
Semantic Portals for Cultural Heritage

Eero Hyvönen

Semantic Computing Research Group
Helsinki University of Technology (TKK) and University of Helsinki
eero.hyvonen@tkk.fi, <http://www.seco.tkk.fi/>

1 Towards Semantic Portals

Cultural content on the web is available in various forms (documents, images, audio tracks, videos, collection items, learning objects etc.), concern various topics (art, history, handicraft, etc.), is written in different languages, is targeted to both laymen and experts, and is provided by different independent memory organizations (museums, archives, libraries) and individuals. The difficulty of finding and relating information in this kind of heterogenous content provision and data format environment creates an obstacle for end-users of cultural contents, and a challenge to organizations and communities producing the contents.

Portals try to ease these problems by collecting content of various publishers into a single site [43]. Portal types include *service portals* collecting a large set of services together (e.g., Yahoo! and other “start pages”), *community portals* [46] acting as virtual meeting places of communities, and *information portals* [36] acting as hubs of data. Much of the semantic web content will be published using *semantic information portals* [31, 36]. Such portals are based on semantic web standards¹ and machine “understandable” content, i.e., metadata, ontologies, and rules, in order to improve structure, extensibility, customization, usability, and sustainability of traditional portal designs.

Cultural heritage is a promising application domain for semantic portals [20, 41, 50]. They are useful from the end-users’ view point in several ways:

- *Global view to heterogeneous, distributed contents.* The contents (e.g., museum collections) of different content providers can be provided through one service as if it were a single, seamless homogenous repository [20]. Only a single user interface has to be learned.
- *Automatic content aggregation.* Satisfying an end-user’s information need often requires *aggregation* of content from several information providers [43, 21], a task suitable for semantic web technologies. For example, when looking for data

¹ <http://www.w3.org/2004/SW/>

about an artist, relevant information may be provided by museum collections, libraries, archives, authority records, ontologies, and other sources.

- *Semantic search.* In traditional portals search is usually based on free text search (e.g., Google), database queries, and/or a stable classification hierarchy (e.g., Yahoo! and dmoz.org). Semantic content makes it possible to provide the end-user with more “intelligent” facilities based on ontological concepts and structures, such as *semantic search* [7], *semantic autocompletion* [19], and *(multi-)faceted semantic search* [35, 15, 24, 40, 51].
- *Semantic browsing and recommendations.* Semantic content also facilitates semantic browsing [13] (cf. Chapter 36) and recommendations [52] (cf. Chapter 34). Here semantic associations between search objects can be exposed to the end-user as recommendation links, possibly with explicit explanations.
- *Other intelligent services.* Also other kind of intelligent services can be created based on machine interpretable content, such as knowledge and association discovery [42], personalization [2], and semantic visualizations based on e.g. historical and contemporary maps and time lines [28].

Semantic portals are very attractive from the content publishers viewpoint, too:

- *Distributed content creation.* Portal content is usually created in a centralized fashion by using a content management system (CMS). This approach is costly and not feasible if content is created in a distributed fashion by independent publishers, e.g. by different of museums and other memory organizations. Semantic technologies can be used for harvesting and aggregating distributed heterogenous content (semi-)automatically into global content portals [20].
- *Automated link maintenance.* The problems of maintaining links up-to-date is costly from the portal maintenance viewpoint. In semantic portals links can be created and maintained automatically based on the metadata and ontologies.
- *Shared content publication channel.* In the cultural domain the publishers usually share the common goal of promoting cultural knowledge in public and among professionals. A semantic portal can provide the participating organizations with a shared, cost-effective publication channel [22].
- *Enriching each other’s contents semantically.* Interlinking content between collaborating organizations enriches the contents of everybody “for free”.
- *Reusing aggregated content.* The content aggregated into a semantic portal can be reused in different applications and cross-portal systems [27, 33].

A cultural semantic information portal includes of the following major components. First, we need a *content model* for representing cultural metadata, ontologies, and rules. Second, a content creation system is needed for creating and harvesting content. Third, the portal publishes semantic services for 1) human end-users as *intelligent user interfaces* and possibly for 2) other portals and applications as *web services*. In the following these components are explained in more detail.

2 Content Models for Semantic Cultural Portals

The semantic web “layer cake model” of Tim Berners-Lee makes the distinction between a syntactic data level based on the Extensible Markup Language XML², and semantic levels above it:

- *Metadata level.* The RDF data model³ (cf. Chapter 3) is used for representing metadata about cultural resources.
- *Ontology level.* The RDF Schema and web ontology language OWL⁴ (cf. Chapter 4) are used for representing ontologies [11] (cf. Chapter 19) that describe vocabularies and concepts concerning the real world and our conception of it.
- *Logic level.* Logic rules (Cf. Chapter 5) can be used for deriving new facts and knowledge based on the metadata and ontologies.
- *Trust level.* At the highest conceptual level issues of e.g. trustworthiness of content, copyrights etc. are of concern.

In the following, metadata, ontology, and logic layers are considered from the viewpoint of semantic cultural portals. Issues related to trust in the cultural domain have thus far not been discussed much in the literature. Trust issues are important e.g. in Web 2.0 spirited social cultural portals, such as Steve⁵, where end-user create and tag content.

2.1 Metadata Schemas

Cultural content in museum collections, libraries, and other content repositories is usually described using *metadata schemas* (also called *annotation schemas* or *annotation ontologies*). These templates specify a set of obligatory and optional elements, i.e. properties, by which the metadata for content items should be described. For example, the Dublin Core (DC) Metadata Element Set⁶ lists 15 standardized⁷ elements, such as dc:title, dc:creator, and dc:subject, with additional elements and element refinements. Encoding guidelines tell how to express the elements in RDF/XML and using HTML/XHTML meta and link elements. Qualifiers, such as encoding schemes, enumerated lists of values, and other processing clues are used to provide more detailed information about a resource. For example, “date” is a DC element that can further be specified as “date published” or “date last modified”. The core elements can be extended in an interoperable way by using the “dumb-down” principle. It means that in any use of a qualified DC element, the qualifier may be dropped and the remaining value of the element should still be a term that is useful for discovery, although with less precision.

