welcome ...



source: jetzt.sueddeutsche.de

Welcome to

THESAURUS

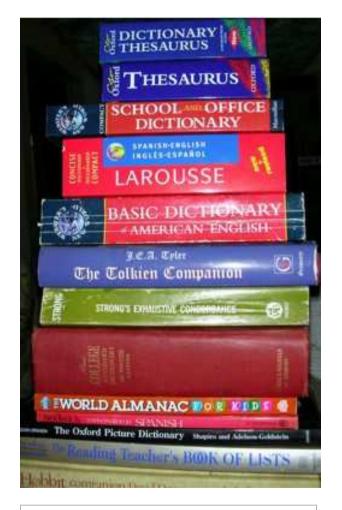
Please drive

carefully, warily, attentively, heedfully, cautiously, conscientiously, thoughtfully

through

the village, community, town, hamlet, settlement.

treasuries of words



source: Lifehack

Authority File

Keyword List

Glossary

Classification Scheme

Synonym List

Thesaurus

Taxonomy

Folksonomy

Controlled Vocabulary

Ontology

Topic Map

Controlled Vocabularies

what they are and how they perform in various contexts

Frankfurt am Main, 30 May 2011

what users may be looking for



- a special item or fact (known item search)
- collection of items or facts as search result
- guidance for discovery (serendipity)

what users want and don't want

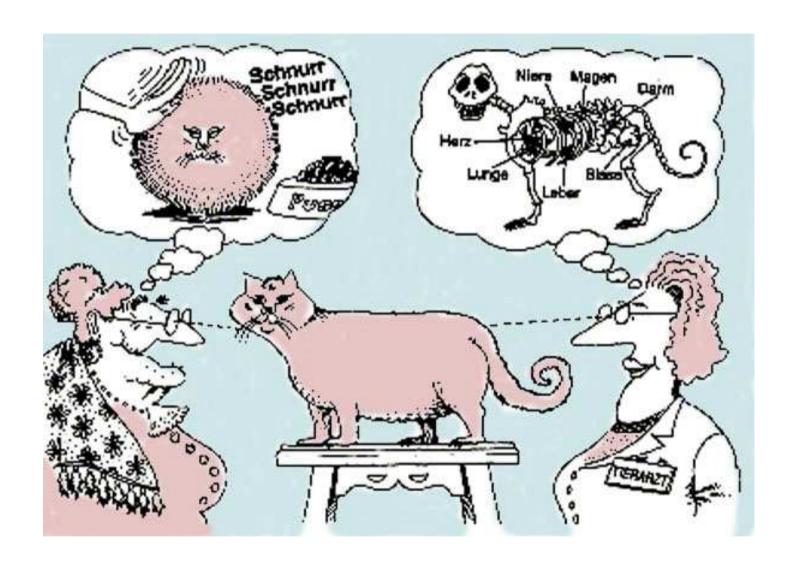
Users want to retrieve

- all interesting documents (recall)
- only interesting documents (precision)

Users don't want to

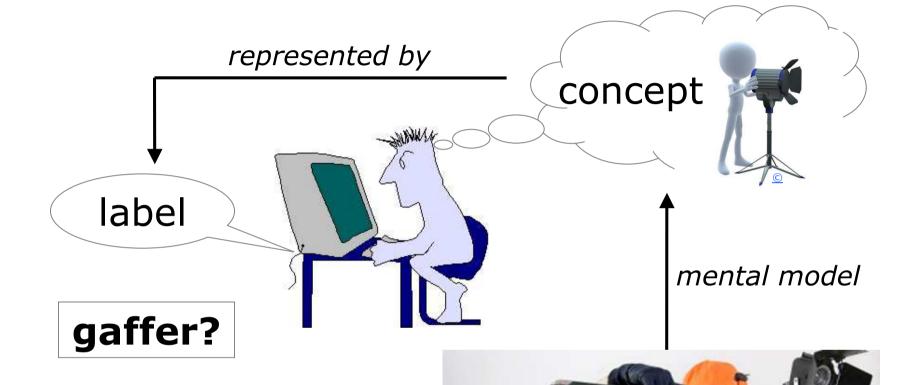
- lose information, that is to miss interesting documents
- retrieve irrelevant documents, they were not looking for

what's a cat ...



beauty is in the eye of the beholder

... and a gaffer?



