Publishing and Using Cultural Heritage Linked Data on the Semantic Web

Synthesis Lectures on Semantic Web: Theory and Technology

Editors

James Hendler, *Rensselaer Polytechnic Institute*
Ying Ding, *Indiana University*

Synthesis Lectures on the Semantic Web: Theory and Application is edited by James Hendler of Rensselaer Polytechnic Institute. Whether you call it the Semantic Web, Linked Data, or Web 3.0, a new generation of Web technologies is offering major advances in the evolution of the World Wide Web. As the first generation of this technology transitions out of the laboratory, new research is exploring how the growing Web of Data will change our world. While topics such as ontology-building and logics remain vital, new areas such as the use of semantics in Web search, the linking and use of open data on the Web, and future applications that will be supported by these technologies are becoming important research areas in their own right. Whether they be scientists, engineers or practitioners, Web users increasingly need to understand not just the new technologies of the Semantic Web, but to understand the principles by which those technologies work, and the best practices for assembling systems that integrate the different languages, resources, and functionalities that will be important in keeping the Web the rapidly expanding, and constantly changing, information space that has changed our lives.
Topics to be included:

- Semantic Web Principles from linked-data to ontology design

- Key Semantic Web technologies and algorithms

- Semantic Search and language technologies

- The Emerging "Web of Data" and its use in industry, government and university applications

- Trust, Social networking and collaboration technologies for the Semantic Web

- The economics of Semantic Web application adoption and use

- Publishing and Science on the Semantic Web

- Semantic Web in health care and life sciences

Publishing and Using Cultural Heritage Linked Data on the Semantic Web
Eero Hyvönen
2012

VIVO: A Semantic Approach to Scholarly Networking and Discovery
Katy Börner, Mike Conlon, Jon Corson-Rikert, and Ying Ding
2012

Linked Data: Evolving the Web into a Global Data Space
Tom Heath and Christian Bizer
2011

Publishing and Using Cultural Heritage Linked Data on the Semantic Web

Eero Hyvönen

www.morganclaypool.com

Publishing and Using Cultural Heritage Linked Data on the Semantic Web

Eero Hyvönen

Aalto University

MORGAN & CLAYPOOL PUBLISHERS

ABSTRACT

Cultural Heritage (CH) data is syntactically and semantically heterogeneous, multilingual, semantically rich, and highly interlinked. It is produced in a distributed, open fashion by museums, libraries, archives, and media organizations, as well as individual persons. Managing publication of such richness and variation of content on the Web, and at the same time supporting distributed, interoperable content creation processes, poses challenges where traditional publication approaches need to be re-thought. Application of the principles and technologies of Linked Data and the Semantic Web is a new, promising approach to address these problems. The development is leading to the creation of large national and international CH portals, such as Europeana, to large open data repositories, such as the Linked Open Data Cloud, and massive publications of linked library data in the U.S., Europe, and Asia. Cultural Heritage has become one of the most successful application domains of Linked Data and Semantic Web technologies.

This textbook gives an overview on why, when, and how Linked (Open) Data and Semantic Web technologies can be employed in practice in publishing CH collections and other contents on the Web. The text first motivates and presents a general semantic portal model and publishing framework as a solution approach to distributed semantic content creation, based on an ontology infrastructure. On the Semantic Web, such an infrastructure includes shared metadata models, ontologies, and logical reasoning, and is supported by shared ontology and other Web services alleviating the use of the new technology and linked data in legacy cataloging systems. The goal of all this is to provide layman users and researchers with new, more intelligent and usable Web applications that can be utilized by other Web applications, too, via well-defined Application Programming Interfaces (API). At the same time, it is possible to provide publishing organizations with more cost-efficient solutions for content creation and publication.

This book is targeted to computer scientists, museum curators, librarians, archivists, and other CH professionals interested in Linked Data and CH applications on the Semantic Web. The text is focused on practice and applications, making it suitable to students, researchers, and practitioners developing Web services and applications of CH, as well as to CH managers willing to understand the technical issues and challenges involved in linked data publication.

KEYWORDS

Semantic Web, linked data, cultural heritage, portal, metadata, ontologies, logic rules, information retrieval, semantic search, recommender system

Contents

Preface

Publishing Cultural Heritage (CH) collections and other content on the Web has become one of the most successful application domains of Semantic Web and Linked Data technologies. After a period of technical research and prototype development, boosted by the W3C Semantic Web Activity kick-off in 2001 and the Linked (Open) Data movement later on, major national and international CH institutions and collaboration networks have now started to publish their data using Linked Data principles and Semantic Web technologies.

