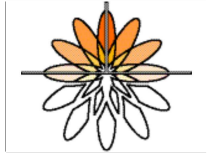


Multimedial N E-Culture behind the scenes

The project team
VU, CWI, UvA, DEN, ICN





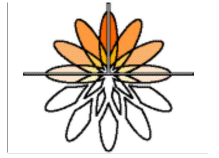
Hypothesis

- Semantic Web technology is in particular useful in knowledge-rich domains

or formulated differently

- If we cannot show added value in knowledge-rich domains, then it may have no value at all

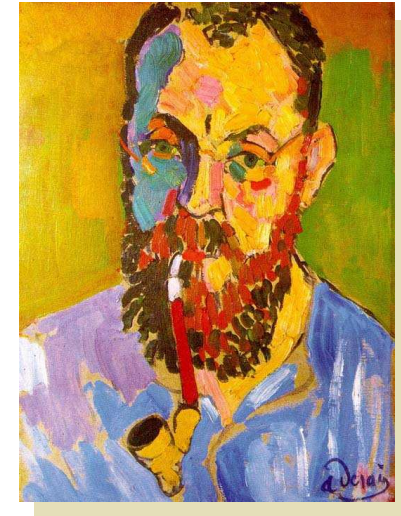
The Web: resources and links



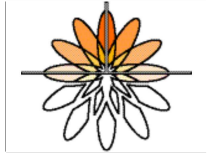
URL



Web link



URL



The Semantic Web: typed resources and links

Painting
"Woman with hat
SFMOMA

Dublin Core
creator

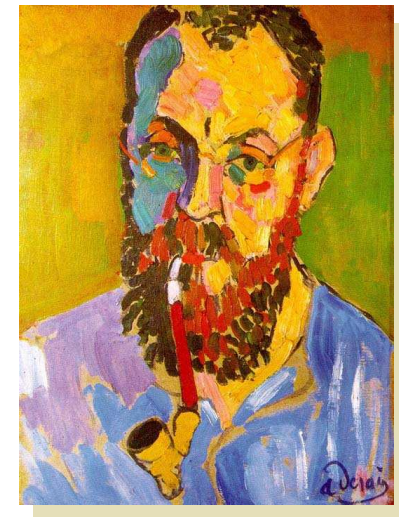
ULAN
Henri Matisse



URL



Web link



URL



Union List of Artist Names® Online

Full Record Display

[New Search](#) [Previous Page](#)

Click the icon to view the full record

ID: 500017300

Matisse, Henri (French)

Related People or Corporate Bodies:

- apprentice was [Jolin, Einar](#) 1911-1913
..... (Swedish painter, 1890-1990) [500014093]
- parent of [Duthuit, Marguerite Matisse](#)
..... (French painter, born ca. 1900) [500075813]
- patron was [Barnes, Dr. Albert C.](#)
..... (American collector, 1872-1951) [500057478]
- student of [Cormon, Fernand](#)
..... (French painter and teacher, 1845-1924) [500115385]
- student of [Moreau, Gustave](#)
..... (French painter, 1826-1898) [500115776]
- teacher of [Boussard, Calixte-Henri](#)

Names:

- Matisse, Henri** artist (**preferred**)
- Henri Matisse** painter
- Matisse, Henri** printmaker
- Matisse, Henri** sculptor
- Matisse, Henri** designer

Nationality:

French

Roles:

- artist
- painter
- printmaker
- sculptor
- designer
- writer

Roles:

- artist (**preferred**)
- painter
- printmaker
- sculptor
- designer
- writer

Gender: male

Birth and Death Places:

Born: [Le Cateau-Cambrésis \(Nord, Nord-Pas-de-Calais, France\)](#) (inhabited place)

Died: [Nice \(Alpes-Maritimes, Provence-Alpes-Côte d'Azur, France\)](#) (inhabited place)



Getty Thesaurus of Geographic Names® Online

Hierarchy Display

[New Search](#)


[Previous Page](#)

[Help](#)

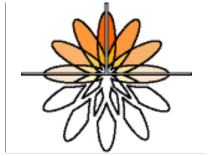
[Vernacular Display](#) | **English Display**

[View Selected Records](#)

[Clear All](#)

Click the  icon to view the hierarchy.
Check the boxes to view multiple records at once.

