



Contents lists available at ScienceDirect

# Web Semantics: Science, Services and Agents on the World Wide Web

journal homepage: [www.elsevier.com/locate/websem](http://www.elsevier.com/locate/websem)

## Semantic annotation and search of cultural-heritage collections: The MultimediaN E-Culture demonstrator

Guus Schreiber<sup>a,\*</sup>, Alia Amin<sup>b</sup>, Lora Aroyo<sup>a</sup>, Mark van Assem<sup>a</sup>, Victor de Boer<sup>c</sup>, Lynda Hardman<sup>b</sup>, Michiel Hildebrand<sup>b</sup>, Borys Omelayenko<sup>a</sup>, Jacco van Osenbruggen<sup>b</sup>, Anna Tordai<sup>a</sup>, Jan Wielemaker<sup>c</sup>, Bob Wielinga<sup>c</sup>

<sup>a</sup> VU University Amsterdam, Amsterdam, The Netherlands

<sup>b</sup> Centre for Mathematics and Computer Science CWI, Amsterdam, The Netherlands

<sup>c</sup> University of Amsterdam, Amsterdam, The Netherlands

### ARTICLE INFO

#### Article history:

Received 30 July 2008

Accepted 19 August 2008

Available online 10 October 2008

#### Keywords:

Semantic search

Digital heritage

Semantic annotation

Virtual collections

### ABSTRACT

In this article we describe a Semantic Web application for semantic annotation and search in large virtual collections of cultural-heritage objects, indexed with multiple vocabularies. During the annotation phase we harvest, enrich and align collection metadata and vocabularies. The semantic-search facilities support keyword-based queries of the graph (currently 20 M triples), resulting in semantically grouped result clusters, all representing potential semantic matches of the original query. We show two sample search scenarios. The annotation and search software is open source and is already being used by third parties. All software is based on established Web standards, in particular HTML/XML, CSS, RDF/OWL, SPARQL and JavaScript.

© 2008 Elsevier B.V. All rights reserved.

### 1. Introduction

The main objective of the MultimediaN E-Culture project is to demonstrate how novel Semantic Web and presentation technologies can be deployed to provide better indexing and search support within large virtual collections of cultural-heritage resources. The architecture is fully based on open Web standards, in particular XML, RDF/OWL and SPARQL. The central hypothesis underlying this work is that the use of explicit background knowledge in the form of ontologies/vocabularies/thesauri is in particular useful for information retrieval in knowledge-rich domains.

The cultural-heritage domain is such a knowledge-rich domain. Collection holders traditionally spent considerable effort on the (manual) indexing process of collection objects. Many institutions use and develop controlled vocabularies to standardize the indexing process. The result is that the domain is dominated by a multitude of vocabularies for different subareas in many different languages. Some efforts have been made to develop collection-spanning vocabularies, such as the Getty vocabularies (see further), but it is clear that the domain is too large and diverse to be covered by a single (set of) vocabulary(ies). There is also significant vari-

ation in the annotation structure for collection objects, although many institutions use a format that is, or can be interpreted as, a specialization of Dublin Core.

Due to the abundance of vocabularies, the availability of existing semantic annotations of cultural objects, and the fact that this is mainly publicly accessible information (or at least a willingness to make it accessible), cultural heritage appears to be an ideal candidate for application of Semantic Web technology. With the growth of the World-Wide Web collection holders have been increasingly interested in making their collections available online. There are large international initiatives to make inter-collection access possible, for example the European “Europeana” initiative.<sup>1</sup> The key problems in inter-collection search lie in the different annotation formats and vocabularies used by collection holders.

The E-Culture project started out with the goal to show that inter-collection search can be achieved at relatively low cost with Semantic Web technology. The approach that we have taken roughly consists of three elements:

- (i) Providing facilities for harvesting, enriching and aligning collection metadata and vocabularies.

\* Corresponding author. Tel.: +31 20 598 7739/7718.

E-mail address: [schreiber@cs.vu.nl](mailto:schreiber@cs.vu.nl) (G. Schreiber).

<sup>1</sup> <http://www.europeana.eu>.

- (ii) Providing facilities for semantic search through the resulting graph, including various presentation mechanisms for the search results.
- (iii) Providing facilities for users to add metadata and/or content.

In this article we report on the results with respect to the first two components; work on the third component is under way and is discussed under future work. The following premises underly our approach:

- The project does not develop new ontologies/vocabularies but solely uses existing ones. The project may develop however vocabulary extensions, in particular through vocabulary alignments.
- The project uses existing metadata of multiple collections.

The online version of the demonstrator can be found at: <http://e-culture.multimedien.nl/demo/search>.