Copyright

looking for gaffer in Google images





Seeking for images about *gaffer* in the sense of a film-making person, most of the retrieved hits are not relevant because of the homonymy of the search term. Results in German Google vary from that in the English language version because of the additional notion of *Gaffer* in German (meaning "rubbernecker").

more concepts

Gaffer (m) gazer gaper gawper gongoozler [sl.] [Br.] Gaffer (pl) gazers gawpers gawkers Gaffer (m) [ugs.] rubberneck [coll.] rubbernecker [coll.] Gaffer (pl) [ugs.] rubbernecks [coll.] rubberneckers [coll.] Gaffer {m} [Schaulustiger] onlooker [gazer]

Homonyms

lead to irrelevant information





```
gaffer
  Chef (m)
  Gevatter (m)
  Alterchen (n)
   film Oberbeleuchter (m) <OB>
gaffer [Br.] [coll.]
  Vorgesetzter (m)
   lobs Meister (m) [Betriebsmeister]
gaffer [coll.: old man]
  Opa (m) [ugs.: alter Mann]
gaffer [coll.: an old man]
  Väterchen (n)
gaffer [chiefly Br.] [coll.]
  Boss (m) [ugs.]
gaffer [foreman]
   lobs Vorarbeiter (m)
gaffer [senior glassblower]
   Glasbläser (m) [Meister]
electrician [member of gaffer's crew]
   Beleuchter (m)
```

Chief Lighting Technician

CLT

chief electrician

gaffa

Oberbeleuchter

OB

Synonyms

lead to information loss





CLT

Charlotte Douglas International Airport

Chemnitzer Linux-Tage

Central limit theorem

Cognitive load theory

Community Land Trust

OB

OB (star)

Oberbürgermeister

Oberbundesanwalt

Oberhausen

Offizierbewerber

Old Badmintonians

Organizational behaviour

Ortsbeauftragter

Outside broadcasting

hence, need for vocabulary control

What we need

- uniquely addressable concepts
- unambigous concepts
- terms or labels representing a concept
- semantic clarity for indexers and users
- permitting interoperability
- easily findable indexing and search terms

guidance for vocabulary management: ISO 25964

ISO 25964: Thesauri and interoperability with other vocabularies

- Part 1: Thesauri for information retrieval
- Part 2: Interoperability with other vocabularies
- updates ISO 2788/5964, with input from BS 8723
- information retrieval as the overall context
- Part 1: monolingual and multilingual thesauri
- Part 2: mapping between thesauri and other types of vocabulary

content of ISO 25964-1

Thesauri for information retrieval

- differentiation between terms and concepts
- guidance on establishing semantic relationships
- guidance on applying facet analysis to thesauri
- recommendations on how to manage compound terms
- guidance on thesaurus development and maintenance
- requirements for software to manage thesauri
- data model and XML schema for data exchange
- multilingual examples

content of ISO 25964-2

Interoperability with other vocabularies

- no normative statements about vocabularies other than thesauri
- however, comparisons are made and key features described
- emphasis is on interoperability, especially mapping
- structural models for mapping
- recommended mapping types
- how to handle pre-coordination
- practical aspects of mapping

- Classification schemes
- File plans
- Taxonomies
- Subject heading schemes
- Ontologies
- Term banks
- Name authority lists
- Synonym rings

ISO 25964-2 – mapping guidance

- Basic mapping types
 - Equivalence
 Laptop computers EQ Notebook computers
 - Hierarchical
 Roads NM Streets; Streets BM Roads
 - Associativee-Learning RM Distance education
- "Exact" or "Inexact" equivalence

Aubergines = EQ Egg-plants Horticulture ~ EQ Gardening

want a copy? want to get involved?

- Part 1 of the standard will be published in 2011
- Get it through your national standards body
- 17 countries already participate: Belgium, Bulgaria, Canada, China, Denmark, France, Germany, Finland, Korea, New Zealand, Russia, South Africa, Spain, Sweden, UK, Ukraine, USA
- Part 2 is still in draft ISO DIS 25964-2
 There is time for you to contribute ideas on interoperability!
- Contact your national standards body, specifically the committee corresponding to ISO TC 46/SC 9/WG8
- Send an email to get informed when the DIS of part 2 is released to <u>stella@lukehouse.org</u>

a word about interface and design considerations

Interface (preferred, en)
User interface (alternative, en)
Schnittstelle (preferred, de)
Nutzerschnittstelle (alternative, de)



What to consider when designing a vocabulary

- how the vocabulary may be used for searching
- how the vocabulary may be used for indexing
- how the vocabulary supports decisions to be made by software
- how well a vocabulary can be aligned with other vocabularies
- how vocabulary items suit metadata statements

methods of indexing or annotating

abstracting

Joyeux Noël (Merry Christmas) is a 2005 film about the World War I Christmas truce of December 1914, depicted through the eyes of French, British and German soldiers.