This work is highly interdisciplinary, involving domain expertise of museum curators, librarians, archivists, and researchers of cultural heritage, as well as technical expertise of computer scientists and Web designers. Applying a new technology in the rapidly evolving Web environment is challenging not only for non-technical personnel in CH institutions, but also for computer scientists themselves.

This book aims at fostering the application of Linked Data and Semantic Web technologies in the CH domain by providing an overview of this fascinating application domain of semantic computing. My own work in this field started in 2001 after the W3C Semantic Web Activity launch by establishing the Semantic Computing Research Group (SeCo) focusing on this field. We first developed a semantic photograph search and recommender system for a university museum, followed by semantic portal prototypes for publishing heterogeneous collections of different kinds, including artifacts in cultural history museums, historical events, folklore, maps, fiction literature, and natural history museum data. This book reflects experiences gained during this work.

From the very beginning in 2002, after developing our first ontologies and transforming the first collection databases into RDF, it became clear that the possibility of reusing existing data, metadata models, and ontologies, and linking it all together in an interoperable way, will be a central benefit of Semantic Web applications. W3C recommendations, such as RDF(S), SKOS, SPARQL, and OWL are the corner stones for facilitating cross-domain, domain-independent interoperability, but this is not enough. We also need domain-dependent metadata-models and domain ontologies based on the generic semantic principles, as well as domain specific datasets. From a practical viewpoint, we also need ontology services so that the shared resources can be published and used in legacy and other application systems in a cost-efficient way. In short, a Semantic Web *content infrastructure* needs to be built in a similar vein as railroad, telephone, and other communication networks were created during earlier technological breakthroughs.

Creating a Semantic Web infrastructure, as well as content for it, requires collaboration between content providers. Co-operation is needed not only for sharing data through joint portals such as Europeana, but also for developing shared metadata models and ontologies used in representing the contents in an interoperable way. Publishing CH content is becoming a game of cross-domain

networking where the traditional boundaries of memory organizations based on content types are breaking down. From a user's viewpoint, the focus is on data, knowledge, and experience, be it based on a book in a library, an artifact in a cultural history museum, a story in an archive, a painting in an art gallery, a photograph taken by a fellow citizen, or a piece of music on a record.

During these years my faith in Semantic Web and Linked Data has become strong even if there are great challenges ahead, too. This is a truly promising way for providing richer content to users through more intelligent and usable interfaces, and at the same time for facilitating memory organization with better tools for collaborative, open content publishing on the Semantic Web.

Eero Hyvönen
October 2012

Acknowledgments

Thanks to the series editors Jim Hendler and Ying Ding for the invitation to write this book, and to Mike Morgan for making the publication possible.

The book's contents are based on collaboration with various students, researchers, and visitors in the Semantic Computing Research Group (SeCo) at the Aalto University and University of Helsinki in different times including (in alphabetical order) Matias Frosterus, Harri Hämäläinen, Tomi Kauppinen, Suvi Kettula, Heini Kuittinen, Jussi Kurki, Nina Laurenne, Aleksi Lindblad, Thea Lindquist, Glauco Mantegari, Eetu Mäkelä, Panu Paakkarinen, Tuomas Palonen, Sini Pessala, Tuukka Ruotsalo, Samppa Saarela, Katri Seppälä, Osma Suominen, Jouni Tuominen, Juha Törnroos, Mika Wahlroos, Mark van Assem, and Kim Viljanen.

Ying Ding, Stefan Gradmann, Patrick Leboeuf, Glauco Mantegari, Katri Seppälä, and Regine Stein made fruitful comments to earlier versions of this manuscript. Special thanks to Jouni Tuominen for several comments, suggestions, and help in proofreading the text. C.L. Tondo's help was invaluable in finalizing the text and layout.

Fruitful collaboration with several museums, libraries, archives, and media organizations in Finland is acknowledged, including (in alphabetical order) Agricola.fi network of historians, Antikvaria Museum Group, Espoo City Museum, Finnish Agriculture Museum, Finnish Broadcasting Company YLE, Finnish Literature Society, Finnish Museum of Photography, Finnish Museum Association, Finnish National Gallery, Finnish Public Libraries (Libraries.fi), Helsinki City Library, Helsinki University Library, Helsinki University Museum, Lahti City Museum, National Board of Antiquities, National Library of Finland, and Suomenlinna Sea Fortress.

