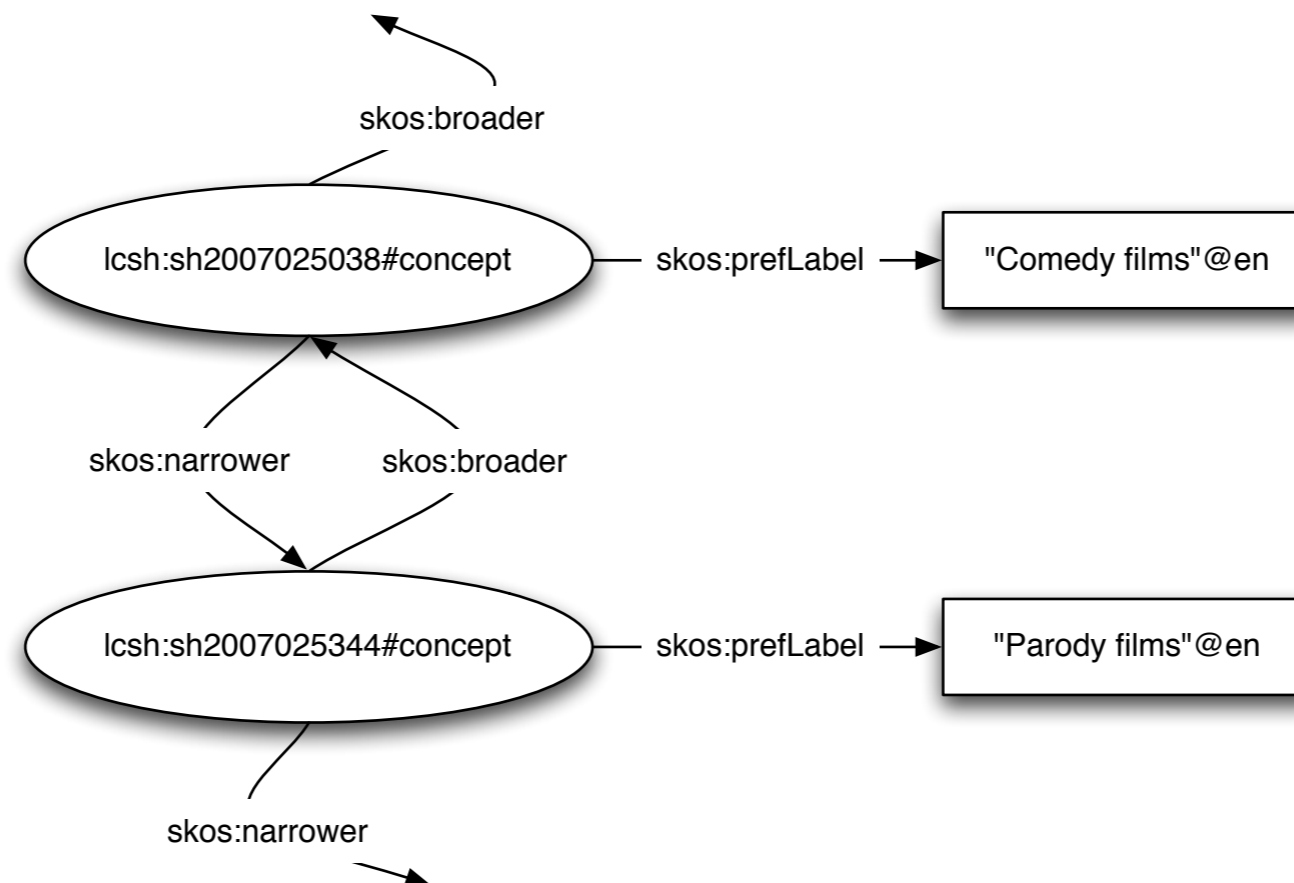
skos:(pref|alt|hidden)Label

- Labels refer to concepts' natural language(s)
 - **skos:prefLabel**: the preferred lexical label
 - **skos:altLabel**: alternative lexical labels (e.g., synonyms)
 - **skos:hiddenLabel**: labels useful for indexing (e.g., typos)