-  [Top of the TGN hierarchy](#) (hierarchy root)
-  [World](#) (facet)
-  [Europe](#) (continent)
-  [Netherlands](#) (nation)
-  [\[view physical features \]](#)
- [Aarkanal](#) (canal)
- [Afsluitdijk](#) (dam)
- [Alblasserwaard](#) (general region)
- [Altena, Land van](#) (general region)
- [Amstelland](#) (general region)
- [Amsterdam Rijn Kanaal](#) (canal)
-  [Aruba](#) (dependent state) [N]
- [Bernisse Molen](#) (mill center)
- [Biesbos](#) (general region)
- [Brouwersdam](#) (dam)
- [Calandkanaal](#) (canal)
- [Delfland](#) (general region)
- [Delfland Oost](#) (general region)



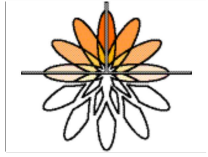
Principle 1: semantic annotation

- Description of web objects with “concepts” from a shared vocabulary



Description:

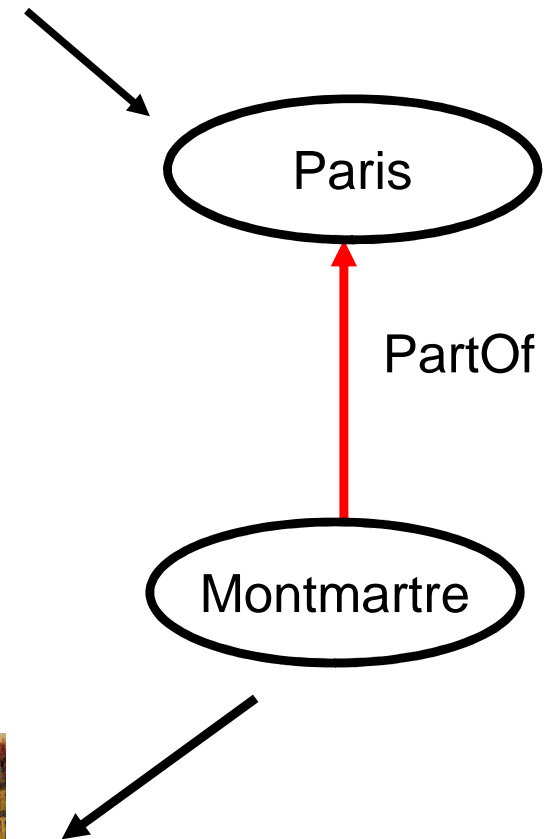
recordnumber	23727;
timestamp	2001-06-22;
type	cultural;original;
collector	Johannes Frederik van Ov Zeldzaamheden; Utagawa
series	360;
Culture	Japans;
Date	1800-1829;
Description	1883 JAPAN aankoop;
Identifier in Current Repository	360-4564;
Creation Site	Japan;
Current Repository	RMV;
Material	papier;
Measurements.Format	oban, 25.5 cm x 37.5 cm;
Style/Period.Period	Edo;
Title	Edo junisho;
Type	prenten;
type	Work;



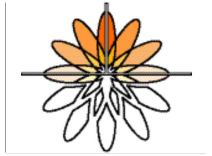
Principle 2: semantic search

- ❑ Search for objects which are linked via concepts (semantic link)
- ❑ Use the type of semantic link to provide meaningful presentation of the search results

Query
"Paris"

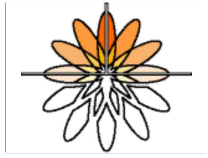


Principle 3: multiple vocabularies or: the myth of a unified vocabulary



- ❑ In large virtual collections there are always multiple vocabularies
 - In multiple languages
- ❑ Every vocabulary has its own perspective
 - You can't just merge them
- ❑ But you can use vocabularies jointly by defining a limited set of links
 - “Vocabulary alignment”
- ❑ It is surprising what you can do with just a few links

Example



“Tokugawa”



AAT style/period
Edo (Japanese period)
Tokugawa



SVCN period
Edo

*SVCN is local in-house
ethnology thesaurus*

MultimediaN E-Culture Demonstrator

Cultural Heritage Data

Generic Multilingual Interfaces

Collections

Artchive.com	36K
Rijksmuseum	30K
Ethnographical	1.1M



Vocabularies

SVCN (ethnographic)	65K
W3C Wordnet	2.2M
Getty: AAT	450K
TGN	500K
ULAN	1.5M



Schemas

- Dublin Core
- VRA Core
- E-Culture Specific

web standards

HTML	OWL
SVG	RDF
AJAX	RDFa



SWI-Prolog

Semantic Web

Server (SPARQL support)

Semantic Grouping

Works created by an artist with matching AAT style (214)



Works created by a teacher of an artist with matching AAT style (2)