The National Funding Agency for Technology and Innovation (Tekes)[1] and consortia of tens of public organizations and companies have supported several research projects of SeCo related to CH, such as Intelligent Catalogs (2002–2004), FinnONTO[2] (2003–2012), Semantic Ubiquitous Services (2009–2012)[3], and Linked Data Finland[4] (2012–). The Finnish Cultural Foundation[5] has supported our research on the CultureSampo system, too.

Thanks to SmartMuseum EU project[6] for funding and collaboration, to European Institute of Technology (EIT) Project EventMAP, as well as to the Network for Digital Methods in the Arts and Humanities (NeDiMAH) (European Science Foundation). Joint work with the University of Colorado regarding war history and linked data is acknowledged. Thanks to collaborations with the

---

[1] http://www.tekes.fi/en/
[2] http://www.seco.tkk.fi/projects/finnonto/
[3] http://www.seco.tkk.fi/projects/subi/
[4] http://www.seco.tkk.fi/projects/ldf/
[5] http://www.skr.fi/
[6] http://www.smartmuseum.eu/

Continuous Access to Cultural Heritage (CATCH) initiative and colleagues at the VU University and other universities in the Netherlands.

Eero Hyvönen
October 2012

CHAPTER 1

Cultural Heritage on the Semantic Web

Cultural Heritage (CH) refers to the legacy of physical objects, environment, traditions, and knowledge of a society that are inherited from the past, maintained and developed further in the present, and preserved (conserved) for the benefit of future generations[1]. This chapter first characterizes the notion of CH and identifies specific challenges encountered when publishing CH contents, especially collection data, on the Web. After this, Semantic Web and Linked Data technologies are introduced as a novel, promising approach to address the problems. The chapter ends with an overview of the book content.

1.1 CHARACTERIZING CULTURAL HERITAGE

CH can divided into three subareas.

1. **Tangible cultural heritage** consists of concrete cultural objects, such as artifacts, works of art, buildings, and books.

2. **Intangible cultural heritage** includes phenomena such as traditions, language, handicraft skills, folklore, and knowledge.

3. **Natural cultural heritage** consists of culturally significant landscapes, biodiversity, and geodiversity.

The key players in preserving CH are *memory organizations* that include libraries, archives, and museums of different kinds specializing in particular areas of CH, such as art museums, archaeological museums, botanical museums and gardens, cultural history museums, medical collections, science museums, theater history museums, geological and mineralogical museums, and zoology museums. Also media organizations often preserve CH materials, especially more recent ones. There are also lots of CH materials maintained by cultural associations of various kinds and individual persons. Tangible CH objects are stored with attached metadata, intangible heritage is documented using textual descriptions, photographs, interviews, and videos, and there are natural history and other museums specializing in storing traces and knowledge of natural history, geology, and environment.

[1]In this book, the ambiguous term "culture" is used to refer to the "the ideas, customs, skills, arts, etc. of a people or group, that are transferred, communicated, or passed along, as in or to succeeding generations" (Webster's New World Dictionary).

The Web has become an increasingly important medium for publishing CH contents of different kinds. For example, libraries and archives are online with their collections, museums show their collections through collection browsers, and documentation of intangible heritage is available as audio and video recordings and as interactive hypertext applications, even as games. There are large national and multi-national CH portal projects active in harvesting and publishing content from different sources via centralized services.

For the layman end-user, such systems provide a single access point to massive heterogeneous collections and an authoritative source of information. In contrast to traditional physical exhibitions, Web services are open all the time, can be accessed without physical presence at an exhibition, the number of exhibits on the Web is not limited by the physical space available, and the exhibits can be linked and accessed flexibly using different strategies, not only the one used in the physical exhibition. Of course, the Web cannot replace the physical experience of visiting a museum or an exhibition in reality but provides a complementary alternative for accessing collection data virtually at any time and from any place.

For researchers in the humanities, availability of CH data in massive amounts in digital machine processable form has opened up a new research paradigm called Digital Humanities.

