Semantic Web Opportunities for Digital Libraries

ELAG 2008

Laura Hollink, Antoine Isaac, Véronique Malaisé, Guus Schreiber

Vrije Universiteit Amsterdam Computer Science Department Why would you want to use Semantic Web techniques?

Two reasons:

- 1. Machine-processable representation of semantic information
 - Your search engine can reason with the knowledge stored in your thesaurus



Use case: "Find paintings of a similar style."





KLIMT, Gustav Portrait of Adele Bloch-Bauer I 1907 Oil and gold on canvas 138 x 138 cm Austrian Gallery, Vienna

MUNCH, Edvard The Scream 1893 Oil, tempera and pastel on cardboard 91 x 73.5 cm National Gallery, Oslo

Why would you want to use Semantic Web techniques?

2. Interoperability

- Link your collection to other collections
- Let other people/institutes link to your collection
- Enable cross-collection access
- Link your KOS to other KOSs
- Use the knowledge encoded in other KOSs

Overview

Semantic Web: for libraries

- Interoperability
- Semantic Search in the E-Culture Web Demonstrator



The Web: resources and links



URL

Web link

URL

The Semantic Web: typed resources and typed links

Painting "The girl with green eyes" MOMA San Francisco

creator

Henri Matisse



URL

Web link

URL

SW Building Blocks I

Resources:

- Everything that can be identified on or outside the web.
 - http://www.artchive.com/artchive/m/matisse/grn_eyes.jpg
 - http://purl.org/dc/elements/1.1/creator
 - http://e-culture.multimedian.nl/ns/getty/ulan#500017300 (matisse)
 - http://dbpedia.org/resource/Beatrix_of_the_Netherlands

Triples of three resources

- Subject predicate object
 - grn_eyes.jpg dc:creator ulan:matisse
 - MyDocument1 date "2008-01-14"
 - Beatrix monarch of The Netherlands
 - The Netherlands part of Benelux

SW Building Blocks II

Graphs of triples



SW Building Blocks III

Standard languages to represent these graphs:

- RDF
- RDF-Schema
- OWL

SW for Libraries and Archives

SW research targeted towards cultural heritage:

- Standard language for library and information sciences: SKOS
- Research projects for access to cultural heritage:
 - CATCH program
 - MultimediaN E-Culture
- Library use-cases in evaluation studies (OAEI)

Interoperability Problem



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

Interoperability: using standards

SKOS

- broaderTerm
- narrowerTerm
- relatedTerm
- exactMatch, narrowMatch, broadMatch, relatedMatch

VRA Core 4.0 Elements: e.g.

- **title**
- Iocation
- date
 - earliestDate (circa)
 - IatestDate (circa)
- material, etc.

Interoperability: reusing vocabularies

4	~	🕲 WordNet Search - 3.0 - Mozilla Firefox
	🕲 т	File Edit Yiew History Bookmarks Tools Help
Ť	<u>File</u>	< - 📄 - 🥑 💿 🏠 💓 http://wordnet.princeton.edu/perl/webwn?o2=&o0=1&o7=&o5=&o1=1&o6=&o4=&o3=&r=1&s=coffee&i= - 🕨
Res		⊻ de Volkskrant 🌮 Getting Started 🔯 Latest Headlines
6	$\mathbf{\underline{V}}$	Google 💽 G Search 🔹 🖗 🥵 🕈 🛛 🗙 RS 🗸 🤣 🕈 Bookmarks* PageRank 🔹 🍄 Check 🔹 🔦 AutoLink 🖺 A
<u>_</u>	ID	WordNet Search - 3.0 - WordNet home page - Glossary - Help
<u>Q</u> 1		Word to search for: coffee Search WordNet
	Coi	Display Options: (Select option to change) 🔽 Change
Cli	La' Lo	Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
IC	No	Noun
œ	11	
-11.41	ро	• S: (n) coffee, <u>lava</u> (a beverage consisting of an infusion of ground coffee beans) "he ordered a cup of coffee"
N:	Na	atrect nyponym i juli nyponym
		 Substance meronym S (n) coffee bean, coffee berry, coffee (a seed of the coffee tree: ground to make coffee)
		• S: (n) caffeine, caffein (a bitter alkaloid found in coffee and tea that is responsible for their stimulating effects)
		 direct hypernym inherited hypernym sister term
	Hie	• S: (n) beverage, drink, drinkable, potable (any liquid suitable for drinking) "may I take your beverage order?"
		• S: (n) food, nutrient (any substance that can be metabolized by an animal to give energy and build tissue)
Na		 <u>S:</u> (n) <u>substance</u> (a particular kind or species of matter with uniform properties) "shigella is one of
		known to man"
		• <u>S:</u> (n) <u>matter</u> (that which has mass and occupies space) " <i>physicists study both the nature of</i>
Ro	Pli	govern it."
		• <u>5:</u> (n) <u>physical entity</u> (an entity that has physical existence)

Interoperability: Links between vocabularies I

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

Interoperability: Links between vocabularies II

Find semantic links between vocabulary terms:

- Van Gogh (WordNet) is Van Gogh (ULAN)
- Painter (WordNet) is more specific than Creator (VRA)
- Derain (ULAN) related-to Fauve (AAT))
- Automatic techniques exists, but performance varies.
- Often combination of automatic and manual alignment.

STICH demonstrator I

2 collections

- National Library of the Netherlands
- French National Library

Goal: access objects from one using the vocabulary of the other and vice versa.