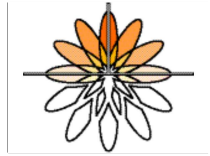


Semantic Timeline



/facet browser

Triples 6,663,465 *and counting...*



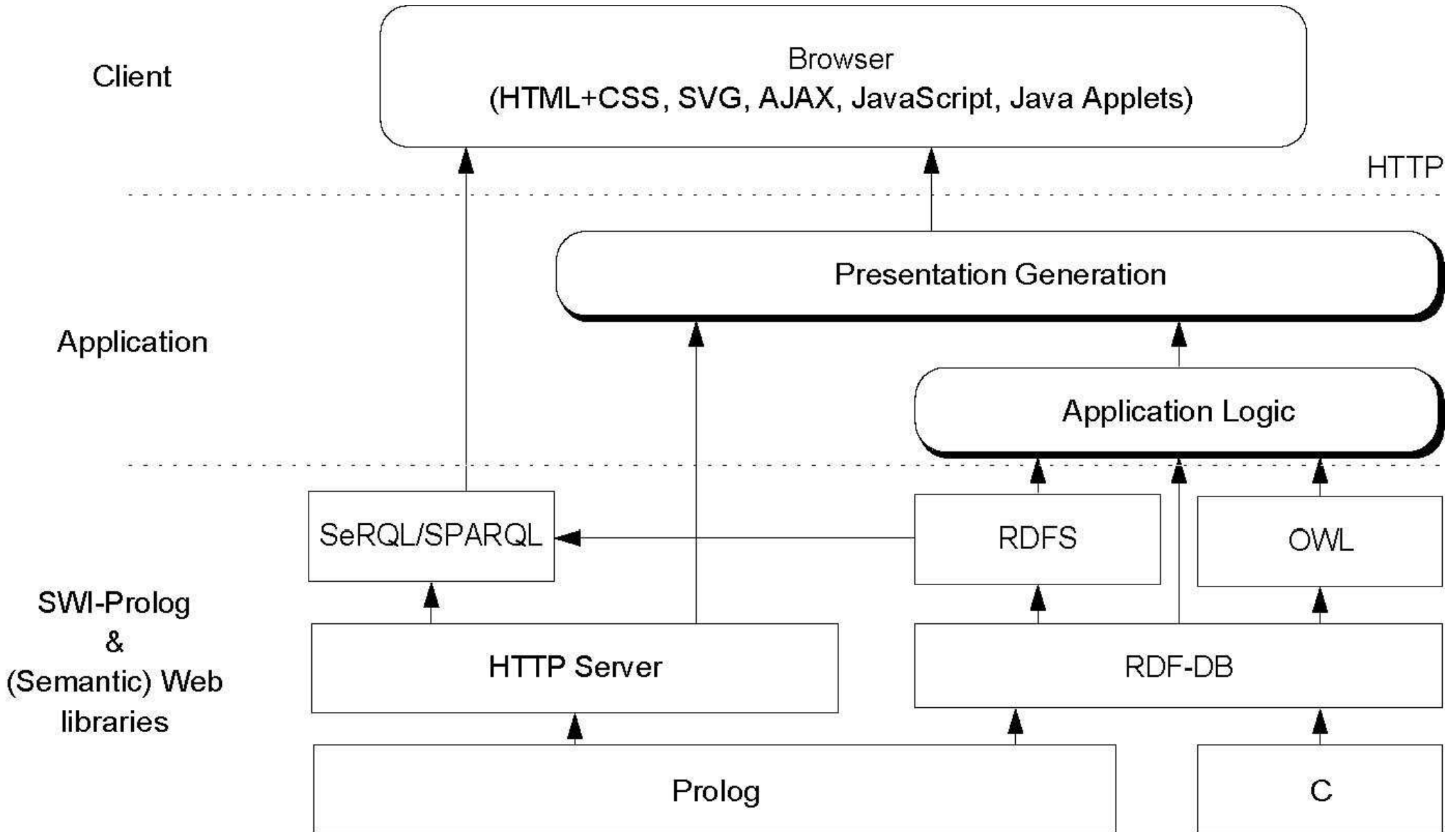
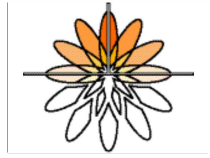
Continuous activity: user studies

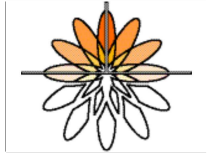
□ First study: 19 users

Use cases (type of user)	Tools	Issue found
Group A. Object handling		
1. Research for object acquisition or release (CU)	RKD, Google, lexicon exhibition catalog, museum database	a. Deal with many systems b. Deal with heterogeneous data
2. Research before object restoration (CU)	old archive, old auction catalog RKD library, museum database	
3. Assessing the value of an object (CU)	RKD, Google, Lexicon	m. Navigational search
Group B. Planning exhibition or publication		
4. Planning (RE)		
5. Research (RE, CU)	catalogs art and history literature museum database artist, gallery website RKD online, RKD library Google, lexicon Internet, Getty database Picarta, municipal archive digital newspaper archive historical library	a. Deal with many systems f. Exploratory search m. Navigational search j. Image search i. Cultural influence and similarities n. Seek relationship e. Many language, alphabet, and spelling b. Deal with heterogeneous data n. Queries with unexpected results k. query by example

Architecture

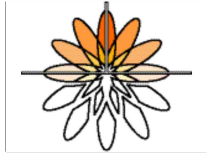
Scalability
is key!