² <http://www.w3.org/XML/>

³ <http://www.w3.org/RDF/>

⁴ <http://www.w3.org/2004/OWL/>

⁵ <http://www.steve.museum/>

⁶ <http://dublincore.org/documents/1998/09/dces/>

⁷ NISO Standard Z39.85-2001 and ISO Standard 15836-2003

DC is used as a basis in more detailed cultural metadata schemas, such as the Visual Resource Association's (VRA) Core Categories⁸. Its element set provides a categorical organization for the description of works of visual culture as well as the images that document them. Most VRA elements are defined as subproperties of corresponding DC elements. An example of an instance of VRA metadata in the CHIP portal [2] is given below in RDF Turtle notation⁹. The schema has properties such as `vra:type` (the type of the art-work as a reference to the VRA vocabulary), `vra:title` (literal title of the art-work), `vra:creator`, `vra:subject`, `vra:culture`, and `vra:material`. Element values with a namespace are references to underlying ontologies.

```
rijks:artefactSK-C-K
  vra:type vra:Work ;
  vra:title "The Night Watch" ;
  vra:date "1642" ;
  vra:creator: 500011051 ;           # Rembrandt
  vra:subject iconclass:45F31 ;     # Call to arms
  vra:culture tgn:7006952 ;        # Amsterdam
  vra:material aat:30015050 .      # Oil paint
```

Syntactic interoperability in semantic portals can be obtained by using shared metadata schemas in metadata descriptions. A schema makes it possible to specify relevant aspects of the search objects, such as the “author”, “title”, and “subject” of a document, and focus search according to these. Syntactic interoperability facilitates, for example, multi- or metasearch¹⁰. Here the user types in a query in a metaportal. The query is then distributed to a set of underlying systems and the results are aggregated for the end-user. Protocols such as Z39.52¹¹ and Search and Retrieve via URL (SRU)¹² of the Library of Congress can be used here. For example, the Australian Museums and Galleries Online¹³ and Artefacts Canada¹⁴ are multi-search engines over nation-wide distributed cultural collections.

Another approach to creating metaportals is to first harvest the content into a global database, and the search the global repository. Protocols such as Open Access Initiative Protocol for Metadata Harvesting (OAI-PMH)¹⁵ can be used for distributed content publishing and harvesting.

A distinctive feature between cultural content aggregation systems is whether they deal with metadata that conforms to a *single metadata schema* or *multiple schemas*. Semantic web portals have tackled the problem of semantic interoperability usually by extending and sharing metadata schemas. For example, in [20] heterogeneous artifact collection databases were made semantically interoperable, but the content was homogenous, i.e., artifacts, and the metadata was based on a single, Dublin Core like metadata schema.

⁸ <http://www.vraweb.org/>

⁹ <http://www.dajobe.org/2004/01/turtle/>

¹⁰ http://en.wikipedia.org/wiki/Metasearch_engine

¹¹ <http://www.cni.org/pub/NISO/docs/Z39.50-brochure/>

¹² <http://www.loc.gov/standards/sru/>

¹³ <http://www.amonline.net.au/>

¹⁴ <http://www.chin.gc.ca/>

¹⁵ <http://www.openarchives.org/>

Ontologies can be applied to addressing the problem of interoperability between different metadata schemas. CIDOC Conceptual Reference Model (CIDOC CRM) [8] (cf. Chapter 19) is an annotation ontology standard¹⁶ developed as an underlying schema into which other metadata schemas in the cultural domain can be transformed for interoperability. This model “provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation”¹⁷. The framework includes 81 classes, such as *crm:Man-Made Object*, *crm:Place*, and *crm:Time-Span*, and a large set of 132 properties relating the entities with each other, such as *crm:Has Time-Span* and *crm:Is Identified By*.

2.2 Vocabularies and Ontologies

Metadata schemas specify data formats but do not tell how to fill the element values in the formats. Additional standard and guidelines are necessary to guide the choice of terms or words (data values) as well as the selection, organization, and formatting of those words (data content). Data value standards have been traditionally specified by constructing controlled vocabularies and thesauri [12, 1]. Examples of cultural thesauri include the Thesaurus for Graphic Materials I (TGM I)¹⁸ for indexing pictorial materials, ICONCLASS¹⁹ for art, Art & Architecture Thesaurus (AAT)²⁰ for fine art, architecture, decorative arts, archival materials, and material culture, Union List of Artist Names (ULAN)²¹, the Thesaurus of Geographic Names (TGN)²², the Library of Congress Authority Files²³, and the terminologies and standards of the MDA (formerly Museum Documentation Association)²⁴. An example of a data content standard is the Cataloging Cultural Objects (CCO) guidelines²⁵.

By filling in the values of metadata elements using shared ontologies semantic interoperability of cultural contents can be obtained. A prototypical example of this approach is the portal “MUSEUMFINLAND—Finnish Museums on the Semantic Web”²⁶ [20]. Here distributed, syntactically heterogeneous museum collection databases are integrated by a set of seven shared ontologies (Object types, Locations, Materials, etc.) [22], and semantic search and browsing services are provided to end-users based on the aggregated RDF knowledge base.

Many cultural thesauri have been transformed [49, 48] into SKOS format²⁷ to be used in cultural semantic portals [41, 50]. However, although a syntactic trans-

¹⁶ Since 2006 it has been an official ISO standard 21127:2006.

¹⁷ <http://cidoc.ics.forth.gr/>

¹⁸ <http://www.loc.gov/rr/print/tgm1/>

¹⁹ <http://www.iconclass.nl/>

²⁰ http://www.getty.edu/research/conducting_research/vocabularies/aat/

²¹ <http://www.getty.edu/vow/ULANSearchPage.jsp>

²² http://www.getty.edu/research/conducting_research/vocabularies/tgn/

²³ <http://authorities.loc.gov/>

²⁴ <http://www.mda.org.uk/stand.htm>

²⁵ <http://www.vraweb.org/ccoweb/cco/index.html>

²⁶ Operational at <http://www.museosuomi.fi> with a tutorial in English.

²⁷ <http://www.w3.org/2004/02/skos/>

formation into SKOS is useful, it is not always enough from a semantic viewpoint. The fundamental problem with traditional thesauri is that its semantic relations have been constructed mainly to help the indexer in finding indexing terms, and understanding the relations needs implicit human knowledge. Unless the meaning of the semantic relations of a thesaurus is made more explicit and accurate for the computer to interpret, the SKOS version is equally confusing to the computer as the original thesaurus, even if semantic web standards are used for representing it.

For example, there are major problems in utilizing the Broader Term (BT) relations of thesauri [23]: 1) BT relations do not necessarily structure the terms into a full-blown hierarchy that would be useful e.g. in faceted search but into a forest of small subhierarchies. 2) The semantics of the BT relation is ambiguous: it may mean either subclass-of-relation, part-of relation (of different kinds, cf. [10]), or instance-of relation. As a result, the BT relation cannot e.g. be used for property inheritance. 3) The transitivity of the BT relation chains is not guaranteed from the instance-class-relation point of view. If x is an instance of class A whose broader term is B , then it is not necessarily the case that x is an instance of B , although this is a basic assumption in RDFS and OWL. For example, assume that x is an instance of “make-up mirror”, whose broader term is “mirror”, and that its broader term is “furniture”. When searching with the concept “furniture” one would expect that instances of furniture are retrieved, but in this case the result would include confusingly make-up mirrors, too, if transitivity is assumed. A solution to these fundamental problems is to actually refine and reorganize the semantic structures of a thesaurus into a light-weight ontology e.g. along the lines proposed in [23].