OFFICIAL SELECTION
CANNES FILM FESTIVAL

classifying

700 Arts & recreation

790 Sports, games & entertainment
791 Public performances
791.4 Motion pictures, radio, television
791.4302 Specific aspects of motion pictures

indexing

World War I

Christmas truce

war drama

Western Front

annotating

hasDirector

isBasedOn

isVersion

hasColour

hasGenre

tasks of retrieval

browsing vs. searching as information seeking strategies

- browsing is always selection-based, requiring that the search system makes sensible proposals
- when the browsing options are well-represented, browsing works better than keyword search
- browsing interfaces should seamlessly integrate keyword querying with navigation of the underlying information structure
- imperative search (keyword search) is prone to typing errors
- requires mastering of search syntax and operators
 - LexisNexis: HEADLINE:("Johnny Depp" w/5 "Chocolat")
 - DIALOG: (Johnny ADJ Depp AND Chocolat) ti
 - Google: "Chocolat" "Johnny Depp"

A keyword list

- is a controlled vocabulary
- may or may not have synonyms
- not structured, simple flat list
- limited number of items
- self-explanatory meaning
- hence no need for a structured vocabulary



- typically presented by a pick list
- simple sort order

keyword

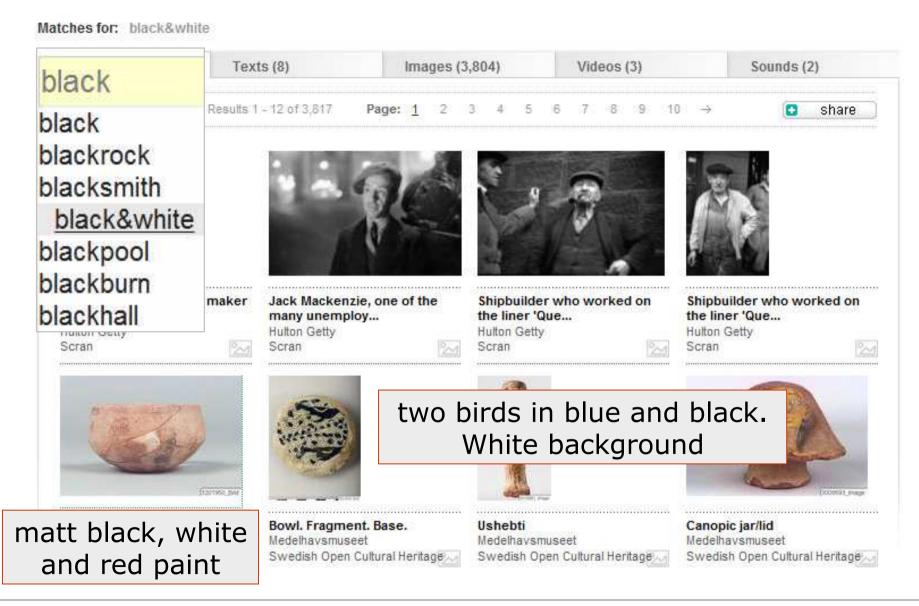
keyword list: colour



Schwarz & Weiss
Farbe
S/W & Farbe
koloriert, viragiert, handkoloriert
Farbe & S/W
k.A.