Means: automatic semantic alignment



STICH demonstrator II

MAPPING: mandraBASE mandraNewNONE These terms define your current search. Click the 🗷 to remove a term. Refine your search further within these categories: Collection IconClass:Religion et Magie 🔀 Title: Baptême du Christ Base Mandragore de la Manuscrits enluminés Found 2869 objects Picture: Bibliothèque Nationale de la Bibliothèque de France (931) Nationale des Pays-Bas (1938) IconClass:all > Religion et Magie Religion religions chrétienne (2021) non-chrétiennes (avec astrologie (877) leurs institutions, magie, croyances coutumes et leurs antiquités) (153) supra-naturelles, occultisme (2869) représentations (symboliques) ~ création, cosmos, cosmogonie, univers, la Image taille réelle: http://visualiseur.bnf.fr/Visu vie (dans le sens le MANDRAGORE: vie pub plus large) (2) MANDRAGORE: s jean baptiste MANDRAGORE ARTS (2050) LITTERATURE ET MANDRAGORE: s esprit AUTRES FICTION (2546) PARAPSYCHOLOGIE, MANDRAGORE: RELIGIONS (1759) poisson BIBLE ET OCCULTISME, MANDRAGORE: nudité APOCRYPHES (1876) DEMONOLOGIE... (2869) BOTANIQUE (318) SCIENCES MANDRAGORE: jourdain CHRISTIANISME (1832) PURES (1181) GEOGRAPHIE (1791) SCIENCES MANDRAGORE: épée HISTOIRE (1722) SOCIALES (2188) MANDRAGORE: TECHNIQUES, cours d'eau SCIENCES PPLIQUEES (1595)



Principle 1: Semantic annotation



Description:

recordnumber
timestamp
type
collector
carios
senes
Culture
Date
Description
Identifier in Current
Repository
Creation Site
Current Repository
Material
Measurements.Format
Style/Period.Period
Title
Туре
type

23727; 2001-06-22; cultural;original; Johannes Frederik van Ov Zeldzaamheden; Utagawa 360; Japans; 1800-1829; 1883 JAPAN aankoop; 360-4564;

Japan; RMV; papier; oban, 25.5 cm x 37.5 cm; Edo; Edo junisho; prenten; Work;

MUNCH, Edvard The Scream 1893 Oil, tempera and pastel on cardboard 91 × 73 5 cm National Gallery, Oslo



Principle 2: Semantic search



Principle 3: Cross-Collection Search

"Tokugawa"



AAT style/period Edo (Japanese period) Tokugawa





E-Culture Web Demonstrator



MultimediaN N9C Eculture Project

MultimediaN

Home

News Multimedia N 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 cultururijzer.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 intermationally renowned academic expertise on Semantic Web and multimedia technology with industrial strength experience in Web design and user interface development.



en nl









Disambiguation of search term









kikvors-fetish Pieter Hovens (RMV)

kikvors-fetish Pieter Hovens (RMV)



kikvors-fetish Pieter Hovens (RMV)



kikvors-fetish Pieter Hovens (RMV)

Works created by an artist with matching name (95)



Head of a Woman (Olga ... Picasso, Pablo



Self Portrait: "Yo Pi ... Picasso, Pablo



Assemblage mounted in ... Picasso, Pablo



kikvors-fetish

Pieter Hovens (RMV)

The Young Painter Picasso, Pablo



Self-portrait with Cloak Picasso, Pablo

Works created by an artist who collaborated with an artist with matching name (31)















Work Vorld SVCN Concept ▼ Europe ▼ France AAT Concept ▼ Île-de-France Place 🔻 Ville de Paris, Dépar ... Person Paris Montparnasse 1 Montmartra facet options

depicted place 🔻

Faceted search

Paris (8)

Results grouped by



Self-Portrait with Se ... Chagall, Marc



Paris Through My Window Chagall, Marc



Homage to Bleriot Delaunay, Robert



Matisse, Henri



>>

A Glimpse of Notre Da ... Avenue de l'Opera: Mo ... Pissarro, Camille

Montmartre (3)



Pissarro, Camille



Boulevard Montmartre: ... Boulevard Montmartre: ... Boulevard Montmartre: Pissarro, Camille



Pissarro, Camille

じ Local view - Mozilla Firefox

File Edit View Go Bookmarks Tools



Done

🌸 http://e-culture.multimedian.nl/demo/local_view?uri=http%3a^

衫 😉 👩

in presentaties

de Volkskrant 🌮 Getting Started 🔂 Latest Headlines

Imayo bijin junikei (Twaalf gezichten op mooie vrouwen van nu)

Help



EN 🔇 🛒 🗞 🛄 🕅 👰 22:43



🐻 Microsoft PowerPoint ...

じ 3 Firefox

Experimental: Relation Search I





Experimental: Relation Search II

Choose URIs of Van Gogh and Gaugain.

Result: what relates them?

- 1. Van Gogh usesStyle impressionism hasPainter Gaugain.
- People network that connects Gauguin and Van Gogh: Van Gogh - studentOf - <u>Cormon, Fernand</u> -teacherOf -<u>Bernard, Emile</u> - colleagueOf Gauguin.

Current 'Hot Topics' in SW

- Scalability
- Conversion to SW standards
- Ontology Alignment
 - Metadata extraction (from text, from semistructured souces)

Final message

- Semantic Web can improve access within and across collections.
 - Structured knowledge is available for reuse.
 - Social barriers have to be overcome:
 - E.g. licences on The Getty vocabulaires



Credits

- The MultimediaN E-Culture project
- The NWO program CATCH
 STICH
 CHOICE
 - MuNCH

The eContentplus TELplus project