Distributed vs. centralized collection data

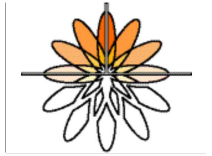
- ❑ Minimal requirement: collection object has image URI
- ❑ Preference for external metadata, accessed through protocol such as OAI
- ❑ In practice, external metadata access is still cumbersome



Search strategies

- Basic search: keyword-oriented
- Advanced search:
 - Tweaking default search parameters
 - Time-related queries
- Faceted search
- Relation search
 - How are two URIs related?

Term disambiguation is key issue in semantic search

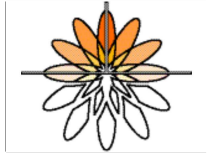


□ Post-query

- Sort search results based on different meanings of the search term
- Mimics Google-type search

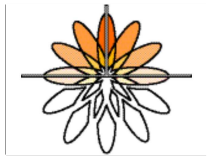
□ Pre-query

- Ask user to disambiguate by displaying list of possible meanings
- Interface is more complex, but more search functionality can be offered



Keyword search with semantic clustering

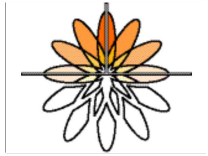
1. Btree of literals plus Porter stem and metaphone index
2. Find resources with matching labels
 - Default resources are “Work”s
3. Find related resources by one-way graph traversal
 - **owl:inverseOf** is used
 - Threshold used for constraining search
4. Cluster results (group instances)