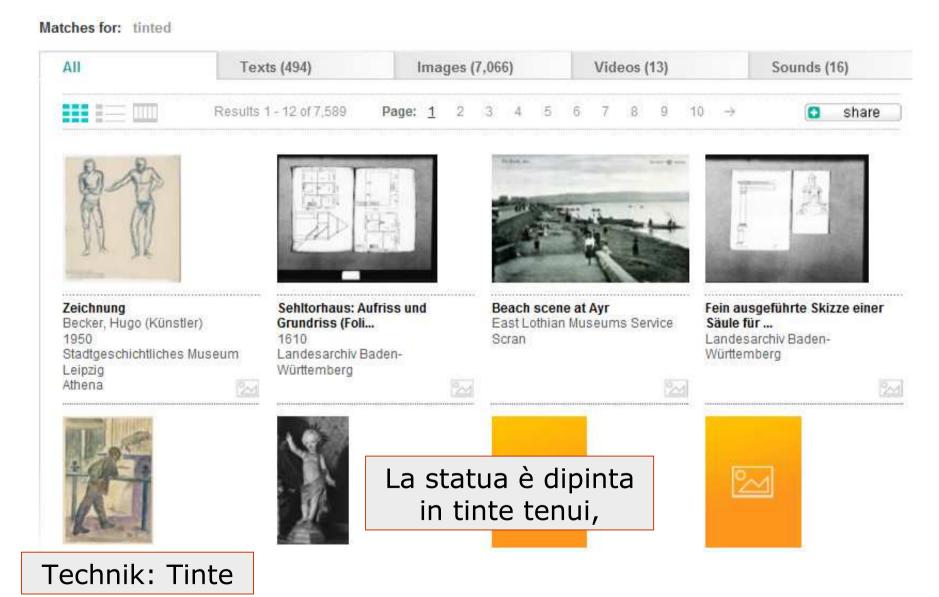


example: looking up black&white in Europeana



source: http://www.europeana.eu/portal/

example: looking up tinted in Europeana



source: http://www.europeana.eu/portal/

A taxonomy

- is a controlled vocabulary
- usually no synonyms
- hierarchically structured
- often with notations
- usually broad categories
- relatively low precision



- top down navigation
- search expansion
- not designed for post-coordination
- but possibility to enable facetted browsing

classification scheme: FIAF Glossary

A Basic Identifiers

A.1 Title

A.2 Date

A.3 Country

B Credits and Cast

B.1 Location of Credits

B.2 Producing

B.3 Directing

B.4 Writing

B.5 Cinematography

B.6 Production Design

B.7 Cast

B.8 Special Effects

B.9 Sound

. . .

C Distribution and Exhibition

C.1 Censorship and Rating

C.2 Copyright and Distribution

C.3 Exhibition and Prizes

D Form and Content

D.1 Form

D.2 Content

E Technical Properties

E.1 Sound Film

E.2 Silent Film

E.3 Language, Original Language

E.4 Subtitles, Subtitles by, Captions, Closed Captions

. . .

A thesaurus

- is a controlled vocabulary
- has synonyms
- hierarchically structured
- has related concepts
- higher degree of specificity
- relatively high precision

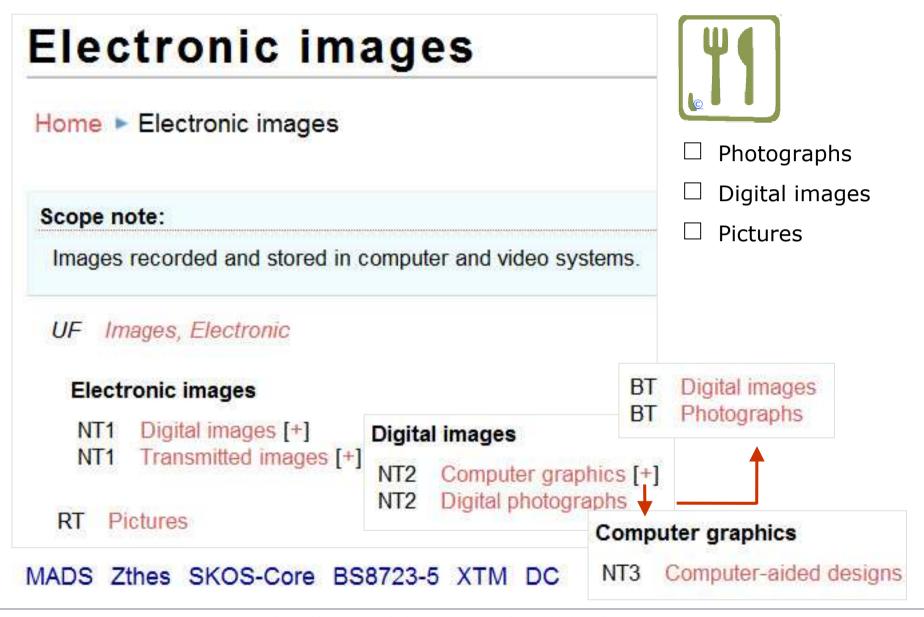


- search expansion
- alternative search terms
- support for refining a search
- suited for post-coordination
- identification of common spelling mistakes

ISO/FDIS 25964-1 (E)

© ISO 2010 — All rights reserved

thesaurus: Thesaurus for Graphic Materials (TGM)



source: http://www.vocabularyserver.com/tgm2/index.php?tema=372

An **ontology**

keyword list

- is a conceptual model of a domain of interest
- provides a vocabulary for concepts and properties
- hierarchically structured, including properties
- following logical rules
- annotation of instances



- Sparql querying
- supports semantic search
- search refinement
- inference based search expansion
- allows for queries across different collections of metadata

example: query for director in OpenCyc

OpenCyc Collection: film director

Unique ID: [Mx4rvVjqs5wpEbGdrcN5Y29ycA]
English ID: [Director-Movie]
English Aliases: ["film directors", "filmmaker", "filmmakers", "movie director",
"movie directors"]

An instance of PersonTypeByOccupation. Each instance of Director_Movie is a person responsible for the development of the production of a film.