Readers are encouraged to first take a look at the demonstrator before reading on. We suggest you consult the tutorial (linked from the online demo page) which provides a sample walk-through of the search functionality. Please note that this is a product of an ongoing project. Visitors should expect the demonstrator to change. We are incorporating more collections and vocabularies and are also extending the annotation, search and presentation functionality. We are incorporating more collections and vocabularies and are also extending the annotation, search and presentation functionality. We are incorporating more collections and vocabularies and are also extending the annotation, search and presentation functionality.

Due to space limitations this article is basically a summary of the key ingredients of the MultimediaN E-Culture demonstrator, which won the Semantic Web Challenge in 2006. Readers should consult the references provided for details. Section 2 describes the semantic annotation process of collections. In Section 3 we discuss the search architecture and some details of the graph-search algorithm. Section 4 provides a peek at the demonstrator through two sample search scenarios. Research issues arising from the endeavour are discussed in Section 5.

## 2. Semantic annotation: collection data, metadata and vocabularies

At this point we have collected descriptions of 200,000 objects from six collections annotated with a range of thesauri and several proprietary controlled keyword lists, which adds up to 20 million triples (detailed statistics are available from <http://e-culture.multimedien.nl/demo/>). The objects in the collections come from the Rijksmuseum Amsterdam,<sup>2</sup> the National Museum of Ethnology,<sup>3</sup> the Royal Tropical Institute,<sup>4</sup> the Netherlands Institute for Art History,<sup>5</sup> the Royal Library,<sup>6</sup> and the Web collection Artchive.<sup>7</sup> We assume this material is representative for the described domain.

The demonstrator hosts four general thesauri, namely the three Getty vocabularies,<sup>8</sup> i.e., the Art & Architecture Thesaurus (AAT), Union List of Artists Names (ULAN) and the Thesaurus of Geo-

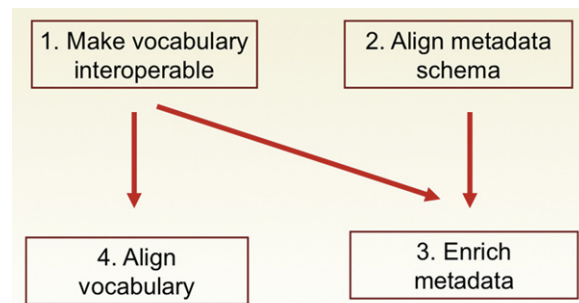


Fig. 1. Four steps of the harvesting, enrichment and alignment process of collection metadata and vocabularies.

graphical Names (TGN), as well as the lexical resource WordNet, version 2.0. The Getty thesauri were converted from their original XML format into an RDF/OWL representation using the conversion methods principles as formulated in Ref. [10]. The RDF/OWL version of the data models is available online.<sup>9</sup> The Getty thesauri are licensed.<sup>10</sup> The RDF/OWL conversion of WordNet is documented in a publication of the W3C Semantic Web Best Practices and Deployment Working Group [9]. It is an instructive example of the issues involved in this conversion process, in particular the recipes for publishing RDF vocabularies [6].

In addition, the MultimediaN E-Culture demonstrator contains collection-specific metadata and vocabularies. We assume that the collection owner provides a link to the actual data object, typically an image of a work such as a painting, a sculpture or a book. When integrating a new collection into the demonstrator we typically receive one or more XML/database dumps containing the metadata and vocabularies of the collection. The harvesting and enrichment process consists of four steps and is summarized in Fig. 1. Details of the process with a full case study can be found elsewhere [8]. The project is developing support software for this process, of which the first version has been released as open source under the name AnnoCultor.<sup>11</sup>

*Step 1: Make vocabulary(ies) interoperable.* Thesauri are translated into RDF/OWL, where appropriate with the help of the SKOS format for publishing vocabularies [7]. The same principles are followed as sketched above for the Getty and WordNet vocabularies.

*Step 2: Align metadata schema.* As a second step, the metadata schema of the collection is mapped to VRA,<sup>12</sup> a specialization of Dublin Core for visual resources.<sup>13</sup> This mapping is realized using the dumb-down principle by means of `rdfs:subPropertyOf` and `owl:equivalentProperty` relations. A full example can be found in the paper by Tordai et al. [8].

*Step 3: Enrich metadata.* Collection metadata are first transformed in a purely syntactic fashion to RDF/OWL triples, thus preserving the original structure and terminology. Subsequently, the metadata go through an enrichment process in which we process plain-text metadata fields to find matching concepts from thesauri already in the demonstrator. For example, if the `dc:creator` field contains the string “Pablo Picasso”, we will add the concept Pablo Picasso from ULAN

<sup>9</sup> <http://e-culture.multimedien.nl/resources/>.