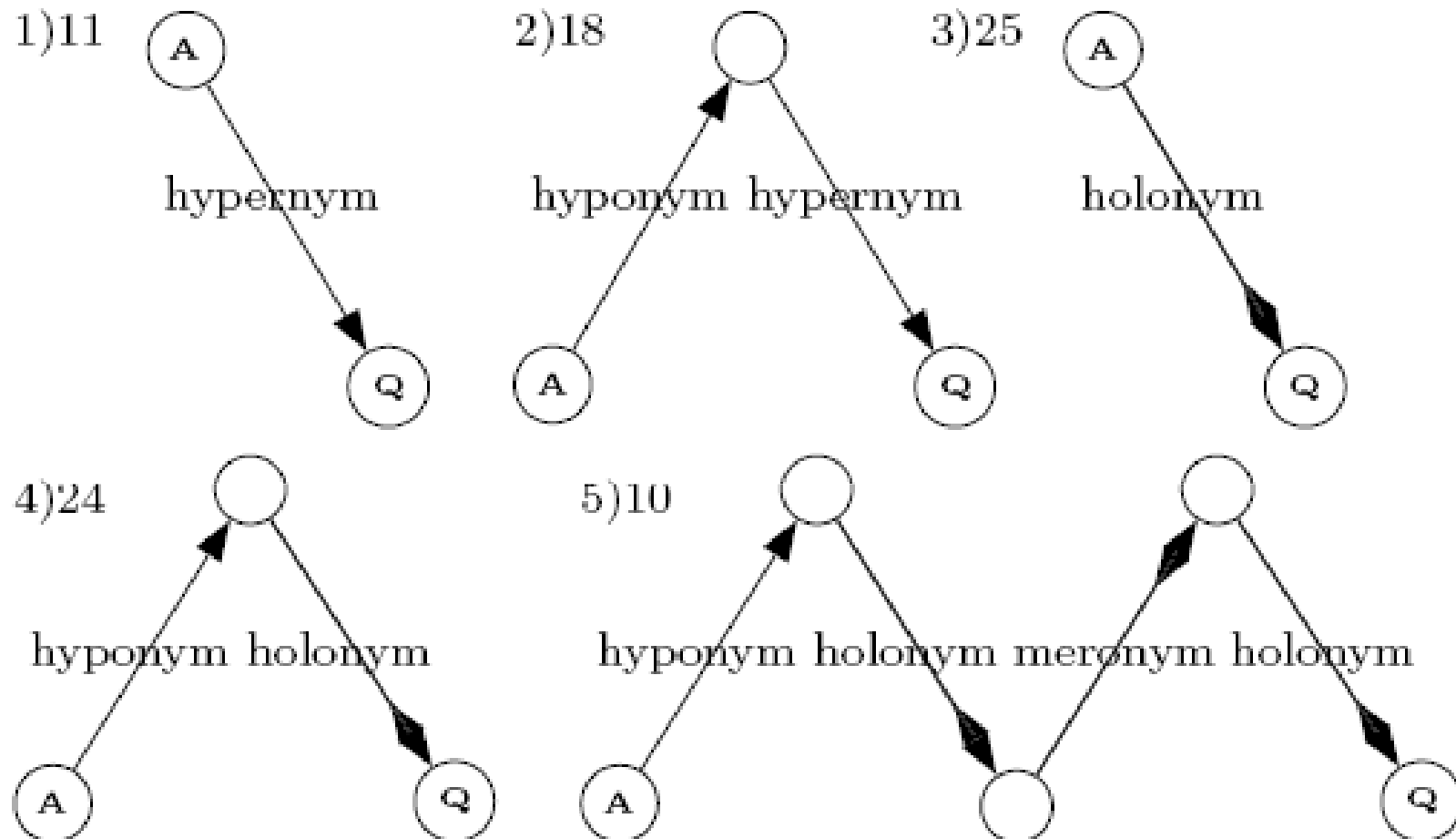
Faceted search

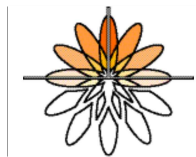
- Use Dublin Core scheme to formulate complex queries
- Navigate through relevant metadata

Main Category	Style/Period	Subject	Culture
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
AAT Concept	All 4	All 30	All 3
Place	Post-Impressionist 1	Paris 8	French 2
Person	Neo-Impressionist 1	cityscapes 7	Jewish 1
SVCN Concept	Cubist 1	Montmartre 3	
Image	Fauve 1	afternoon 2	
Work >1000		day 2	
		self-portraits 2	
		night 1	
	▼ facet options	▼ facet options	▼ facet options



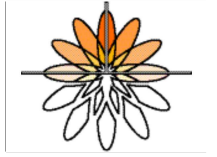
Search: WordNet patterns that increase recall without sacrificing precisions





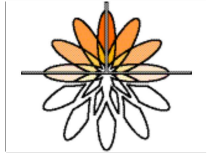
Triple statistics

Document	# Sources	# Triples
Schemas		
RDFS/OWL	2	358
Annotation	6	769
Vocabularies	9	1,225
Collections	1	29,889
Vocabularies		
TGN	4	425,517
ULAN	16	1,896,936
AAT	1	249,162
WordNet	18	2,579,206
Collections		
Artchive	4	74,414
Rijkmuseum	1	27,933
RVM	1	3,662,257



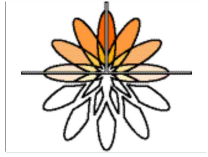
Knowledge engineering principles

- ❑ Be modest!
- ❑ Think large!
- ❑ Don't strive for perfection!
- ❑ Use open Web standards!