A Type of: artist, leader

Instance of: type of profession

Subtypes: amateur film director

Instances: Brian De Palma, Martin Scorsese, Blake Edwards, Orson Welles, Roman Polanski, John Derek, Frank Capra, Michael Moore, Alfred Hitchcock, Rocco Siffredi, George Lucas, Quentin Tarantino, John Glenn, Sergio Leone

Same as:

http://umbel.org/umbel/sc/Director_Movie

Copyright @ 2001-2009 Cycorp, Inc.

ontology: Vocabulary Mapping Framework (VMF)

Scope of the mapped vocabularies is:

- Resource categories (eg CD, Ebook, Photograph)
- Resource-to-Resource relators (eg IsVersionOf, HasTranslation)
- Resource-to-Party relators (eg Author, EditedBy)
- Party-to-Party relators (eg AffiliatedTo)
- Party categories

Example of the Concept Family for "Adapt":

Table 4: Concept Family example

Concept Type	no	Term name	Definition
Verb or ContextType	1	Adapt / AdaptingEvent	To Derive an Adaptation.
Agent	1-n	Adapter	A Deriver of an Adaptation.
Patient	1-n	Adaptation	A Derivation made by changing an existing Creation.

source: The Vocabulary Mapping Framework (VMF): an introduction

faceted based navigation with Flamenco Search

PEOPLE

 agency (245)
 designer (270)

 architect (16268)
 developer (91)

 artist (1773)
 historical figure (225)

 author (289)
 instructor (725)

 culture (690)
 photographer (103)

PERIODS

<u>17 & 18th C</u> (1400) <u>Islamic-Hegira, 622 CE</u> (1391) <u>19th Century</u> (2261) <u>Modern</u> (5834)

LOCATIONS

20th Century (17827)

 Africa (1440)
 Middle East (1813)

 Antarctica (37)
 North America (11585)

 Asia (2864)
 South America (654)

 Australasia & Pacific Islands (186)
 Southeast Asia (465)

 Central America (70)
 Western Europe (11744)

 Eastern Europe (1165)

STRUCTURE TYPES

architectural elements (5068)

book elements (192)

buildings (by design) (863)

buildings (by function) (20168)

buildings (by height) (205)

buildings (by location or context) (560)

buildings (by massing or shape) (774)

cultural landscapes (435)

details (3239)

human settlements (by condition) (481)

human settlements (by economic base) (107)

human settlements (by form) (179)

more...

MATERIALS

 animal material (38)
 ceramic tile (207)

 artists' materials (31)
 chalk (8)

 asphalt (2)
 chemical compounds (16)

 brick (1380)
 clay (20)

 building materials (727)
 coating (482)

 cement (5)
 more...

STYLES

 African (599)
 Early Near Eastern (777)

 Ancient (74)
 European (13899)

 Asian (3755)
 Iron Age (28)

 Australian & Pacific Island (187)
 Islamic (583)

 Bronze Age (28)
 North American (14232)

 Central American (2)
 more...

 Early Mediterranean (2112)

VIEW TYPES

 axonometrics (119)
 design drawings (318)

 city aerial views (702)
 drawings (178)

 city details (356)
 elevations (306)

 city general views (1566)
 exterior details (2087)

 city maps and plans (934)
 exterior views (10619)

 construction views (170)
 more...

 decorative elements (42)

CONCEPTS

 access (1)
 in the arts (3792)

 barrier-free design (4)
 legal (24)

 circulation (2)
 philosophical (21)

 cultural (145)
 political (70)

 economic (17)
 psychological (118)

 environmental (471)
 more...

 housing (1188)