<sup>10</sup> The partners in the project have acquired licenses for the thesauri. People using the demonstrator do not have access to the full thesauri sources, but can use them to annotate and/or search the collections.

<sup>11</sup> <http://sourceforge.net/projects/annocultor>.

<sup>12</sup> Visual Resource Association core categories, see <http://www.vraweb.org/projects/vracore4/>.

<sup>13</sup> An unofficial OWL specification of the VRA elements, including links to Dublin Core, can be found at <http://e-culture.multimedien.nl/resources/>.

<sup>2</sup> <http://www.rijksmuseum.nl>.

<sup>3</sup> <http://www.volkenkunde.nl>.

<sup>4</sup> <http://www.kit.nl>.

<sup>5</sup> <http://www.rkd.nl>.

<sup>6</sup> <http://www.kb.nl>.

<sup>7</sup> <http://www.artchive.org>.

<sup>8</sup> <http://www.getty.edu/research/conductingresearch/vocabularies/>.

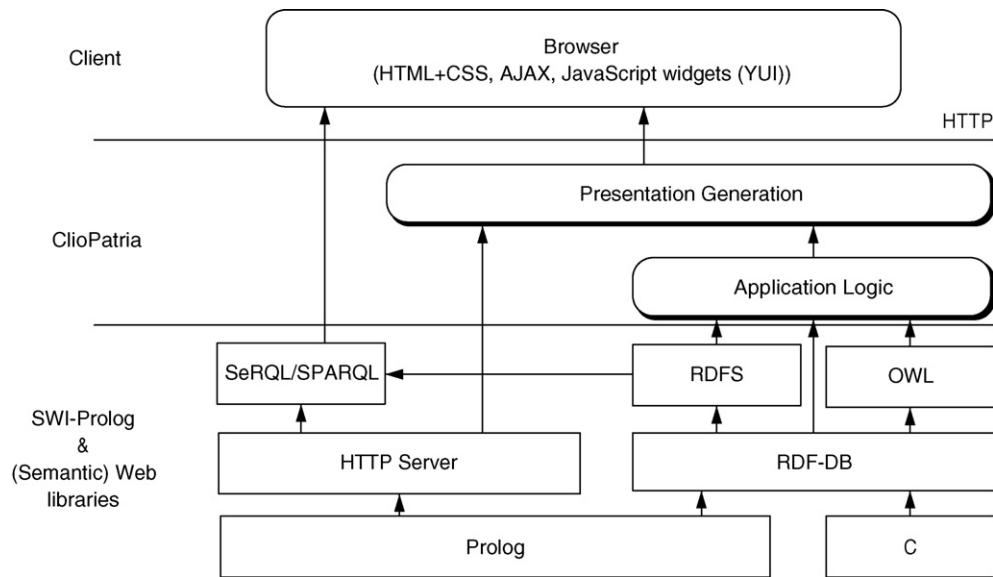


Fig. 2. ClioPatria architecture of the demonstrator.

to the metadata. Most enrichments concern people, places and materials.

*Step 4: Align vocabulary(ies).* Finally, the thesauri are aligned using `owl:sameAs` and `skos:exactMatch` relations. For example, the art style Edo from a local ethnographic collection was mapped to the same art style in AAT (see the second search scenario for an example of why such mappings are useful). Our current database (April 2008) contains 38,508 `owl:sameAs` and 9635 `skos:exactMatch` triples and these numbers are growing rapidly. Within the Getty vocabularies one set of links is systematically maintained: places in ULAN (e.g., place of birth of an artist) refer to terms in TGN. Within the project we are adding important sets of links. For example, links between art styles in AAT (e.g., “Impressionism”) and artists in ULAN (e.g., “Monet”) have a high added value for certain search strategies. de Boer [3] has worked on deriving these semi-automatically from texts on art history.

After this harvesting process we have a graph representing a connected network of works and thesauri lemmas that provide background knowledge. VRA and SKOS provide a – albeit weak – semantics, and underneath the richness of the original data is still preserved.

### 3. Semantic search

#### 3.1. Technical architecture

The technical baseline of the MultimediaN E-Culture demonstrator is formed by the ClioPatria software, built on top of SWI-Prolog and its (Semantic) Web libraries.<sup>14</sup> Fig. 2 gives an overview of the architecture. The reader is referred elsewhere for detailed information about ClioPatria [11–13]. The software is freely available under a GPL license.<sup>15</sup>

ClioPatria provides two APIs on top of the SWI-Prolog Semantic Web libraries:

- (i) A SPARQL API which supports database queries of the RDF graph.
- (ii) A graph-search API which provides limited RDF/OWL reasoning.