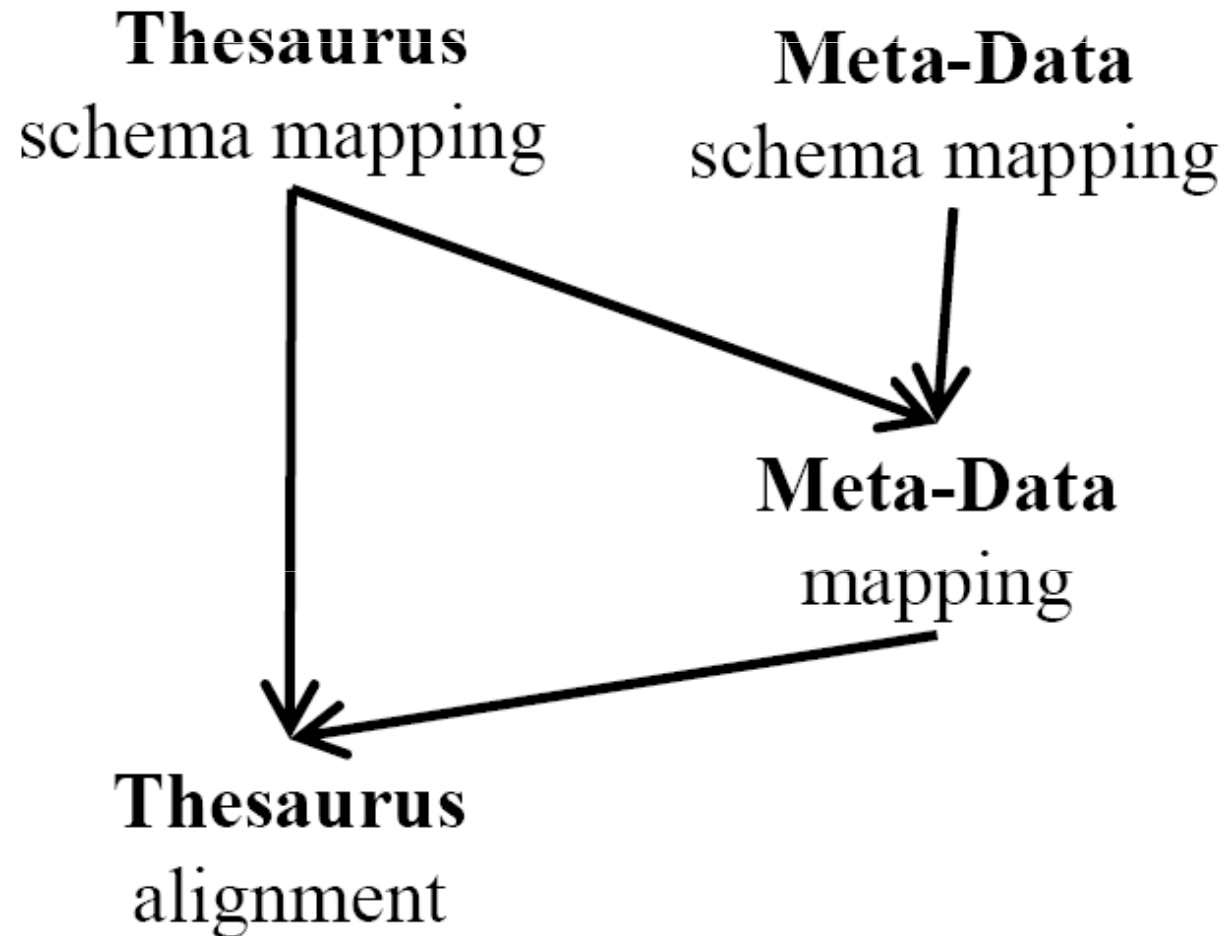


Levels of interoperability

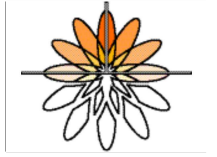
- Syntactic interoperability
 - using data formats that you can share
 - XML family is the preferred option
- Semantic interoperability
 - How to share meaning / concepts
 - Technology for finding and representing semantic links



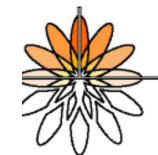
From metadata to semantic metadata



Activity 1: syntactic vocabulary interoperability



- ❑ Making vocabularies available in the Web standard RDF
- ❑ Many organizations already do this
- ❑ W3C provides the SKOS template to make this almost straightforward
- ❑ Effort required: at most a few days



RDF/OWL Representation of WordNet

W3C Working Draft 19 June 2006

This version:

<http://www.w3.org/TR/2006/WD-wordnet-rdf-20060619/>

Latest version:

<http://www.w3.org/TR/wordnet-rdf/>

Previous version:

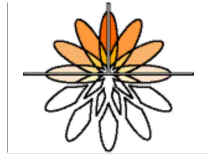
This is the first published version

Editors:

[Mark van Assem](#), Vrije Universiteit Amsterdam

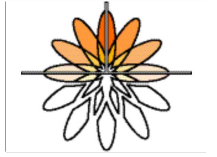
[Aldo Gangemi](#), ISTC-CNR, Rome

Activity 2: metadata schema mapping



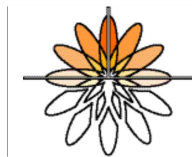
Data Item	Function	Activity	Source and Target Property/Class
NUMMER	Record Id	Create URI and additional project specific triples (&vra;Work)	<i>source:</i> 6 <i>target:</i> bp:6 rdf:Type vra:Work .
TITEL	Title in Dutch	Create literal and language tag	<i>source:</i> Delftse Bijbel... <i>target:</i> vra:title "Delftse Bijbel..."@nl ;
TITEL_EN	Title in English	Create literal and language tag	<i>source:</i> Delft Bible... <i>target:</i> vra:title "Delft Bible..."@en ;
MAKER	Creator and his marker for role	Extract name and role marker, create URI and label for name and convert marker to role, create role as subproperty of vra:creator	<i>source:</i> Yemantszoon, Mauricius : d (d stands for drukker meaning printer) <i>target:</i> bp:drukker bp:Yemantszoon_Mauricius ; bp:Yemantszoon_Mauricius rdf:type ulan:person ; rdfs:label "Yemantszoon Mauricius" .
OBJECT	Object type	Map to AAT or create local extension to AAT and mapping	<i>source:</i> tekstbladzijde (text page) <i>target:</i> vra:type bp:object_tekstbladzijde ; bp:tekstbladzijde rdf:type skos:concept . skos:prefLabel "tekstbladzijde"@nl ; skos:broader AAT:pages ;