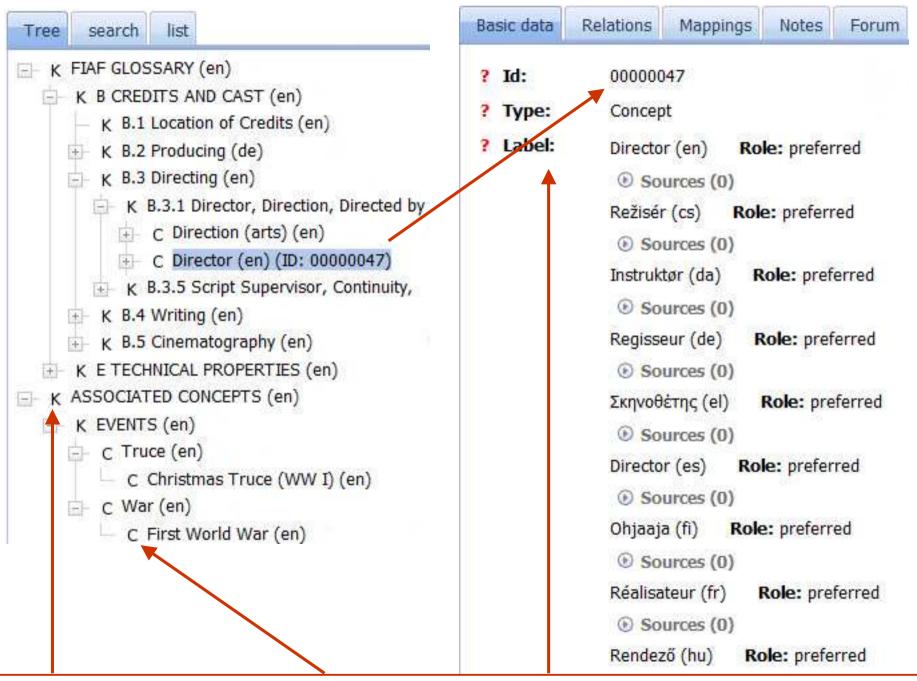
SOURCE

 book (13937)
 periodical (3953)

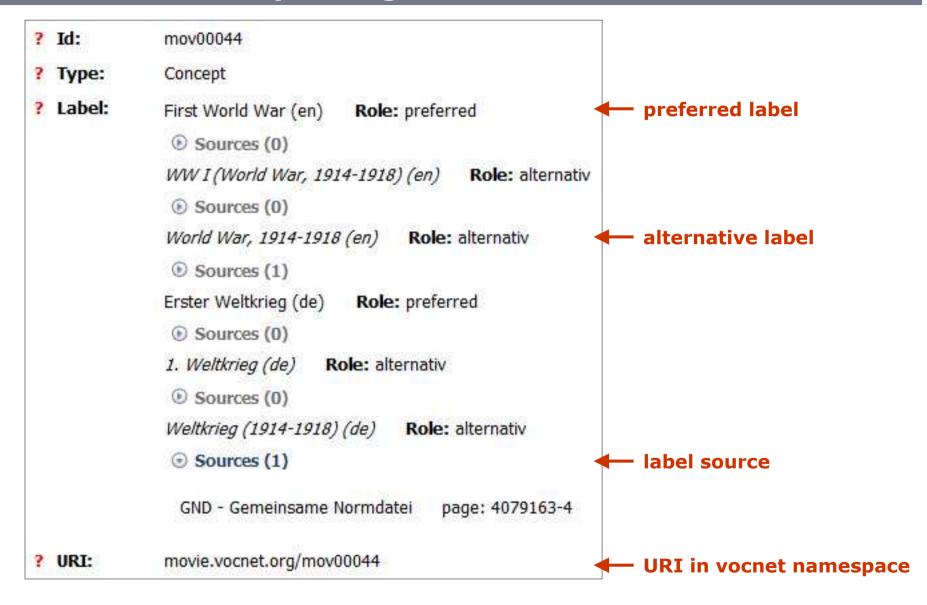
 donor (10956)
 vendor (7731)

source: http://orange.sims.berkeley.edu/cgi-bin/flamenco.cgi/spiro/Flamenco; s.a. http://www.museosuomi.fi/

- web-based
- suitable for distributed work
- concept-based unique URI
- preferred and non-preferred labels, customisable for display
- supports all standard relationships of KOS
- consistency checks (doublet control, refuse of circles, etc.)
- webservice based on vocnet exchange format
- exchange format compatible with SKOS, BS8723-5 model
- future: term management, systematic display
- contact: L. Landwehr, A. Vitzthum digiCULT-Verbund eG http://www.digicult-verbund.de/index.php?p=Kontakt