1.2   INFORMATION PORTALS FOR CULTURAL HERITAGE

There are several kind of CH publications on the Web. First, there is a large variety of well-curated systems that have been hand-crafted for a specific purpose with a focused closed theme, dataset, and interfaces. Such systems are often implemented using tools such as Adobe Flash with a beautiful game-like appearance. For example, the Lewis and Clark Expedition (1803–1806) is documented on the Web in great detail by several applications. The portal in Figure 1.1[2] provides the end-user with several thematic perspectives to the journey by selecting the buttons on the left, such as "overview," "American nation," "geography," "journal excerpts," "natural history," and "technology used." Such systems may also be available on CD/DVD as stand-alone applications.

On the other end of the spectrum, there are collection search services and browsers providing access to large open collection databases whose content is not thematically focused, and curated access paths and interfaces may be missing. In return, large collection databases originating possibly from several institutions can be accessed. For example, a variety of Australian CH collections can be accessed using the Collections Australia Network system[3]. Similar federated portals for searching and browsing collections can be found in many countries and internationally. A flagship application here is Europeana[4], based on millions of collection objects originating from memory organizations all over Europe. For example, in Figure 1.2 the user has typed in the keyword "chair" in the search field of Europeana and the system has found various chairs in participating collections. The search can be refined further by selecting additional filters on the facets on the left, such as "media type,"

[2]http://lewis-clark.org/
[3]http://www.collectionsaustralia.net/
[4]http://www.europeana.eu/

Figure 1.1: A portal exhibiting versatile content related to the Lewis and Clark expedition (1803–1806) in the U.S. from different perspectives. (Fort Mandan Foundation, North Dakota)

"language," "date," "country," and whether content contributed by users should be included or not. Another portal example, harvesting library data, is WorldCat[5] that contains metadata (without the primary sources) of about 1.5 billion books, DVDs, CDs, and articles in the participating libraries. The World Digital Library[6] is yet another international portal, operated by UNESCO and the United States Library of Congress, that makes available, free of charge, significant multilingual primary materials, such as manuscripts, maps, rare books, musical scores, recordings, films, prints, photographs, and architectural drawings.

In this book, the main focus is on *information portal systems* of the latter kind: CH portals based on large heterogeneous collection datasets are considered, where organizing the contents by hand into a focused thematic application with application-specific visualizations and interfaces is not usually feasible. Such shared publication portals facilitate exchange of knowledge for CH researchers, librarians, and archivists. For the contributing memory organizations, such systems are

---

[5]http://www.worldcat.org/
[6]http://www.wdl.org/

Figure 1.2: Faceted search in Europeana portal exhibiting chairs from different European collections.

an opportunity to reach out to wider audiences on the Web with new ways of interaction, and to collaborate with other organizations. From a societal perspective, publishing CH on the Web stimulates cultural tourism, creative economy, and enhances friendly relationships and unity between parties and nations involved in such initiatives.

1.3   CHALLENGES OF CULTURAL HERITAGE DATA

CH collection data has many specific characteristic features, such as the following.

- **Multi-format**. The contents are presented in various forms, such as text documents, images, audio tracks, videos, collection items, and learning objects.

- **Multi-topical**. The contents concern various topics, such as art, history, artifacts, and traditions.

- **Multi-lingual**. The content is available in different languages.

- **Multi-cultural**. The content is related and interpreted in terms of different cultures, such as religions or national traditions in the West and East.

- **Multi-targeted**. The contents are often targeted to both laymen and experts, young and old.

As a result, a fundamental problem area in dealing with CH data is to make the content mutually *interoperable*, so that it can be searched, linked, and presented in a harmonized way across the boundaries of the datasets and data silos. The problem occurs on a syntactic level, e.g., when harmonizing different character sets, data formats, notations, and collection records used in different collections. Even more importantly, there is the problem of *semantic interoperability*: different metadata formats may be interpreted differently, data is encoded at different levels of precision, vocabularies and gazetteers used in describing the content are different, and so on. The Semantic Web standards[7] and best practices, especially those advocated by the World Wide Web Consortium (W3C)[8], provide a shared basis on which interoperable Web systems can be built in a well-defined manner. The new technologies are of course no panacea for all problems but rather a tool set by which the hard issues can be tackled arguably more effectively than before.

A major reason for interoperability problems in CH content publishing is the *multi-organizational* nature in which CH content is collected, maintained, and published. The content is provided by different museums, libraries, and archives with their own established standards and best practices, by media organizations, cultural associations, and individual citizens in a Web 2.0 fashion. The success of the WWW is very much due to its simple distributed many-to-many publishing paradigm that has few restrictions and shared standards, with the HTML mark-up language combined with the HTTP protocol and the idea of URL addressing as core technologies. However, things get more complicated on the Semantic Web, where content is not published only for human users in HTML form but also as data for machines to use. An additional standard base is needed for the Web of Data. In application domains such as CH more coordinated collaboration is needed between CH publishers and the technical WWW developer community than before.