Activity 3: semantically enriching the metadata

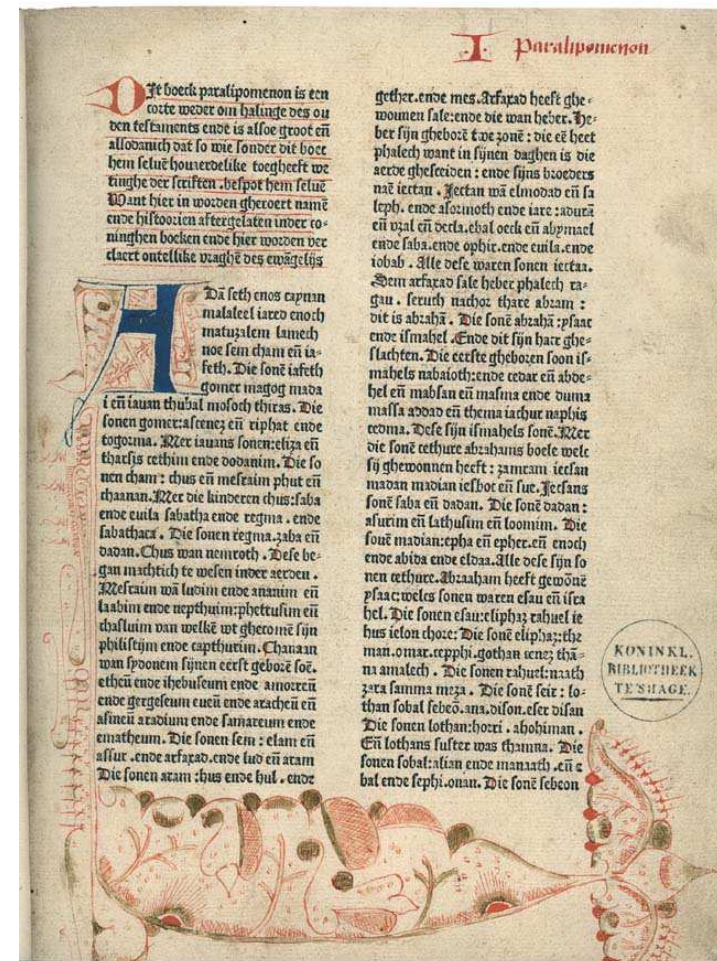


- ❑ Extracting additional concepts from an annotation
 - Matching the string “Paris” to a vocabulary term
- ❑ Information-extraction techniques exists (and continue to be developed)
- ❑ Effort required can be up to a few weeks
 - The more concepts, the better, but no need to be perfect!

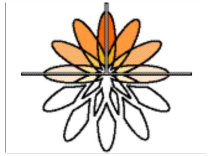
Example textual annotation



```
<inm:Record>
  <inm:NUMMER>6</inm:NUMMER>
  <inm:TITEL>Delftse Bijbel...</inm:TITEL>
  <inm:TITEL_EN>Delft Bible...</inm:TITEL_EN>
  <inm:MAKER>Yemantszoon, Mauricius : d</inm:MAKER>
  <inm:OBJECT>tekstbladzijde</inm:OBJECT>
  <inm:TECHNIEK>boekdruk</inm:TECHNIEK>
  <inm:DATERING>10 jan. 1477</inm:DATERING>
  <inm:CLASSIFICATIE>D</inm:CLASSIFICATIE>
  <inm:ORIGINEEL>Bijbel. Oude
    Testament...</inm:ORIGINEEL>
</inm:REPRODUCTIE>
<inm:TWNAAM/>
<inm:TWOND>typografische vormgeving</inm:TWOND>
<inm:TWOND>bijbels</inm:TWOND>
<inm:TWGEO>Delft</inm:TWGEO>
<inm:OMSCHRIJVING>Eerste bijbel die in het
  Nederlands verscheen...</inm:OMSCHRIJVING>
<inm:OMSCHRIJVING_EN>The first Bible to
  appear in the Dutch language...</inm:OMSCHRIJVING_EN>
<inm:AFMETINGEN>27 x 20 cm</inm:AFMETINGEN>
  ...
</inm:Record>
```



