



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

E-Culture MultimediaN

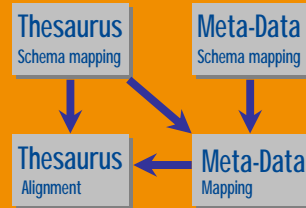
Guus Schreiber (schreiber@cs.vu.nl), Jacco van Ossenbruggen (jacco.van.ossenbruggen@cwi.nl)
Alia Amin, Lora Aroyo, Mark van Assem, Victor de Boer, Lynda Hardman,
Michiel Hildebrand, Borys Omelayenko, Anna Tordai, Jan Wielemaker, Bob Wielinga

Thesaurus and Meta-Data Meta-Data Extraction

Thesaurus and Meta-Data Alignment
We develop a novel process to convert datasets to an interoperable Semantic Web format. The datasets typically consist of meta-data about cultural objects and (extractable) thesauri of concepts used in the meta-data.

- **Thesaurus schema mapping.** Thesauri are converted to SKOS.
- **Metadata schema mapping.** Meta-data properties are converted to VRA, a specification of Dublin Core to the cultural domain.
- **Metadata mapping.** Meta-data values are converted, interpreted, and looked up in the thesaurus or external vocabularies using the information extraction techniques.
- **Thesauri alignment.** Thesauri are aligned to external standard vocabularies with the ontology alignment techniques.

We use semi-automatic methods to extract additional relations between concepts from different thesauri from heterogeneous documents on the Web. A redundancy-based method is used to extract instances of relations between Artists and Art Styles. A second method extracts temporal information about Art Styles.



Collections

(works)

Tropen Museum	80 K
Rijksmuseum Leiden	25 K
Archive.com	3.5 K
Rijksmuseum Amsterdam	3 K
Bibliopolis	2.4 K
Total:	2.7 M triples



Vocabularies

(terms)

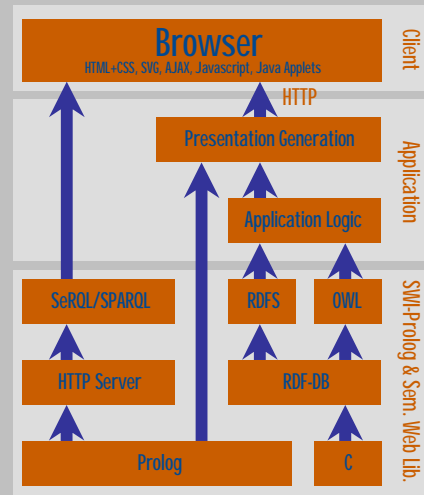
Getty TGN	894.1 K
Wordnet	464.8 K
Getty ULAN	155.4 K
Getty AAT & AATNED	32.8 K
SVCN	22.1 K
Bibliopolis terms	1.1 K
Total:	12.9 M triples



Back-End Technology

The demonstrator is built on top of the SWI-Prolog semantic web libraries and its SeRQL/SPARQL web front-end. Within MN9C we have improved the core libraries for better scalability, concurrency, persistency and fast searching for literals based on words, prefix, stem and 'sounds-like'. Multiple languages are enabled through full support for Unicode.

The RDF store can handle up to 25 million triples on 32-bit hardware while we plan to store 150 million triples on a 64-bit machine with 32GB main memory. We also developed libraries and best-practices to enhance reusability and modularity of application code in the logic-programming based environment.



User Study

We aim to conduct series of user studies to:

- Investigate how different types of users can benefit from our applications
- Evaluate our applications from different aspects such as interface, support different information seeking tasks, provenance and annotation usage

First User Study

1. Interview
 - 18 Cultural heritage experts
 - 7 Cultural heritage institutions

2. (Partial) Results

Understanding the cultural heritage experts main tasks, main sources of information and identifying the issues they face within searching for information for their professional work.

Semantic Interface

We develop interfaces that give intelligent access to multiple collections from a single entry point. The semantics play a role in three stages of the search process:

- **Query construction:** text-based search, semantic auto-completion, faceted browsing
- **Search algorithms:** graph search, domain-specific suggestions
- **Presentation of the search results:** clustering, geographical map, timeline

The integrated annotation interfaces allows users to edit existing metadata and add new resources and easily use metadata from controlled vocabularies. This implies that the system needs to keep track of the original sources for all annotations and clearly convey this in the user interface.



vrije Universiteit amsterdam



UvA UNIVERSITEIT VAN AMSTERDAM