1.4   PROMISES OF THE SEMANTIC WEB

Semantic Web technologies[9] [34] (SW) are a promising new approach for addressing the problems of publishing CH content on the Web. The term "semantic" here refers to *Semantics*, a discipline studying relations between *signifiers*, such as words, phrases, signs, and symbols, and what they stand for, i.e., *denotata*. In Computer Science semantics refers to the formal meaning and interpretation (declarative or procedural) that has been given to syntactic structures, such as programming languages or symbolic data structures.

---

[7]Called "recommendations" by the W3C.
[8]http://www.w3.org/
[9]http://www.w3.org/standards/semanticweb/

The Semantic Web can be seen as a new layer of *metadata* being build inside the Web. According to the traditional definition, metadata is data about data. For example, a metadata record of a book (data) may tell its title, author, subject, and publishing year. However, the term "metadata" is used more widely in the Semantic Web context as a synonym for machine processable or interpretable data. The key idea is that syntactic metadata structures make Web content "understandable" to the machines, based on shared semantic specifications founded on formal logic. This makes it possible to create more interoperable and intelligent Web services. A computer that cannot interpret the data it is dealing with is like a telephone just passing information, and cannot be very helpful in more complicated information processing tasks dealing with the meanings of the contents.
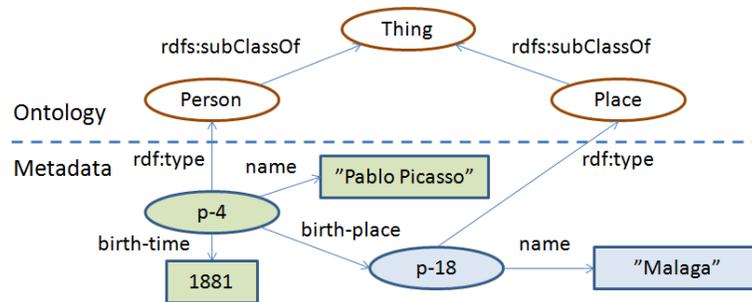
Figure 1.3: The data model of RDF is a directed labeled graph.

The methodology for representing metadata and ontological concepts[10] on the Web is based on a simple data model: a directed labeled graph, i.e., a *semantic net*. For example, Figure 1.3 depicts an RDF graph telling on a metadata level that the identity *p-4* is an individual of the class *Person* (denoted by the arc *rdf:type*) with name "Pablo Picasso" born in 1881 at an instance *p-18* of the class *Place* whose name is "Malaga." In the RDF graph, classes such as places and persons are represented as subclasses (arc *rdfs:subClassOf*) of the class *Thing* on an ontology level, while the individuals of the classes are considered metadata. Both metadata and ontologies are represented uniformly in the same graph. In the figure, identities that may have properties, i.e., may have out-going arcs, are depicted as ovals while literal terminal atomic values without further properties (here strings and numbers) as rectangular boxes.

The figure illustrates that actually there are several levels of descriptions needed on the Semantic Web.

1. **Real world**. On the bottom, there is the real world, i.e., the domain of discourse, such as persons, artifacts, and places.

---

[10]The notion of "concept" is a complex philosophical notion referring to a general idea or something conceived in the mind. On the Semantic Web, the term "concept" is used for any entity on the Web or outside of it with an identity specified by a URI.

2. **Data level**. Then there is the data level, since real world items have to be represented as data of some kind in a computer. For example, images and documents are data as well as a URI reference to a person.

3. **Metadata level**. After data, there is metadata about the data, e.g., records in a collection database about images, persons, or artifacts.

4. **Ontology level**. Next, ontology level defines the generic classes and properties used in describing a domain, i.e., the vocabularies in terms of which the metadata is represented. The metadata schema used in cataloging and controlled vocabularies of subject headings are part of this level. For example, in Figure 1.3 persons are described in terms of their name, birth time, and birth place, and instantiated from the classes defined on the ontology level. The same ontologies can be used for representing collection metadata of a similar domain area in different memory organizations (e.g., books in libraries).