Several domain ontologies are used in describing cultural metadata. This raises up the problem of making ontologies mutually interoperable. There are solution approaches for this [23], such as ontology mapping and alignment [14] (cf. Chapter 26), sharing common foundational logical principles DOLCE²⁸ (cf. Chapter 16), and using shared horizontal top ontologies, such as the IEEE SUMO²⁹. It is likely, that in many cases several identifiers (URIs) will be in use for denoting a single concept even if this is not desirable in general. For example, registries of same geographical locations are maintained at different countries and by different service providers using their own identifiers. In such cases, dereferencing services will be needed to map resource identifiers denoting same concepts with each other.

A form of semantic interoperability is encountered when dealing with contents conforming to different metadata schemas (ontologies). CIDOC CRM addressed this problem by providing over 200 concepts and relations for cultural heritage documentation, but the standard does not contain domain ontologies for filling in property values. CULTURESAMPO [21, 38] addresses the problem of semantic interoperability of metadata schemas on a deeper domain vocabulary level. The cultural content types in the system include a wide variety of cultural objects, such as artifacts in museum collections, paintings, photographs, videos, music, narratives (e.g. biographies and epics), cultural processes (e.g., semantic descriptions of farming, booth

²⁸ <http://www.loa-cnr.it/DOLCE.html>

²⁹ <http://suo.ieee.org/>

making), cultural buildings and sites, and historical events. The metadata used conforms to several schemas, including CIDOC CRM. Content integration is performed by transforming content into a light-weight knowledge representation scheme based on domain ontology events and their thematic roles [45], such as agent, goal, and place. For example, the DC metadata of a painting tells that there has been a painting event with the dc:creator in the agent role. This event instance can be used for enriching the painter's biography, that is also represented in terms of underlying events, such as the agent "being born" or "dying" at a certain place in a certain time. In the same vein, relational embedded meanings of metadata schemas, such as the meaning of the CIDOC CRM property *P13B.was_destroyed_by*, can be represented as a destroying event.

2.3 Rules for Cultural Heritage

A collection of cultural metadata and related ontologies constitute a knowledge base. On the logical level, rules can be used for deriving new facts and knowledge based on the repository, i.e., for explicating the implicit content of the repository, and enriching the content semantically. Some examples illustrating different ways of using rules in semantic cultural portals and systems are given below.

- *Explicating content of metadata schemas.* Many metadata formats contain implicit knowledge embedded e.g. in the relational meaning of the element names. In [38] rule sets for three cultural metadata schemas are presented for explicating such knowledge in terms of events.
- *Enriching semantic content.* Common sense rules may be used for enriching annotations, thus extending the machine's understanding about culture. In [24], for example, family relation rules (and others) we used to explicate implicit family relations, such as "grand father of", between persons in order to link photographs of relatives together while browsing the repository.
- *Semantic recommendations with explanations.* In [20] some 300 rules and associations, such as "doctoral hats are related to academic ceremonies" or "distaffs are related to spinning events", were used to represent simple common sense knowledge and associations between ontological concepts. A semantic recommendation service was then established that, based on additional logical rules, could 1) dynamically find out chained semantic associations between cultural objects based on ontologies and the common sense relations, and 2) at the same time construct literal explanations of why the association would be of interest. In [26] semantic process descriptions of cultural processes, such as traditional farming and fishing, were used as basis for relating cultural objects with each in meaningful ways. For the same reason, this system also included semantic narrative descriptions of a few poems of the national epic Kalevala that is related in many ways with the underlying cultural objects, such as paintings inspired by the epic and artifacts related to the images.
- *Projecting search facets.* In faceted search, rules can be used for constructing facet hierarchies based on ontological structures, such as the rdfs:subClassOf of

part-of-relations [20]. Furthermore, rules can be used to solve problem of projecting search items to facet categories, which may be complicated [52, 25, 24]. From a software engineering viewpoint, using logic rules for projections separates facets from the annotation ontologies and annotations, which makes it possible to apply the same faceted search engine [32] to knowledge bases based on different kind of ontologies and annotation schemas.

- *Association discovery*. Association discovery can be based on rules trying to find paths between resources in a knowledge base [41, 21].

3 Cultural Content Creation

Several kinds of content need to be created for a semantic portal, including 1) ontologies, 2) terminologies, and 3) semantic annotations [22]. Also creating rules for e.g. semantic recommendations can be seen as a form of content to be created. In below, ontology, terminology, and annotation creation are discussed in some more detail.

3.1 Ontology and Terminology Creation

The core of a semantic cultural heritage portal is a set of domain ontologies describing the cultural concepts (e.g., artifacts, places, materials, persons etc.) of the domain of discourse. In creating domain ontologies it is advisable to try to re-use existing ontologies or transform existing thesauri into SKOS or other semantic web formats, as discussed earlier. Ontologies can also be created or enhanced manually using an ontology editor such as Protégé-2000³⁰. With instance-rich large ontologies, such as location ontologies (geographical names) and actor ontologies (persons and organizations) *ontology population* is often used. It is a process, where the class structure of the ontology already exists, and is extended with individuals from e.g. a database.

A challenge of ontology construction is dealing with *free indexing* keywords (and corresponding concepts) . These include e.g. names of persons, organisms and places that cannot be enumerated exhaustively as ontological concepts beforehand. When a previously unknown free indexing term is encountered, the corresponding ontology need be updated with a new concept so that it can be referred to later on, if the concept encountered again [22].

For example, when constructing the location ontology of [20] a small ontology of concepts such as “country”, “city”, “lake”, etc. was first constructed manually. This class ontology was then populated semi-automatically first from official data sources, such as the list of Finnish cities and counties. However, most of the location instances had to be populated from the collection databases, since they included lots of place names that were not available in the official data sources, such as names of estates and historical locations. For some locations, the part-of relations constituting a backbone of the ontology, could be identified automatically during population. This is because many collection data entries contain both a general and a more particular

³⁰ <http://protege.stanford.edu/>

location term, in the style “Paris in Texas” or “Paris in France”. However, in many most cases, the `rdf:type` and `part-of` properties of new unknown instances had to be edited or at least checked by hand by a human editor.

The content providers often use different literal terms to refer to same resources (synonymy) when describing metadata in legacy systems. For example, literals “Nice” and “Nizza” may be used to refer to the French Mediterranean city. On the other hand, the same term may be used to refer different concepts (homonymy), such as “bank”. To resolve the meaning(s) of terms used in metadata, a terminology mapping is needed for defining the relation between linguistic words and expressions and the corresponding ontological concepts. It may be wise to create a separate term mapping for each content provider, in order to make the portal flexible with respect to the variance in terminologies used by different content providers. In this way local terminological conventions can be used as long as the local term mapping tells the meaning of the literal terms by a URI reference to the domain ontologies [22].