The graph-search algorithm for keyword-based search is briefly described in the next subsection. OWL reasoning is limited to three OWL features: symmetry (`owl:inverseOf`, `owl:SymmetricProperty`), transitivity (`owl:TransitiveProperty`), and resource equivalence (`owl:sameAs`). The algorithm also interprets similar SKOS relations (`skos:broader`, `skos:exactMatch`).

ClioPatria provides the application logic for constructing the MultimediaN E-Culture demonstrator. Basically, it provides a client-server architecture which supports the search and presentation facilities with the help of standard Web components, in particular HTML + CSS, AJAX and the Yahoo! Widget library.

Example search scenario’s with the demonstrator are shown in Section 4. Third parties are using the ClioPatria search API for other applications, for example the CHIP Rijksmuseum tour wizard [2] and the European digital heritage portal “Europeana”.<sup>16</sup> For details of ClioPatria the reader is referred to the ISWC’08 paper of Wielemaker et al. [11].

#### 3.2. Keyword search with semantic clustering

One of the goals of the demonstrator is to provide users with a familiar and simple keyword search, but still allow the user to benefit from all background knowledge from the underlying thesauri and taxonomies. The underlying search algorithm consists of several steps, that can be summarized as follows (for details see [11]). First, it checks all RDF literals in the repository for matches on the given keyword. Second, from each match, it traverses the RDF graph until a resource of interest is found, we refer to this as a *target resource*. Finally, based on the paths from the matching literals to their target resources, the results are clustered.

To improve performance in finding the RDF literals that form the starting points, the RDF database maintains a btree index of words appearing in literals to the full literal, as well as a Porter-stem and metaphone (sounds-like) index to words. Based on these indexes, the set of literals can be searched efficiently on any logical combination of word, prefix, by-stem and by-sound matches.<sup>17</sup>

<sup>14</sup> <http://www.swi-prolog.org>.

<sup>15</sup> See <http://e-culture.multimedien.nl/software/ClioPatria.shtml>.

<sup>16</sup> <http://www.europeana.eu>.

<sup>17</sup> See <http://www.swi-prolog.org/packages/semweb.html#sec:3.8>.

In the second step, which resources are considered of interest is currently determined by their type. The default settings return only resources of type artwork (`vra:Work`), but this can be overridden by the user. To avoid a combinatorial explosion of the search space, a number of measures had to be taken. Graph traversal is done in one direction only: always from the object in the triple to the corresponding subject. Only for properties with an explicit owl: `inverseOf` relation is the graph also traversed in the other direction. While this theoretically allows the algorithm to miss out many relevant results, in practice we found that this is hardly an issue. In addition to the direction, the search space is kept under control by setting a threshold. Starting with the score of the literal match, this score is multiplied by the weight assigned to the property being traversed (all properties have been assigned a (default) weight between 0 and 1), and the search stops when the score falls under the given threshold. This approach not only improves the efficiency of the search, it also allows filtering out results with paths that are too long (which tend to be semantically so far apart, that users do not consider them relevant any more). By setting the weights to non-default values, the search can also be fine tuned to a particular application domain.

In the final step, all results are clustered based on the path between the matching literal and the target result. When the paths are considered on the instance level, this leads to many different clusters with similar content. We found that clustering the paths on the schema level provides more meaningful results. For example, searching on keyword “fauve” matches works from Fauve painters Matisse and Derain. On the instance level, this results in different paths:

```
dc:creator ->
  ulan:Derain ->
    glink:hasStyle ->
      aat:fauve ->
        rdfs:label -> "Fauve"

dc:creator ->
  ulan:Matisse ->
    glink:hasStyle ->
      aat:fauve ->
        rdfs:label -> "Fauve"
```

while on the schema level, this becomes a single path:

```
dc:creator ->
  ulan:Person ->
    glink:hasStyle ->
      aat:Concept ->
        rdfs:label -> "Fauve"
```

The paths are translated to English headers that mark the start of each cluster, and this already gives users an indication why the results match their keyword. The path given above results in the cluster title “Works created by an artist with matching AAT style”. To explain the exact semantic relation between the result and the keyword searched on, the instance level path is displayed when hovering over a resulting image.

#### 4. Sample search scenario's

In this section we give two sample scenario's of the use of the MultimediaN E-Culture demonstrator. The reader is invited to try these out him/herself (see the link in Section 1). It should be noted that the collection is continuously extended, so the actual search results are likely to vary over time.

##### 4.1. Scenario 1: “Picasso”

Assume a user is typing in the query “Picasso”. Although the name Picasso is reasonably unique in the art world, the user may

still have many different intentions with this simple query: a painting by Picasso, a painting depicting Picasso, the styles Picasso has worked in? Without an elaborate disambiguation process it is impossible to tell in advance.