5. **Metaontology level**. Finally, there are the general cross-domain modeling principles of ontologies that are domain-independent. For example, the notions of subclass-of relation and class are generic and not restricted to a particular domain. Such generic principles are specified by the Semantic Web standards, such as RDF(S) and OWL, and facilitate cross-domain interoperability of contents.

On a global WWW scale, the Semantic Web forms a *Giant Global Graph* (GGG) of connected data resources. The GGG can be used and browsed in ways analogous to the WWW, but while the WWW links associated Web pages with each other for human use, the GGG links associated underlying concepts and data resources together. For example, the GGG may tell that ducks are birds, and that Donald is an instance of a duck (and therefore a bird) while the related WWW pages may constitute a comics book about Donald Duck.

A key idea of linked data is that the different parts of the GGG can come from different data sources. For example, in Figure 1.3 metadata about persons, such as Pablo Picasso, may come from an authority database, information about places, such as Malaga, may be provided by a land survey organization, and the class ontology can be based on an existing keyword thesaurus in use in a library. Different data sources are illustrated in the figure by different colors/densities.

Based on harmonized RDF-based representations of data, more "intelligent" Web applications can be built and with less effort. From a technical application perspective, Semantic Web technologies have many promising features:

- *More accurate content descriptions*. The technology is based on globally unique Universal Resource Identifiers (URI), which makes it possible to refer to meanings more accurately than using literal expressions. For example, person and place names can be disambiguated: there are lots of "John Smiths" around, "Paris" can be found in France, Texas, and in many other places, and the names can have different transliterations in different language systems. In libraries, the notion of, e.g., Shakespeare's play "Hamlet" can refer to the abstract story, its manifestation

as a text or a video of the play, different translations of it, variants of the story, editions of these, and finally individual books or DVDs on the library shelves. Modeling such semantic distinctions can be done using novel "ontology-based" CH standards to be presented in this book.

- *Interoperability*. Semantic Web technologies provide a novel approach to creating interoperable linked data.

- *Simple data model for aggregation*. Two (interoperable) RDF graphs can be joined together technically in a trivial way by simply making the union of them (i.e., the corresponding triple sets).

- *Data aggregation by linked data*. By combining data sources in an interoperable way, data from one source can be enriched with additional linked data from another source. A notable international initiative toward this goal is Linked Data[11] [53], where open datasets such as Wikipedia/DBpedia[12] and Freebase[13] for common knowledge, GeoNames[14] for millions of place names, or Gutenberg project[15] for over 40,000 free ebooks are described in terms of Semantic Web standards and interlinked with each other.

- *Semantic Web services*. Semantic linked data is published not only as passive datasets, but as operational services than can be utilized by legacy and other CH applications via open and generic Application Programming Interfaces (API). By utilizing shared ready-made services, application programmers can re-use work done by others, and save their own programming effort and resources. This idea can be paralleled with Google and Yahoo! Maps that provide map services on a global basis to applications via easy-to-use APIs for mash-up development.

Publishing CH on the Web is not only a technical challenge; issues of trustworthiness of content, copyrights, and licensing are also of concern. Much of CH content is protected by copyright, and there are also other reasons why organizations cannot publish their data openly, e.g., issues of personal privacy. However, based on the ideas of Linked Open Data, the WWW world is clearly taking steps toward publishing open data and free of charge when feasible. The idea is that CH content should be maximally shared. It is also usually produced by public funding and in this sense already paid by the public. Free open data also fosters interoperability and creates a basis on which commercial applications can be built more easily. Trust and copyright issues are important, e.g., in Web 2.0 spirited social cultural portals, where end-users create, tag, and publish content of their own and the others'.

[11]http://linkeddata.org/
[12]http://www.dbpedia.org/
[13]http://www.freebase.com/
[14]http://www.geonames.org/
[15]http://www.gutenberg.org/

1.5   OUTLINE OF THE BOOK

This book is an introduction to publishing CH contents on the Semantic Web as Linked Data. The idea is to provide a kind of cook book on how to create semantic portals of CH, where heterogeneous content is produced by a multitude of distributed organizations, and is harvested, harmonized, validated, and published as a service for human and machine users.

The text starts (Chapter 2) with presenting a motivating "business model" for this prototypical semantic portal scenario that can be considered a kind of standard model for publishing CH on the Semantic Web. In Chapter 3 requirements for publishing Linked Data are considered. The Semantic Web is based on the "layer cake model" of W3C that adds new standards above the XML[16] standard family, the *lingua franca* of the Web.