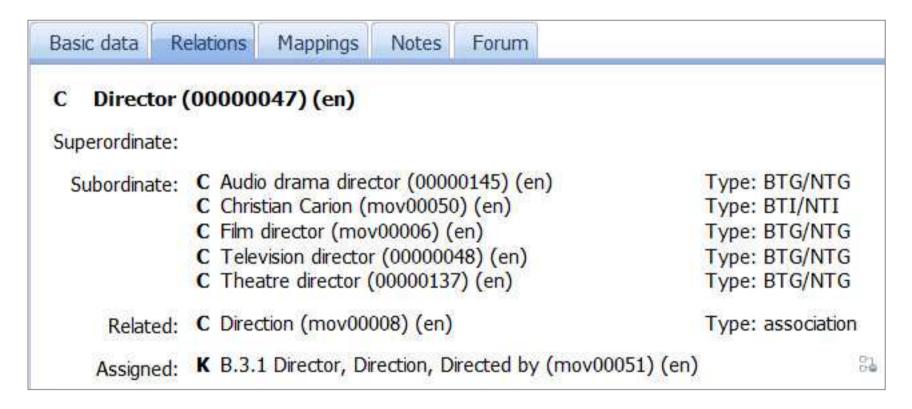
Categories for systematic display; Concepts for indexing; Labels representing a concept



Preferred and alternative labels for concept "First World War"

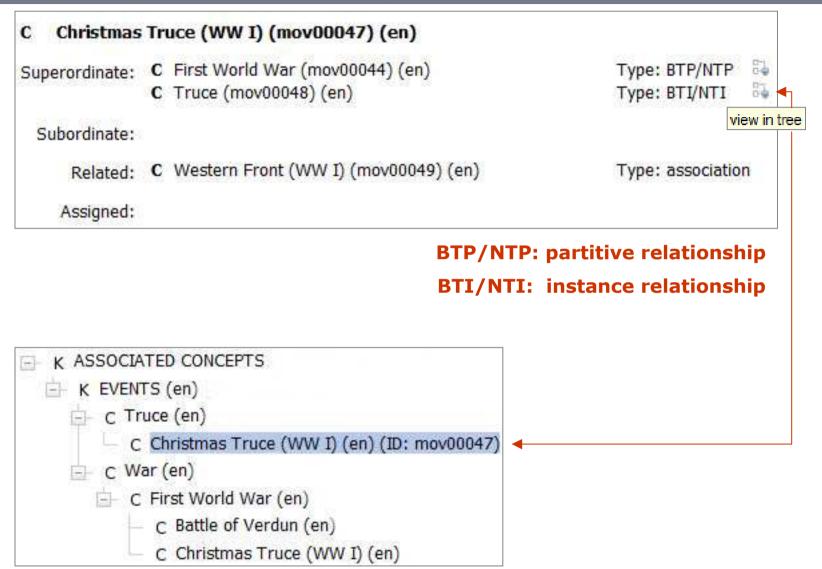
Basic data Relatio	ns Mappings Notes Forum	
?Identifier:	http://www.getty.edu/vow/AAT	http://www.imdb.com/glossary/D#director
?Type of identifier:	URI	URI
?Type of mapping:	exactMatch	exactMatch
?source:	AAT	IMDb
?Label:	directors (performing arts)	Director
?Identifier:	http://d-nb.info/gnd/4049050-6	http://stitch.cs.vu.nl/vocabularies/rameau/
?Type of identifier:	URI	URI
?Type of mapping:	exactMatch	narrowMatch
?source:	GND	Rameau
?Label:	Regisseur	Réalisateurs de cinéma
?Identifier:	http://id.loc.gov/authorities/sh85088067	Mx4rvVjqs5wpEbGdrcN5Y29ycA
?Type of identifier:	URI	Local
?Type of mapping:	broadMatch	narrowMatch
?source:	LSCH	OpenCyc
?Label:	Motion picture producers and directors	Director-Movie