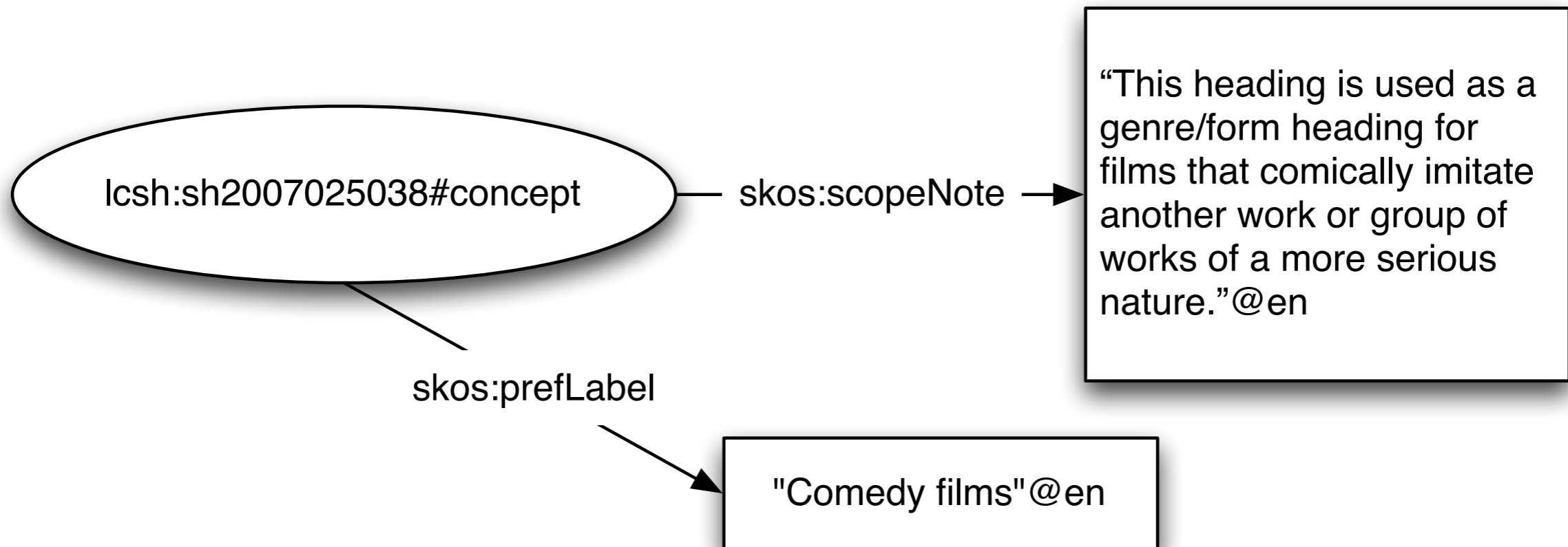
SKOS Semantic Relationships

- The meaning of a concept is also defined by its links to other concepts
 - **skos:broader**: hierarchical link to a more general concept
 - **skos:narrower**: hierarchical link to a more specific concept
 - **skos:related**: associative (non-hierarchical) link



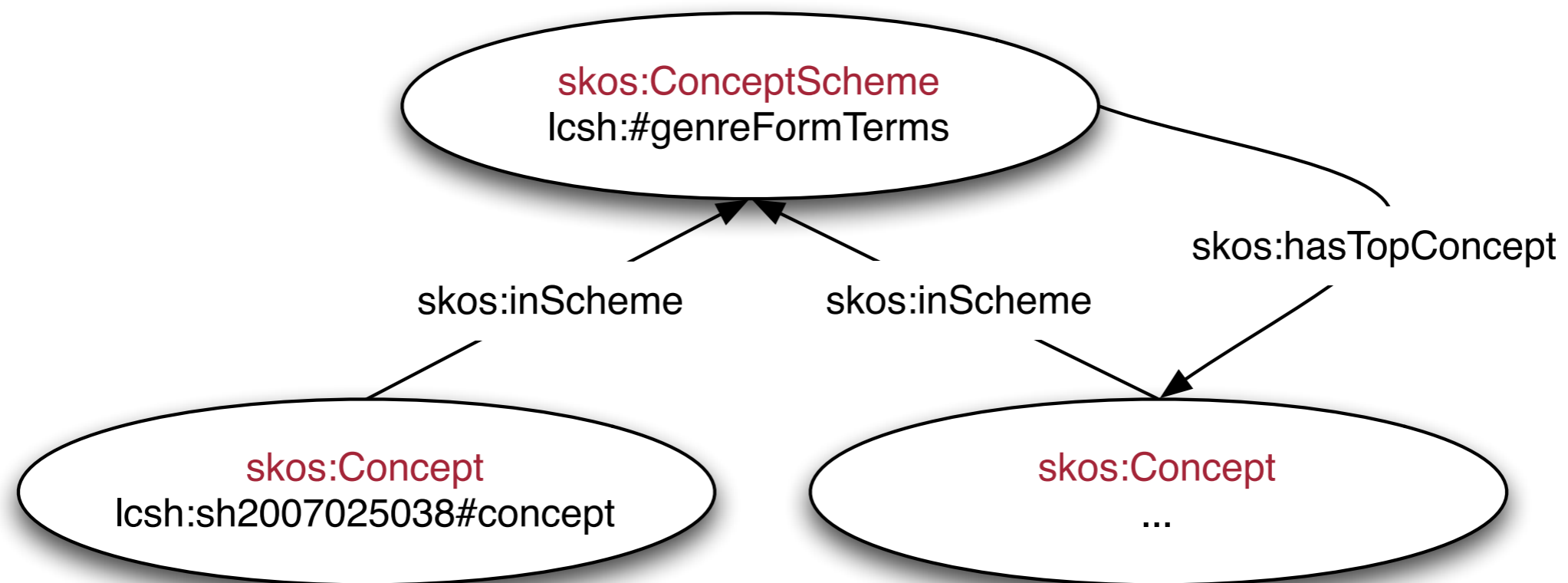
SKOS Documentary Notes

- Add further **human-readable** documentation
 - **skos:scopeNote**: info about intended meaning
 - **skos:definition**: complete explanation of meaning
 - **skos:example**: example concept use



skos:ConceptScheme

- Allow the organization of skos:Concepts in some Knowledge Organization Scheme (KOS)



Another Real-World Examples

- NY Times topics:
<http://topics.nytimes.com/>
- Dereference and analyze
“Jack Nicholson” @ New York Times
 - <http://data.nytimes.com>
 - [http://data.nytimes.com/
N5761411277431266513](http://data.nytimes.com/N5761411277431266513)

QUESTIONS & HOUSEKEEPING