3.2 Creating Semantic Annotations

Semantic annotations can be created either by transforming metadata from existing databases into RDF, or by annotating new material semantically.

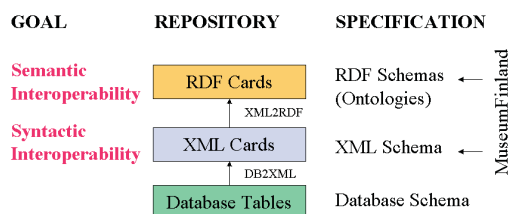


Fig. 1. Transforming legacy museum collection data from a database into RDF.

Transforming Legacy Literal Metadata

Content in memory organizations is usually available as relational legacy databases, whose annotations are based on literal terms and free text descriptions. Such annotations are often indented for human usage, use various syntactic conventions, are often semantically ambiguous, and may contain syntactic typing errors. Transforming such metadata records into semantically interoperable RDF format needed for the semantic web involves two major steps (cf. figure 1). First, syntactic interoperability among all data sources is obtained by transforming the metadata into a format defined, e.g., by an XML metadata schema that is shared by the co-operating content providers. It is possible to transform relational databases directly into RDF, but in many cases the intermediate XML level is helpful because 1) such syntactic metadata specifications

may already be in use and 2) the database-to-XML transformation can be made fairly easily at content provider organizations without referencing to ontologies. The XML level lets every museum to decide which database fields to use in filling the XML element values. Second, semantic interoperability between metadata sources is obtained by transforming the XML descriptions into the final RDF metadata schema format used by the portal. During the XML-to-RDF transformation the essential task is to move from term space into concept space by changing literal terms, used at the XML level as certain element values, into corresponding concept URIs of the domain ontologies. For example, the value “Finland” in the `dc:spatial` element could be transformed into the URI `http://www.ontology.fi/locations#Finland`. The URIs created in this phase connect metadata RDF with ontology RDF resulting into a single large semantic RDF triple store. This store can be used for querying and as a basis for logical reasoning.

A major problem in the RDF transformation is how to disambiguate the meanings of homonyms. The type of the metadata element in which a homonymous expression is used can often be used for semantic disambiguation effectively [22]. However, when dealing with the `dc:subject` element (or similar ones) that can have all kinds of ontological values, such contextual disambiguating information is not available, and human intervention is needed.

Another practical problem is spelling errors encountered frequently in legacy databases, and the variance of correct syntactic encoding practices used at different organizations at different times, in different languages, and even by different catalogers. For example, the name of Ivan Aivazovsky (Russian painter, 1817–1900) has 13 different labels in ULAN (Ajvazovskij, Aivazovski, Aiwassoffski etc.), and the first, middle, and last names can be ordered and shortened in many different ways. Still another problem is complicated free text descriptions that may be used as elements values, such as material description “cow leather with some copper decorations and painted figures”.

Annotating New Content Semantically

Free text descriptions in metadata are in general difficult search for due their syntactic variance, and for the same reason, difficult to transform into URI references automatically. The problem can be approached by using in indexing controlled vocabularies that can be represented as ontologies or reorganized into such. However, even then the problem of dealing with free indexing terms remains. In a distributed content creation environment, new free indexing concepts have to be populated into ontologies and be shared, too. For example, when a painting of a new, formerly unknown artist is cataloged in a museum, the other museums should be made aware of her/him in order to prevent creation of multiple identifiers for the artist and later confusion of identities.

A solution approach to this is to connect annotation creation tools to centrally maintained *ontology library services* that provide the clients with up-to-date information about the vocabulary resources available, and facilitates creation and sharing of new resources collaboratively. An implementation of such a service is the ONKI

Ontology Server³¹ intended for homogenizing and sharing URI references at a national level in the Finnish semantic web infrastructure [23].

Sharing unique URIs for concepts is preferable on the semantic web, but in practice there will be multiple URIs referring to a single resource, at least before global standards are adopted everywhere. For example, different countries are likely to use their own identifiers already in use for their geographical locations. For such situations, global dereferencing services will be needed in the future telling e.g. that the Finnish concept “Lontoo” refers to the same concept as “London” in U.K. and “Londres” in France.

After creating semantically interoperable RDF metadata, content harvesting and aggregation can be done either 1) off-line before starting the portal or 2) online dynamically when answering end-user queries. The online approach is more dynamic. However, from the viewpoint of creating intelligent end-user services, the off-line approach seems more promising: 1) By creating a global knowledge base first off-line, reasoning can be easily done at the global scale across local contents, which facilitates e.g. generation of recommendation links between the content of different content providers. 2) Knowledge can be compiled and critical reasoning tasks performed off-line for faster response times by performing critical reasonings tasks beforehand (knowledge compilation). For example, the `rdf:type` instance-class-relations can be explicated as RDF-triples based on the transitive closures of the `subClassOf`-hierarchies. 3) The portal is independent on the content providers possibly unreliable web services when running the system.

4 Semantic Portal Services

The goal of semantic information portals for cultural heritage is to provide the end-user with intelligent services for finding and learning the right information based on her own conceptual view to culture and the context of using the system. In the following, some possibilities of providing the end-users with intelligent services based on semantically annotated metadata in portals are shortly reviewed.

Semantic Search

In traditional information retrieval [3] search is usually based on finding occurrences of words on documents. On the semantic web, semantic search can be based on finding the concepts related to the documents at the metadata and ontology levels, in addition to the actual text or other features of the data. With concept-based methods document meanings and queries can be specified more accurately which in general leads to better recall and precision, especially if both the query and the underlying content descriptions are concept-based.

With non-textual cultural documents, such as paintings, photographs, videos, and multimedia, metadata-based search techniques are a must in practice, although also

³¹ <http://www.seco.tkk.fi/services/onki/>

content-based information retrieval methods [37] (CBIR) and multimedia information retrieval (MIR) [30] can be used as complementary techniques. Here the idea is to utilize actual document features (at the data level), such as color, texture, and shape in images, as a basis for information retrieval. For example, an image of Abraham Lincoln could be used as a query for finding other pictures of him, a piece of music could be searched for by humming it, and goal making events on a video can be searched for by learning their characteristic features. Bridging the “semantic gap” between low level image and multimedia features and semantic annotations is an important but challenging research theme [18].

Still another approach to do “semantic search” is to analyze and build search on the content using linguistic and/or statistical methods, without using annotated semantic metadata [5].

A key problem of semantic search is mapping the literal search words used by humans to underlying ontological concepts, used by the computer. Depending on the application, only queries expressed by terms that are relevant to the domain and content available are meaningful, other queries result in frustrating “no hits” answers.