Concept "Director" mapped to a concept or class in different schemes

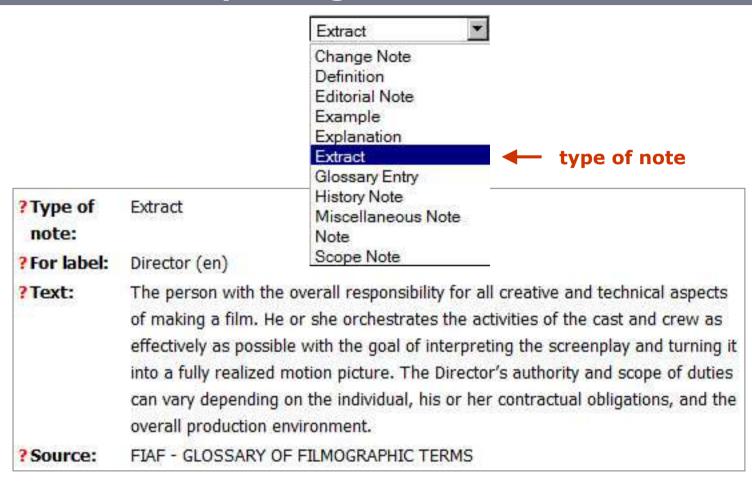


BTG/NTG: generic relationship

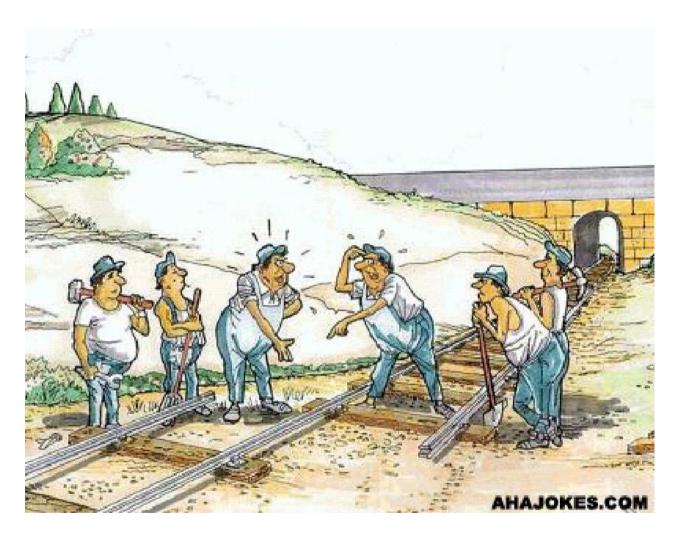
BTI/NTI: instance relationship



Concept "Christmas Truce" and its hierarchical and associative relationships



Good luck!



- Look for authorities, rather than start from scratch
- Make things as simple as possible, but not simpler

Thank you!



resources

xTree - vocabulary management tool

Contact: digiCULT-Verbund eG: Lütger Landwehr, Axel Vitzthum, Björn Schill http://www.digicult-verbund.de/index.php?p=Kontakt

or Jutta Lindenthal

jutta.lindenthal@googlemail.de

Stella Dextre Clarke

ISO 25964 - the new standard for thesauri and interoperability with other vocabularies http://eurovoc.europa.eu/drupal/sites/all/files/conference2010/EuroVocConference ISO25964preview.ppt

Draft Format for exchange of thesaurus data conforming to ISO 25964-1 http://www.niso.org/schemas/iso25964/

Gordon Dunsire, Centre for Digital Library Research - University of Strathclyde, UK

The Vocabulary Mapping Framework and its potential for improving metadata interinteroperability in the Semantic Web

http://eurovoc.europa.eu/drupal/sites/all/files/conference2010/Dunsire%20Eurovoc2010.ppt

The Vocabulary Mapping Framework (VMF): an introduction, v1.0, December 12, 2009 http://cdlr.strath.ac.uk/VMF/documents/VocabularyMappingFrameworkIntroductionV1.0%28091212%29.pdf

resources

FIAF - International Index to Film Periodicals

http://www.fiafnet.org/uk/publications/iifp subjectHeadings.cfm

FIAF - GLOSSARY OF FILMOGRAPHIC TERMS

http://www.fiafnet.org/publications/Glossary%20of%20Filmographic%20Terms%20%28English%20Version%292008%20revision.pdf

IMDb Movie Terminology Glossary

http://www.imdb.com/glossary/

IMDb Movie Keyword Analyzer (MoKA)

http://www.imdb.com/Sections/Keywords/

Taxonomy Warehouse

http://www.taxonomywarehouse.com/index.asp

TaxoBank - List of Vocabularies

http://www.taxobank.org/terminologies

MARS Authority Control

 $\frac{http://ac.bslw.com/community/blog/2011/01/were-expanding-rbms-and-tgm-databases-available-for-searching/$

resources

National Monuments Record Thesauri

http://thesaurus.english-heritage.org.uk/frequentuser.htm

MDA Terminology Initiatives

http://www.mda.org.uk/spectrum-terminology/termwork

Vocabularies and Thesauri

http://www.vrafoundation.org/ccoweb/cco/vocab.html

Guidelines for Constructing a Museum Object Name Thesaurus

http://www.mda.org.uk/spectrum-terminology/holm

Thesaurus for Graphic Materials I & 2

 $\underline{http://www.vocabularyserver.com/tgm1/}$

http://www.vocabularyserver.com/tgm2/

The Moving Image Genre-form Guide

http://www.loc.gov/rr/mopic/migintro.html

Further links (German)

http://www.jlindenthal.de/mb/index.html