Fig. 3 shows part of the results of this query in the MultimediaN demonstrator. We see several clusters of search results. The first cluster contains works from the Picasso Museum. The second cluster contains works by Pablo Picasso (only first five hits shown; clicking on the arrow allows the user to inspect all results). Further down we see clusters of surrealist and cubist paintings (styles that Picasso worked in; details not shown for space reasons), and works by George Braque (a prominent fellow cubist painter, but the works shown are not necessarily cubist). Other clusters (not present in the figure) are works made from Picasso marble and works with Picasso in the title (includes two self-portraits). We are aiming to create clusters such that the user can afterwards choose herself what she is interested in. We have found that even in relatively small collections of 100 K objects users discover interesting results that they were not aware of that existed. We have termed this type of search tentatively “post-query disambiguation”: in response to a simple keyword query the user gets (in contrast to, for example, Google image search) semantically grouped results that enable further detailing of the query. It should be pointed out that the knowledge richness of the cultural-heritage domain allows this approach to work. In less rich domains this approach is less likely to provide added value.

Next to the result clustering ClioPatria support various other presentation facilities, such as showing the results on a Google map.

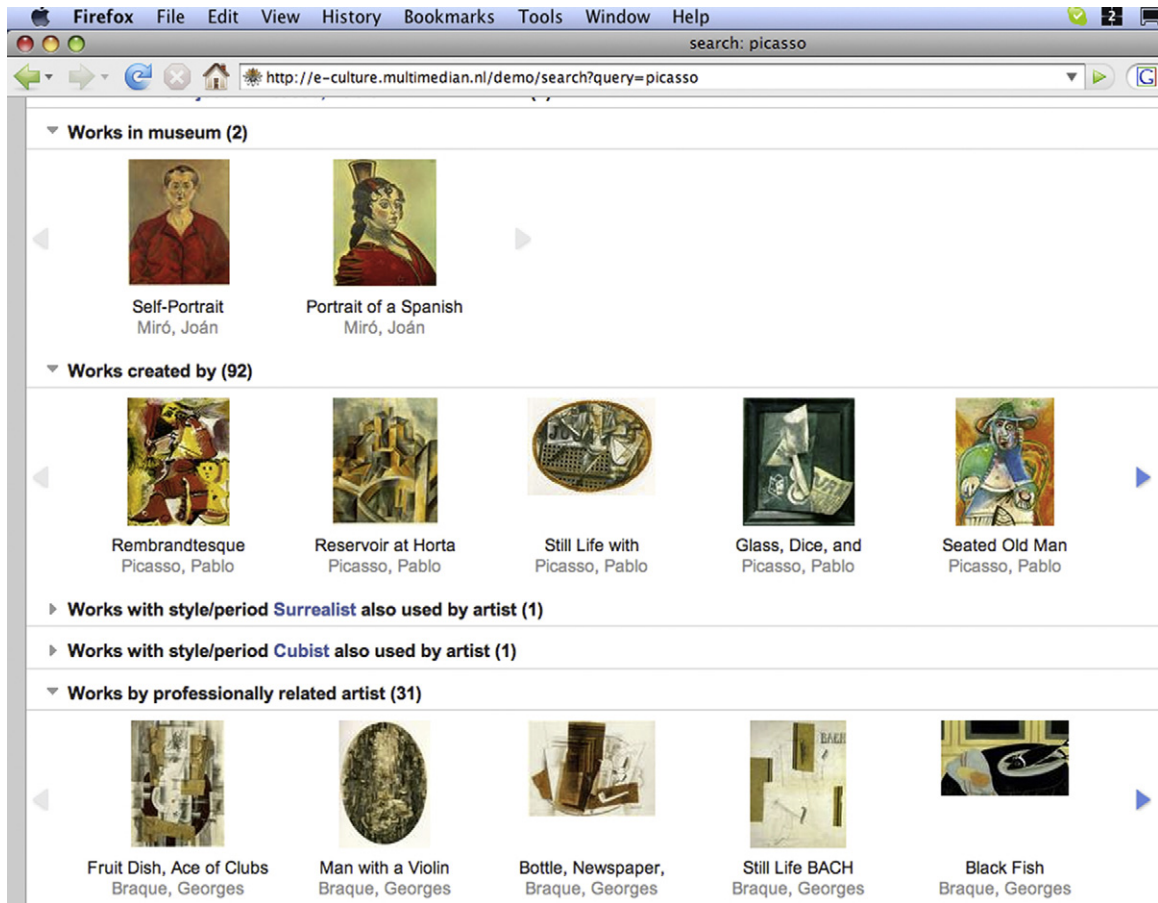
##### 4.2. Scenario 2: “Tokugawa”