Resulting semantic annotation (rendered as HTML with RDFa)



Description:

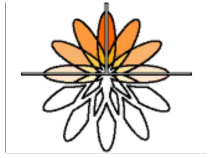
classificatie	Geschiedenis van de boekdrukkunst;
drukker	Meer, Jacob Jacobszoon van der; Yemantszoon, Maucius;
origineel	Bijbel. Oude Testament. - Delft: Jacob Jacobszoon van der Meer en Maucius Yemantszoon, 10 jan. 1477, dl. 2, p. 1;
Date	10 jan. 1477;
Description	The first Bible to appear in the Dutch language, known as the Delft Bible. It consists of the Old Testament only and is an anonymous adaptation of the - again anonymous - History Bible of 1360. It is an example of an incunabulum where the hand-written book still served as an example for lay-out and design. Contrary to many other incunabula, the place of origin, the names of the printers and even the day of its completion are mentioned in the colophon.;
Measurements.Dimensions	27 x 20 cm;
rights.copyright	Den Haag Koninklijke Bibliotheek;
Source	Bibliopolis;
Subject	bibles; incunabula; initials; ornamental borders; rubrications; typographical design;
subject.geographicPlace	Delft;
Technique	letterpress printing;
Title	Delft Bible, printed in Delft by Jacob Jacobszoon van der Meer and Maucius Yemantszoon, 1477;
Type	tekstbladzijde;
type	Work;

Used as value to describe other resources:

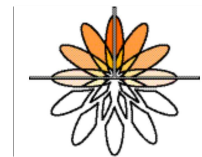
BBB_169E56_1477_P1.JPG;

relation.depicts

Activity 4: vocabulary alignment



- ❑ Find semantic links between vocabulary links
 - Derain (ULAN) related-to Fauve (AAT))
- ❑ Automatic techniques exist, but performance varies
- ❑ Often combination of automatic and manual alignment
- ❑ Effort strongly dependent on vocabularies
 - But “a little semantics goes a long way” (Hendler)

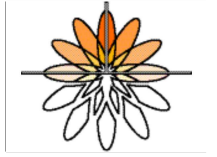


Learning alignments

- Learning relations between art styles in AAT and artists in ULAN through NLP of art historic texts

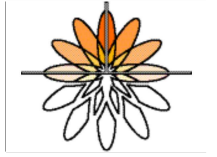


<i>Artist Name</i>	<i>IS</i>	<i>In GS</i>
edgar degas	0.0699	1
edouard manet	0.0548	1
pierre-auguste renoir	0.0539	1
morisot, berthe	0.0393	1
gogh, vincent van	0.0337	0
cassatt, mary	0.0318	1
cezanne, paul	0.0302	1



Planned: annotation interfaces

- Brings in issues w.r.t. trust and access control
 - Quality issue
- Involving users in the annotation process
 - Stimulating people to contribute
- User profiling

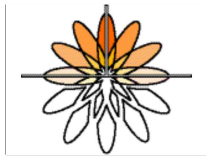


Planned: large-scale deployment

- ❑ Moving to 100-150M triples, 30+ collections
- ❑ Scalability is key
 - 64-bit machines
 - Search algorithms
- ❑ Methodology for including new collections

On-line demo

<http://e-culture.multimedian.nl>



Home [en](#) [nl](#)

News Multimedian N9C Eculture project homepage

Posters & Publications

Online demo

FAQ

People

Credit

Internal

Contact us

XHTML VALID

Despite the amounts of public funding devoted to both cultural heritage and ICT infrastructure, online access to even the most important aspects of our past is still limited and highly fragmented. The objective of this project is the development of a set of e-culture demonstrators providing multimedia access to distributed collections of cultural heritage objects. The demonstrators are intended to show various levels of syntactic and semantic interoperability between collections and various types of personalized and context-dependent presentation generation. The demonstrators will all be developed as components of the portal cultuurwijzer.nl of DEN (Foundation Cultural Heritage Netherlands). This portal will serve as a joint application domain for the demonstrators. The portal provides access to a relatively large set of key culture-heritage collections in The Netherlands. The demonstrators will typically focus on subsets of the collections to demonstrate the use of semantic interoperability, semantic information access and visualization, and context-specific presentation generation. The project integrates results from computer science in the areas of: semantic Web technology, multimedia indexing and search, and web interfacing and data visualization to facilitate display of (part of) our cultural heritage. The team combines a balanced mix of internationally renowned academic expertise on Semantic Web and multimedia technology with industrial strength experience in Web design and user interface development.