A way solve the problem is to provide the end-user with a vocabulary as a subject heading category tree, a facet, as in Yahoo! and dmoz.org. By selecting a category, related documents are retrieved. Faceted search [15, 24, 51] is a natural of generalization of this, where the user can make several simultaneous selections from *multiple* orthogonal facets. The facet views are exposed to the end-user in order to 1) provide her with the right query vocabulary, and 2) for presenting the repository contents and search results and the amounts of hits along different facets. The result set can be presented to the end-user according to the view hierarchies for better readability. This is in contrast with traditional search where results are typically presented as a list of decreasing relevance. By precalculating always the number of hits for possible category selection the user can make next, it is possible eliminate selections leading to “no hits” dead-ends, and to guide the user in making next constraining selections on the facets.

Faceted search has been integrated with the notion of ontologies and the semantic web [24, 20]. The idea of such *semantic faceted search* is to construct facets algorithmically from a set of underlying ontologies that are used as the basis for annotating search items. Furthermore, the mapping of search items onto search facets could be defined using logic rules. This facilitated more intelligent semantic search of indirectly related items. A method for ranking the search results in faceted search based on fuzzy logic is presented in [17], and [47] develops a card sorting approach for specifying user facets independently from the indexing ontologies.

The faceted (view-based) search paradigm [35, 15, 24, 40] is based on *facet analysis* [34], a classification scheme introduced in information sciences by S. R. Ranganathan already in the 1930’s. The idea of faceted search has been invented and developed independently by several research groups, and is also called view-based search³² [35] and dynamic taxonomies [40].

³² A short history of the view-based search is presented in <http://www.view-based-systems.com/history.asp>

Semantic Autocompletion

Faceted search is becoming more and more popular, but is not a panacea for all information retrieval tasks. Google-like keyword search interface is usually preferred if the user is capable of expressing her information need terms of accurate keywords [9]. Keyword search can be integrated with semantic search. For example, in MUSEUMFINLAND [20] keyword search is integrated with faceted search in the following way: First, search keywords are matched against category names in the facets in addition to text fields in the metadata. The result set of hits is shown containing all objects in any of the categories matched in addition to all objects whose metadata directly contains the keyword. The hits are grouped by the categories found. Second, a new dynamic facet is created in the user interface for disambiguating the different possible ontological interpretations of the keyword. This facet contains all facet categories whose name (or other property values) matches the keyword. They tell the end-user the different interpretations of the keyword, and by selecting one of them the user is able to disambiguate the meanings and constrain search further. For example, keyword “Nokia” matches in the portal to the mobile phone company resource in the “Manufacturer” facet, and to the city of Nokia in the facets “Place of manufacturing” and “Place of usage”. In addition also some other interpretations of the literal keyword can be found and presented to the user for disambiguating her intent.

The idea of searching facet categories for disambiguating intended meanings has been generalized into the notion of semantic autocompletion [19]. This approach generalizes traditional text autocompletion by trying to guess, based on ontologies and reasoning, the search concept the user is trying to formulate after each input character in an input field. For example, the user may type in the query in French and the semantic autocompletion service finds the possible intended search concepts in English after each input character. Autocompletion as a way to find meaningful keywords in large search vocabularies has become popular after Google Suggest³³ was released, and is used in semantic cultural portals, such as [27, 21, 41, 50].

Semantic Browsing and Recommending

In addition to semantic search, semantic content facilitates *semantic browsing*. Faceted search is already a kind of combination of searching and browsing because search is based on selecting links on facets. However, in semantic browsing the general idea is not to constrain the results set but rather to expand it by trying find objects of potential interest outside of the hit list. The idea is to support browsing documents through associative links that are created based on the underlying metadata and ontologies, not on hardwired anchor links encoded by humans in HTML pages.

A simple form a semantic browser is RDF browsers and tabulators [4]. Their underlying idea has been explicated as the “linked data”³⁴ principle proposing that

³³ <http://www.google.com/webhp?complete=1&hl=en>

³⁴ <http://www.w3.org/DesignIssues/LinkedData.html>

when an RDF resource (URI) is rendered in a browser, the attached RDF links to related resources should be shown. When one of these links is selected, the corresponding a new resource is rendered, and so on.

A more developed related idea is *recommendation systems* [6]. Here, however, the logic of selecting and recommending of related resources can be based of also other principles than the underlying RDF graph. For example, collaborative filtering [16] is based on browsing statistics of other users. In [24] and [20] logic rules for creating semantic recommendation links [52] to related historical photographs were employed. The rules, written in SWI-Prolog³⁵ on top of an RDF knowledge base, not only associated collection item pages with each other, but at the same time created *explanations* telling the user why the recommendation link was selected in this context. In [2] explanations for recommended art works can be obtained based on a user profile of interest and features of the artworks. In [26] complex ontological models of processes and stories in the society, such as fishing or slash farming or the structure of narrative stories, were used as a basis for creating recommendation links between cultural resources. Still another approach to create recommendation links is to use similarity measures of event-based annotations [39].

Relational Search

Semantic recommending is related to relational search, where the idea is to try to search and discover serendipitous semantic associations between different content items [42, 41, 21]. The idea is to make it possible for the end-user to formulate queries such as "How is *X* related to *Y*" by selecting the end-point resources and the result is a set of semantic connection paths between *X* and *Y*. For example, in figure 2 the user has specified two historical persons, the Finnish artist Akseli Gallen-Kallela (1865–1931) and the French emperor Napoleon I (1769–1821) in a prototype of the portal CULTURESAMPO [21]. The underlying knowledge base contains an ontologized version of the ULAN vocabulary with some 100,000 persons and organizations, and semantic autocompletion based on an ONKI ontology server is used for finding the right query resources. The system has discovered an association between the persons based of eight patronOf, teacherOf, knows, and studentOf properties.

Personalization and Context Awareness

In many occasions the functioning of a semantic portal should not be static but adapt dynamically according to the 1) personal profile of the end-user and 2) the context of usage, especially the time and location [44].

Visitors in semantic cultural portals, like in physical museums, are usually not interested in everything found in the underlying collections, and would like to get information at different levels of detail. An important aspect of a semantic cultural portal is then adaptation of the portal to different personal information needs and

³⁵ <http://www.swi-prolog.org/>

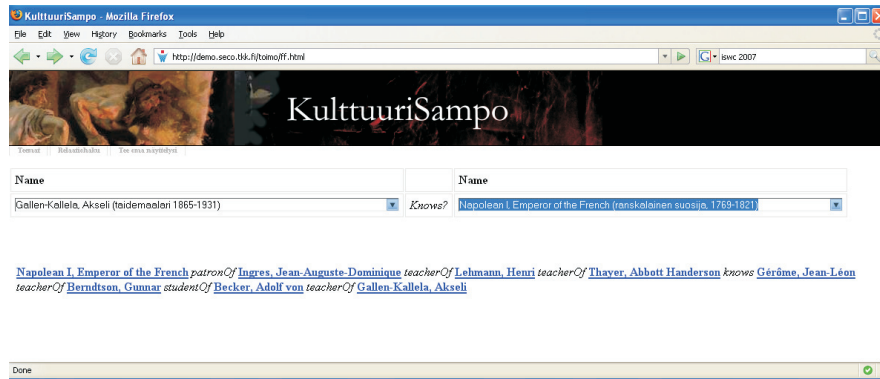


Fig. 2. An example of relational search in CULTURESAMPO [21] using the ULAN vocabulary. A serendipitous semantic association between two persons, who lived at different times in different countries and represented different occupations, is discovered.

interests. An example of a personalized cultural semantic portal is [2], where user profiling and personalization is based on metadata obtained by asking the users about her interests by rating pieces of artworks.