Another typical search scenario concerns the exploitation of vocabulary alignments. As mentioned before, many collection owners have their own homegrown vocabulary variants. Consider the situation in Fig. 4, which is based on real-life data. A user is searching for “Tokugawa”. This Japanese term has actually two major meanings in the heritage domain: it is the name of a 19th century shogun and it is a synonym for the Edo style period. Assume for a moment that the user is interested in finding works of the latter type. The National Museum of Ethnology in Leiden actually has works on this style in its digital collection, such as the work shown in the top-right corner. However, the Dutch ethnographic thesaurus SVCN, which is being used by the museum for indexing purposes, only contains the label “Edo” for this style. Fortunately, another thesaurus in our collection, the aforementioned AAT, does contain the same concept with the alternative label “Tokugawa”. In the harvesting process we learned this equivalence link (quite straightforward: both are Japanese styles with matching preferred labels). The existence of this link allows us to retrieve the painting as a result of the “Tokugawa” query, despite the fact that it is not indexed with this term.

Although this is actually an almost trivial alignment, it is still extremely useful. The cultural-heritage world (like any knowledge domain) is full of such small local terminology differences. Multilingual differences should also be taken into consideration here. If Semantic Web technologies can help making such matches, there is a definite added value for users.

#### 5. Discussion

Over the past 2.5 years the E-Culture demonstrator has grown from 4000 to 200,000 objects. We are now planning large-scale deployment in the context of the European digital heritage portal

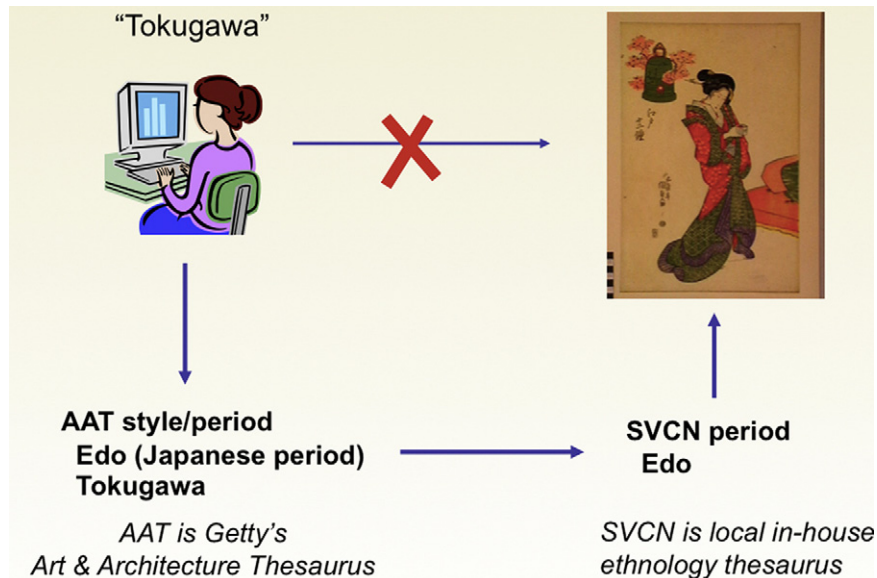


**Fig. 3.** Selection of clustered search results for query “Picasso”: works from the Picasso Museum, works by Picasso, works of art styles used by Picasso (cubist, surrealist, for space reasons only heading shown), works by professional relations of Picasso (George Braque, colleague cubist painter).

europaana.eu where we intend to grow to a collection of 12–14 M objects from musea, libraries and archives. We discuss here the lessons we learned so far, including the main research challenges we see from our perspective.

5.1. Semantic annotation

When we started the semantic annotation process with the first collection (the Artchibe collection) it was mainly a manual



**Fig. 4.** A user searches for “Tokugawa”. The Japanese painting in the top-right matches this query, but is indexed with a thesaurus that does not contain the synonym “Tokugawa” for this Japanese style. Through a “same-as” link with another thesaurus that does contain this label, the semantic match can be made.



Fig. 5. Autocompletion facility: potential matches are grouped in respective types.

art. However, while adding more collections we got more grip on the process for making a collection “ready” for the Semantic Web. The four-step process model (Fig. 1) supported by the AnnoCultor toolkit is the tangible result. In general, it takes us now 1–3 weeks to include a new collection. This may seem like a long time, but the result is a set of tools that can be run automatically each time a collection owner has to update the collection data. Only in case of (in practice relatively infrequent) changes in the schema of the metadata or of the vocabularies, additional manual work is needed.

Thesaurus conversion (step 1) is usually simple, certainly given the fact that vocabulary owners are routinely starting to “skossify” their vocabularies. Representing the interoperability of metadata schema’s through a RDF property hierarchy (step 2) is a big plus for the cultural-heritage field, where up till now cumbersome XSLT techniques prevail.

The enrichment of metadata (step 3) is basically an information-extraction task, which is by its nature simpler than general IE in document collections, due to the structured nature of the data. Recognizing “Amsterdam” as a particular location is easier when it is a string value in a dc: location field. The main research challenge here is identity resolution of works (e.g., the Night Watch appears in multiple collections). This is actually a complex problem, as we see not just simple “same-as” relations, but also “X detail of Y”, work series, etc.

With respect to vocabulary alignment (step 4) we are at the moment just looking at the “low hanging fruit”, e.g., simple syntactic alignments such as the Tokugawa example. Much more can be done here; we see this as a second critical area of research. However, Hendler’s adagium “a little semantics goes a long way” is certainly true in this domain: the current limited set of alignments boosts already the search results. The information-retrieval nature of our task helps here: the results do not have to be perfect, as long as a sufficiently large set is relevant.

One statistical fact is worthwhile to mention: on average we 17–18 metadata triples per collection object. It should be noted that these are mainly museum objects: for library and archive objects the numbers may be different.

## 5.2. Semantic search

The current RDF/OWL graph with 200 K objects and multiple (some quite large) vocabularies already poses enormous search challenges. The vocabulary concepts generate many potential graph paths (for example, check the website of ULAN to see for yourself

how much information is linked to an artist). For the moment we are still using a relatively straightforward graph-search algorithm [11], but this will likely need rethinking when the number of collection objects goes up an order of magnitude. This is definitely an important research challenge. We mention two avenues one could explore. Hollink et al. [5] has done an experiment where she tried to discover graph patterns in WordNet which increase recall without jeopardizing precision. This led to six preferred WordNet path patterns, most of these using some combination of hyponym and meronym relations. Such patterns may also exist for other vocabularies or combinations of vocabularies. Secondly, one can try to exploit metaknowledge of the hierarchy of metadata schema properties to drive the search.

In a sense, semantic search in large collections is still for the most part unexplored terrain. The problems we are facing are similar to the issues in the Linked Data initiative.<sup>18</sup> Of course, our dataset is smaller, but on the other hand the branching factor in the graph is likely to be much higher due to the knowledge-rich nature of the area.

In this article we have only addressed keyword-based search. We have experienced that users tend to prefer this type of search, because they have grown accustomed to the Google-type search. This does not mean we think this should necessarily be the only search paradigm. For example, we have experimented with faceted search [4]. We have also tentatively explored relation search: find interesting relationships between two URLs, e.g., between two artists or between an artist and a location. This is potentially an area where semantics can provide functionality that cannot be provided by standard IR techniques.

## 5.3. User involvement

Cultural-heritage partners have been sitting at our work table from the beginning. This has been an enormous help in steering the project. We have done a number of user studies, for example examining search behaviour of cultural-heritage experts [1].

User involvement is a key theme in cultural heritage. Many musea are interested in tagging (cf. Steve Museum,<sup>19</sup> Powerhouse Museum,<sup>20</sup> but are unsure how to combine this with their in-house

<sup>18</sup> <http://linkeddata.org/>.

<sup>19</sup> <http://www.steve.museum/>.

<sup>20</sup> [www.powerhousemuseum.com](http://www.powerhousemuseum.com).

annotation practices. Musea have many objects in their collection which require annotation, for which they do not have the necessary resources. At the same time they are afraid of the quality of user annotations. We are now exploring some mixed schemes in which Web users can annotate collection objects with “semantic” tags. When a user types in a term, an autocompletion facility allows her to pick the right concept. Fig. 5 shows an example of this (the autocompletion mechanism is here used in combination with keyword search, but is essentially the same). We are currently performing a case study with the Rijksmuseum to explore interactive annotation facilities. The main problems are not technical, but social: how should external user annotations be handled? This is a subject where Web 2.0 issues get intermingled with issues related to quality and trust. For example, it requires mechanisms for external annotations to be “approved” by museum professionals and for Web users to be acknowledged as experts by musea. There is still a lot of work to be done here, which brings us well out of the context of the present paper.

We view the work described in this paper as a step towards showing that the Semantic Web endeavour has a chance of succeeding, at least in knowledge-rich web “islands” such as cultural heritage. It is fair to say that in some areas we have only scratched the surface, but there are sufficient positive pointers to continue this effort. Similar encouraging experiences have been reported by the MuseumFinland project<sup>21</sup> which won the second prize in the 2004 Semantic Web Challenge.