An example of location-based adaptability is the mobile phone user interface of the MUSEUMFINLAND portal [20]. It is possible to connect the system of a geolocation service providing the coordinate information of the phone. By pushing a special button on the interface, collection artifacts either manufactures or used nearby can be retrieved. It can be envisioned that new kinds of context sensitive services will be available in future portals based on modern phones supporting GPS positioning and radio-frequency identification (RFID) tags.

Also time is an important parameter for contextualizing portal services. For example, recommending the end-user to visit a beach during winter may not be wise, or a museum when it happens to be closed.

Visualization and Mash-ups

Visualization is an important aspect of the semantic web dealing with semantically complicated and interlinked contents. In the cultural heritage domain, e.g. maps, time lines, and methods for visualizing complicated and large semantic networks are of special interest.

Maps are useful in both searching content and in visualizing the results. A simple approach to using maps in cultural portals is to use a mash-up map-services. For example, there is a Google Maps³⁶ mash-up service³⁷ available for [20], showing the places, such as cities and villages, of the underlying location ontology on the map as interactive buttons. By selecting one of them, a query is executed by which all

³⁶ <http://maps.google.com/>

³⁷ Available online at <http://users.tkk.fi/tomik/geo/karttahaku.html>.

items manufactured or used in the selected place are retrieved. At the same time, additional search links to seven different traditional portals are shown. For example, by selecting the Wikipedia link, an article about the location (if available) is opened.

In the cultural heritage domain, historical maps are of interest of their own. For example, they depict old place names and borders not available anymore in contemporary maps. To visualize the historical changes, Temp-O-Map [29] makes it possible to lay old maps semi-transparently on top of the contemporary maps and satellite images of Google Maps [28]. To demonstrate the idea, the Karelia region of Finland was selected as a test case. This region was annexed to Soviet-Union as a result of the Second World War, after which most old Finnish place names in the region were changed into Russian ones making it difficult to the end-user to bridge the semantic gap between old and new names. The system is connected into an ontology modeling over 1000 historical regional changes of Finnish municipalities in 1860-2007. Historical municipalities of different time periods are available as facets for finding historical places on the maps. By selecting a category, the tool focuses the map view to the center point of the region.

Another important dimension for visualizing cultural content is time. A standard approach for temporal visualization is to use here a time line and project search objects on it. A generic mash-up tool for creating time lines is the Simile time line³⁸ used e.g. in [21].

4.1 Cross-portal Re-use of Content

The content of semantic portals can re-used easily in other web applications and portals due to the open semantic web standards. Re-using semantic content in this way is a generalization of the idea of “multi-channel publication” of XML, where a single syntactic structure can be rendered in different ways. In a similar vein, semantic metadata can be re-used without modifying it through *multi-application publication*.

One possibility to facilitate cross-portal re-use is to merge triple stores, and provide services to end-users based on the extended knowledge base. This approach was used e.g. in the Orava portal [27]. Another way of re-using content is to keep the portals separate and publish their functionalities as web services to be used by other semantic portals. Both traditional web services or light-weight mash-ups based on the REST principle can be used [33]. Here portal functionalities can be used in other portals on the HTML user interface level with just a pair of additional Javascript code added. For example, the semantic search functionality of MUSEUMFINLAND can be used as an additional information window, a “floatlet”, on another portal’s page [33]. If the page contains e.g. a video about skating, then the floatlet can show dynamically, using AJAX, images and semantic links to skates and related objects in the museum collections. The approach is related to the idea of using Google AdSense³⁹ advertisements, but generalized on a semantic level.

³⁸ <http://simile.mit.edu/timeline/>

³⁹ <http://www.google.com/adsense/>

5 Conclusions

Cultural heritage provides semantic web research with a semantically rich application domain in which useful vocabularies and collection contents are available, and where the organizations are eager to make their content easily and publicly accessible. A major application type in the area has been semantic portals, often aggregating content from different organizations, thus providing cultural organizations with a shared cost-effective publication channel and the possibility of enriching collaboratively the contents of each other's collections. For the end-user, new kinds of intelligent semantic services and ways of visualizing content can be provided. We envision that in the near future ever larger cultural semantic portals crossing geographical, cultural, and linguistic barriers of content providers at different countries will be developed. Also more systems for enriching the content by end-user created content in the spirit of Web 2.0 will be seen.

A major practical hinder for publishing cultural content on the semantic web is that current legacy cataloging system do not support creation of ontology-based annotations. If semantic annotations cannot be created in memory organization when cataloging content, then costly manual work is needed when transforming and disambiguating literal legacy metadata into ontological references in semantic portals. A practical approach to solve this fundamental problem is to provide ontologies as publicly available ontology services [23], and to re-use them as ready-to-use functionalities in legacy systems using mash-up techniques.

Acknowledgements

This work was supported the national semantic web ontology project FinnONTO⁴⁰ 2003-2007 in Finland, funded by the National Funding Agency for Technology and Innovation (Tekes) and a consortium of 36 public organizations and companies.

References

1. J. Aitchison, A. Gilchrist, and D. Bawden. *Thesaurus construction and use: a practical manual*. Europa Publications, London, 2000.
2. L. Aroyo, R. Brussee, L. Rutledge, P. Gorgels, N. Stash, and Y. Wang. Personalized museum experience: The Rijksmuseum use case. In *Museums and the Web 2007: Proceedings*, April 2007.
3. Ricardo Baeza-Yates and Berthier Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley, 1999.
4. Tim Berners-Lee, Yuhsin Chen, Lydia Chilton, Dan Connolly, James Hollenbach Ruth Dhanaraj, Adam Lerer, and David Sheets. Tabulator: Exploring and analyzing linked data on the semantic web. In *The 3rd International Semantic Web User Interaction Workshop (SWUI 2006)*, Nov 2006.

⁴⁰ <http://www.seco.tkk.fi/projects/finnonto/>

5. W. Buntine, K. Valtonen, and M. Taylor. The ALVIS document model for a semantic search engine. In *The 2nd Annual European Semantic Web Conference, Demos and Porters*, May 2005.
6. Robin Burke. Knowledge-based recommender systems. In A. Kent, editor, *Encyclopedia of Library and Information Systems*, volume 69, Supplement 32. Marcel Dekker, New York, 2000.
7. Stefan Decker, Michael Erdmann, Dieter Fensel, and Rudi Studer. Ontobroker: Ontology based access to distributed and semi-structured information. In R. Meersman et al., editor, *DS-8: Semantic Issues in Multimedia Systems*, pages 351–369. Kluwer, New York, 1999.
8. M. Doerr. The CIDOC CRM—an ontological approach to semantic interoperability of metadata. *AI Magazine*, 24(3):75–92, 2003.
9. J. English, M. Hearst, R. Sinha, K. Swearingen, and K.-P. Lee. Flexible search and navigation using faceted metadata. Technical report, University of Berkeley, School of Information Management and Systems, 2003.
10. C. Fellbaum, editor. *WordNet. An electronic lexical database*. The MIT Press, Cambridge, Massachusetts, 2001.
11. D. Fensel. *Ontologies: Silver bullet for knowledge management and electronic commerce (2nd Edition)*. Springer-Verlag, 2004.
12. D. J. Foskett. Thesaurus. In *Encyclopaedia of Library and Information Science, Volume 30*, pages 416–462. Marcel Dekker, New York, 1980.
13. C. Goble, S. Bechhofer, L. Carr, D. De Roure, and W. Hall. Conceptual open hypermedia = the semantic web? In *Proceedings of the WWW2001, Semantic Web Workshop, Hongkong*, 2001.
14. A. Hameed, A. Preese, and D. Sleeman. Ontology reconciliation. In S. Staab and R. Studer, editors, *Handbook on ontologies*, pages 231–250.
15. M. Hearst, A. Elliott, J. English, R. Sinha, K. Swearingen, and K.-P. Lee. Finding the flow in web site search. *CACM*, 45(9):42–49, 2002.
16. Jonathan H. Herlocker, Joseph A. Konstan, and John Riedl. Explaining collaborative filtering recommendations. In *Computer Supported Cooperative Work*, pages 241–250. ACM, 2000.
17. Markus Holi and Eero Hyvönen. Fuzzy view-based semantic search. In *Proceedings of the 1st Asian Semantic Web Conference (ASWC2006), Beijing, China*. Springer-Verlag, Berlin, September 3–7 2006.
18. Laura Hollink. *Semantic annotation for retrieval of visual resources*. PhD thesis, Free University of Amsterdam, 2006. SIKS Dissertation Series, No. 2006-24.
19. E. Hyvönen and E. Mäkelä. Semantic autocompletion. In *Proceedings of the first Asia Semantic Web Conference (ASWC 2006), Beijing*. Springer-Verlag, New York, August 4–9 2006.
20. E. Hyvönen, E. Mäkelä, M. Salminen, A. Valo, K. Viljanen, S. Saarela, M. Junnila, and S. Kettula. MuseumFinland—Finnish museums on the semantic web. *Journal of Web Semantics*, 3(2):224–241, 2005.
21. E. Hyvönen, T. Ruotsalo, T. Häggström, M. Salminen, M. Junnila, M. Virkkil, M. Haaramo, T. Kauppinen, E. Mäkelä, and K. Viljanen. CultureSampo—Finnish culture on the semantic web. The vision and first results. In *Semantic Web at Work—Proceedings of STeP 2006. To appear in: Klaus Robering (Ed.), Information Technology for the Virtual Museum. LIT Verlag, 2007, Nov 2006*.
22. E. Hyvönen, M. Salminen, S. Kettula, and M. Junnila. A content creation process for the Semantic Web, 2004. Proceeding of OntoLex 2004: Ontologies and Lexical Resources in Distributed Environments, May 29, Lisbon, Portugal.

23. E. Hyvönen, K. Viljanen, E. Mäkelä, T. Kauppinen, T. Ruotsalo, O. Valkeapää, K. Seppälä, O. Suominen, O. Alm, R. Lindroos, T. Käsälä, R. Henriksson, M. Frosterus, J. Tuominen, R. Sinkkilä, and J. Kurki. Elements of a national semantic web infrastructure—case study Finland on the semantic web (invited paper). In *Proceedings of the First International Semantic Computing Conference (IEEE ICSC 2007)*, Irvine, California, September 2007. IEEE Press, forth-coming.
24. Eero Hyvönen, Sampa Saarela, and Kim Viljanen. Application of ontology techniques to view-based semantic search and browsing. In *The Semantic Web: Research and Applications. Proceedings of the First European Semantic Web Symposium (ESWS 2004)*, 2004.
25. Eero Hyvönen, Arttu Valo, Kim Viljanen, and Markus Holi. A logic-based semantic web html generator—a poor man’s publishing approach. In *Proceedings of WWW2004, New York, Alternate Track Papers and Posters*, May 2004.
26. M. Junnila, E. Hyvönen, and M. Salminen. Describing and linking cultural semantic content by using situations and actions. In *Semantic Web at Work—Proceedings of STeP 2006. To appear in: Klaus Robering (Ed.), Information Technology for the Virtual Museum. LIT Verlag, 2007*, Nov 2006.
27. T. Käsälä and E. Hyvönen. A semantic view-based portal utilizing Learning Object Metadata, August 2006. 1st Asian Semantic Web Conference (ASWC2006), Semantic Web Applications and Tools Workshop.
28. T. Kauppinen, R. Henriksson, J. Väättäinen, C. Deichstetter, and E. Hyvönen. Ontology-based modeling and visualization of cultural spatio-temporal knowledge. In *Semantic Web at Work—Proceedings of STeP 2006*, Nov 2006.
29. Tomi Kauppinen, Christine Deichstetter, and Eero Hyvönen. Temp-o-map: Ontology-based search and visualization of spatio-temporal maps. In *Demo track at the European Semantic Web Conference ESWC 2007, Innsbruck, Austria, June 4–5 2007*. <http://www.seco.hut.fi/publications/>.
30. Michael S. Lew, Nicu Sebe, Chabane Djeraba, and Ramesh Jain. Content-based multimedia information retrieval: state of the art and challenges. *ACM Transactions on Multimedia computing, communications, and applications*, pages 1–19, Feb 2006.
31. A. Maedche, S. Staab, N. Stojanovic, R. Struder, and Y. Sure. Semantic portal—the SEAL approach. Technical report, Institute AIFB, University of Karlsruhe, Germany, 2001.
32. E. Mäkelä, E. Hyvönen, and S. Saarela. Ontogator—a semantic view-based search engine service for web applications. In *Proceedings of the 5th International Semantic Web Conference (ISWC 2006)*, Nov 2006.
33. Eetu Mäkelä, Kim Viljanen, Olli Alm, Jouni Tuominen, Onni Valkeapää, Tomi Kauppinen, Jussi Kurki, Reetta Sinkkilä, Teppo Käsälä, Robin Lindroos, Osmo Suominen, Tuukka Ruotsalo, and Eero Hyvönen. Enabling the semantic web with ready-to-use mash-up components, May 2007. Submitted for review.
34. Amanda Maple. Faceted access: A review of the literature. Technical report, Working Group on Faceted Access to Music, Music Library Association, 1995.
35. A. S. Pollitt. The key role of classification and indexing in view-based searching. Technical report, University of Huddersfield, UK, 1998. <http://www.ifla.org/IV/ifla63/63polst.pdf>.
36. Dave Reynolds, Paul Shabajee, and Steve Cayzer. Semantic Information Portals. In *Proceedings of the 13th International World Wide Web Conference on Alternate track papers & posters*, New York, NY, USA, May 2004. ACM Press.
37. Y. Rui, T. Huang, and S. Chang. Image retrieval: current techniques, promising directions and open issues. *Journal of Visual Communication and Image Representation*, 10(4):39–62, April 1999.