### Acknowledgements

We are grateful to Marco de Niet, Annelies van Nispen (Digital Heritage Netherlands<sup>22</sup>), Marie-France van Orsouw and Annemiek Teasing (Netherlands Institute for Cultural Heritage<sup>23</sup>) for their valuable input. This research would not have been possible without the gracious support of the collection owners: the Rijksmuseum Amsterdam, the National Museum of Ethnology, the Royal Tropical Institute, The Netherlands Institute for Art History, the Royal Library, and the Web collection Artchive. The E-Culture project is a subproject of the MultimediaN (“Multimedia Netherlands”<sup>24</sup>) project funded by the Dutch BSIK Programme.

### References

- [1] A. Amin, J. van Ossenbruggen, L. Hardman, A. van Nispen, Understanding cultural heritage experts’ information seeking needs, in: Proceedings of the Eighth ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL’08, ACM, New York, NY, USA, 2008.
- [2] L.M. Aroyo, N. Stash, Y. Wang, P. Gorgels, L.W. Rutledge, CHIP demonstrator: semantics-driven recommendations and museum tour generation, in: Proceedings of the Sixth International Semantic Web Conference and Second Asian Semantic Web Conference, ISWC2007 + ASWC 2007, vol. 4825 of LNCS, Busan, Korea, November 11–15, 2007, Springer, Berlin, 2007.
- [3] V. de Boer, M. van Someren, B. Wielinga, Extracting instances of relations from web documents using redundancy, in: Proceedings of the Third European Semantic Web Conference (ESWC’06), Budvar, Montenegro, 2006.
- [4] M. Hildebrand, J. van Ossenbruggen, L. Hardman, Facet: a browser for heterogeneous semantic web repositories, in: Proceedings of International Semantic Web Conference ((ISWC2006), LNCS, 2006.
- [5] L. Hollink, G. Schreiber, B. Wielinga, Patterns of semantic relations to improve image content search, *J. Web Semant.* 5 (3) (2007) 195–203.
- [6] A. Miles, T. Baker, R. Swick, Best practice recipes for publishing RDF vocabularies, Working draft, W3C, <http://www.w3.org/TR/2006/WD-swbp-vocab-pub-20060314/> (March 14, 2006).
- [7] A. Miles, S. Becchofer, SKOS simple knowledge organization system reference, W3c working draft, World-Wide Web Consortium (January 25, 2008).
- [8] A. Tordai, B. Omelayenko, G. Schreiber, Semantic excavation of the city of books, in: Proceedings Semantic Authoring, Annotation and Knowledge Markup Workshop (SAAKM2007), vol. 314, CEUR-WS, 2007, <http://ceur-ws.org/Vol-314>.
- [9] M. van Assem, A. Gamgemi, G. Schreiber, Conversion of wordnet to a standard RDF/OWL representation, in: Proceedings of LREC 2006, 2006.
- [10] M. van Assem, M. Menken, G. Schreiber, J. Wielemaker, B. Wielinga, A method for converting thesauri to RDF/OWL, in: S.A. McLraith, D. Plexousakis, F. van Harmelen (Eds.), Proceedings of the Third International Semantic Web Conference ISWC2004, vol. 3298 of LNCS, Hiroshima, Japan, Springer-Verlag, Berlin/Heidelberg, 2004.
- [11] J. Wielemaker, M. Hildebrand, J. van Ossenbruggen, G. Schreiber, Thesaurus-based search in large heterogenous collections, in: A. Sheth, et al. (Eds.), Proc. The Semantic Web - ISWC 2008 7th International Semantic Web Conference, Karlsruhe, October 2008, vol. 5318 of LNCS, Springer-Verlag, Heidelberg, 2008.
- [12] J. Wielemaker, G. Schreiber, B. Wielinga, Using triples for implementation: the Triple20 ontology-manipulation tool, in: Y. Gil, E. Motta, R. Benjamins, M. Musen (Eds.), The Semantic Web—ISWC2005: Proceedings of the Fourth International Semantic Web Conference, vol. 3729 of Lecture Notes in Computer Science, Galway, Ireland, November 6–10, 2005, Springer-Verlag, 2005.
- [13] J. Wielemaker, G. Schreiber, B.J. Wielinga, Prolog-based infrastructure for rdf: performance and scalability, in: D. Fensel, K. Sycara, J. Mylopoulos (Eds.), The Semantic Web—Proceedings of ISWC’03, vol. 2870 of Lecture Notes in Computer Science, Sanibel Island, FL, Springer-Verlag, Berlin/Heidelberg, 2003, ISSN 0302-9743.

<sup>21</sup> <http://www.museosuomi.fi>.

<sup>22</sup> <http://www.den.nl>.

<sup>23</sup> <http://www.icn.nl>.

<sup>24</sup> <http://www.multimedien.nl>.