38. Tuukka Ruotsalo and Eero Hyvönen. An event-based method for making heterogeneous metadata schemas and annotations semantically interoperable. Helsinki University of Technology, 2007. Submitted for review, <http://www.seco.tkk.fi/publications>.
39. Tuukka Ruotsalo and Eero Hyvönen. A method for determining ontology-based semantic relevance. In *Proceedings of the International Conference on Database and Expert Systems Applications DEXA 2007, Regensburg, Germany*. Springer, September 3–7 2007.
40. G. M. Sacco. Dynamic taxonomies: guided interactive diagnostic assistance. In N. Wickramasinghe, editor, *Encyclopedia of Healthcare Information Systems*. Idea Group, 2005.
41. G. Schreiber, A. Amin, M. van Assem, V. de Boer, L. Hardman, M. Hildebrand, L. Hollink, Z. Huang, J. van Kersen, M. de Niet, B. Omelayenko, J. van Ossenbruggen, R. Siebes, J. Taekema, J. Wielemaker, and B. Wielinga. MultimediaN E-Culture3 demonstrator. In *Proceedings of the Fifth International Semantic Web Conference (ISWC 2006)*, pages 951–958. Springer-Verlag, Berlin, November 2006.
42. Amit Sheth, Boanerges Aleman-Meza, I. Budak Arpinar, Clemens Bertram, Yashodhan Warke, Cartic Ramakrishnan, Chris Halaschek, Kemafor Anyanwu, David Avant, F. Sena Arpinar, and Krys Kochut. Semantic association identification and knowledge discovery for national security applications. *Journal of Database Management on Database Technology*, 16(1):33–53, Jan–March 2005.
43. T. Sidoroff and Hyvönen. Semantic e-government portals—a case study. In *Proceedings of the ISWC-2005 Workshop Semantic Web Case Studies and Best Practices for eBusiness SWCASE05*, Nov 2005.
44. Spiros Sirmakessis, editor. *Adaptive and Personalized Semantic Web*. Springer-Verlag, Berlin, 2006.
45. J. Sowa. *Knowledge Representation. Logical, Philosophical, and Computational Foundations*. Brooks/Cole, 2000.
46. S. Staab, J. Angele, S. Decker, M. Erdmann, A. Hotho, A. Maedche, H.-P. Schnurr, R. Studer, and Y. Sure. Semantic Community Web Portals. In *Proceedings of the 9th International World Wide Web Conference*, Amsterdam, The Netherlands, May 2000. Elsevier.
47. Osmo Suominen, Kim Viljanen, and Eero Hyvönen. User-centric faceted search for semantic portals. In *Proceedings of the European Semantic Web Conference ESWC 2007, Innsbruck, Austria*, pages 356–370. Springer-Verlag, Berlin, June 4–5 2007.
48. M. van Assem, V. Malaise, A. Miles, and G. Schreiber. A method to convert thesauri to skos. In *Proceedings of the Third European Semantic Web Conference (ESWC'06)*, Lecture Notes in Computer Science, 2006.
49. M. van Assem, M. R. Menken, G. Schreiber, J. Wielemaker, and B. Wielinga. A method for converting thesauri to RDF/OWL. In *Proceedings of ISWC 2004, Hiroshima, Japan*. Springer-Verlag, Berlin, October 2004.
50. J. van Ossenbruggen, A. Amin, L. Hardman, M. Hildebrand, M. van Assem, B. Omelayenko, G. Schreiber, A. Tordai, V. de Boer, B. Wielinga, J. Wielemaker, M. de Niet, J. Taekema, M.-F. van Orsouw, and A. Teasing. Searching and Annotating Virtual Heritage Collections with Semantic-Web Techniques. In *Proceedings of Museums and the Web 2007*, San Francisco, California, March 2007. Archives & Museum Informatics.
51. Michiel Hildebrand Jacco van Ossenbruggen and Lynda Hardman. /facet: A browser for heterogeneous semantic web repositories. In *Proceedings of the ISWC 2006*. Springer-Verlag, Berlin, Nov 2006.
52. K. Viljanen, T. Känslä, E. Hyvönen, and E. Mäkelä. ONTODELLA—a projection and linking service for semantic web applications. In *Proceedings of the 17th International Conference on Database and Expert Systems Applications (DEXA 2006), Krakow, Poland*. IEEE, September 4–8 2006.

Index

- AAT, 5
- annotation ontology, 3
- annotation schema, 3
- automated link maintenance, 2

- CBIR, 12
- CCO, 5
- CIDOC CRM, 5, 6
- common sense rules, 7
- community portal, 1
- content aggregation, 1
- content model, 2, 3
- content-based information retrieval, 12
- cultural heritage, semantic portal, 2
- CultureSampo, 6

- data value standard, 5
- dereferencing, 6
- distributed content creation, 2
- DOLCE, 6
- Dublin Core, 3
- dumb-down principle, 3
- dynamic taxonomies, 12

- encoding guideline, 3

- faceted search, 2, 12
- FinnONTO, 17
- free indexing, 8, 10

- GPS, 15

- IEEE SUMO, 6
- information portal, 1

- intelligent user interface, 2

- layer cake model, 3
- logic level, 3

- MDA, 5
- meta-search, 4
- metadata level, 3
- metadata schema, 3
- MIR, 12
- multi-search, 4
- multimedia information retrieval, 12
- MuseumFinland, 5

- ONKI Ontology Server, 11
- ontology level, 3
- ontology library service, 10
- ontology population, 8

- precision, 11

- recall, 11
- recommendation, 2
- recommendation systems, 14
- RFID, 15
- rules, 7

- semantic autocompletion, 2
- semantic browsing, 2, 13
- semantic enriching, 2
- semantic gap, 12
- semantic interoperability, 10
- semantic search, 2
- semantic visualization, 2

service portal, 1

SKOS, 5

syntactic interoperability, 4, 9

TGM I, 5

TGN, 5

thematic role, 7

trust level, 3

Turtle notation, 4

ULAN, 5, 10

view-based search, 12

VRA, 4

Web 2.0, 3

Author Index

Hyvönen, Eero, 1