- *Metadata level.* The RDF data model[17] is the basis of the Semantic Web and Linked Data, and is used for representing metadata as well as other forms of content on the Web of Data. Metadata models for CH data are considered in Chapter 4.

- *Ontology level.* The RDF Schema and the Web Ontology Language OWL[18] are used for representing ontologies that describe vocabularies and concepts concerning the real world and our conception of it. Domain vocabularies and ontologies for CH are in focus in Chapter 5.

- *Logic level.* Logic rules, to be discussed in Chapter 6, can be used for deriving new facts and knowledge based on the metadata and ontologies. This can be used, e.g., to minimize cataloging work, make searching and browsing more effective, and to find serendipitous semantic links between CH objects.

After presenting technical foundations and models, issues related to annotating and harvesting CH content for a portal are presented in Chapter 7. Chapter 8 discusses intelligent services based on semantic linked data. The book is finally concluded in Chapter 9.

1.6   BIBLIOGRAPHICAL AND HISTORICAL NOTES

The idea of the World Wide Web (WWW) was proposed first in 1989 by Tim Berners-Lee, and more formally with Robert Cailliau in 1990. History of the early WWW is documented in the book "Weaving the Web" [14]. Already in the early days of the WWW the idea of a "Semantic Web," i.e., a web of machine interpretable data, has been around. However, the first generation of the WWW was targeted to humans, and was based on three simple technologies for mediating Web pages between human users: HTML, HTTP, and URLs.

From a scientific viewpoint, the Semantic Web is based on results of Artificial Intelligence, where semantic networks and logic-based knowledge representation have been studied from the

---

[16] http://www.w3.org/XML/
[17] http://www.w3.org/RDF/
[18] http://www.w3.org/2004/OWL/

late 50's; see, e.g., [126] for a thorough overview of this field. The first Semantic Web standard in use, Resource Description Framework (RDF), was published by W3C already in 1999, only a year after the XML recommendation. As another approach for the Semantic Web, Topic Maps [114] has been developed and published as the ISO standard ISO/IEC 13250:2003[19]. This standard is intended for the representation and interchange of knowledge, with an emphasis on the findability of information. The system originated from the idea of creating semantic indexes for publications. However, Semantic Web development really got off using the W3C standard stack after the publication of the seminal article "The Semantic Web" [15] in *Scientific American,* and the launch of the Semantic Web Activity at W3C.

The semantic technology did not penetrate the market as quickly as many other Web developments, say XML. A reason for this is complexity of some standards and their foundations in logic not so familiar in mainstream computing. In around 2005, the ideas on Linked Data and Web of Data started to gain momentum as a simple approach to the Semantic Web focusing on publishing large existing datasets, and using only simple RDF and lightweight ontologies. Combined with idea of Open Data, the idea of the Semantic Web has been adopted especially by the public sector [158], and several national initiatives have been started in the U.K.[20], U.S[21], and in smaller countries, such as Finland [67].

A thorough overview of Linked Data and Web of Data is presented in [53]. Semantic Web and linked data standards and technology, with pointers to related research and applications, can be accessed at W3C Web pages[22], and at the home pages of the Linked Data community[23]. The W3C Linked Library Data Incubator Group has evaluated the current state of library data management, outlined the potential benefits of publishing library data as Linked Data, and formulated next-step recommendations for library standards bodies, data and systems designers, librarians and archivists, and library leadership in a final report[24]. Another report "Linked Data for Libraries, Museums, and Archives: Survey and Workshop Report" with related goals was published at the same date, based on a workshop at the Stanford University[25]. Major international Semantic Web conferences include the International Semantic Web Conference (ISWC) and Extended Semantic Web Conference (ESWC). The World Wide Web conference (WWW) is the main yearly event for general Web research with a W3C focus.

A wide variety of Web applications in the museum domain have been presented in the proceedings of the Museums and the Web conference series since 1997, with papers available online[26]. The International Federation of Library Associations and Institutions (IFLA)[27] organizes a large annual World Library and Information Congress for libraries, and the International Council on

---

[19]http://www.isotopicmaps.org/
[20]http://data.gov.uk/
[21]http://www.data.gov/
[22]http://www.w3.org/standards/semanticweb/
[23]http://linkeddata.org/
[24]http://www.w3.org/2005/Incubator/lld/
[25]http://www.clir.org/pubs/abstract/reports/pub152
[26]http://www.archimuse.com/conferences/mw.html
[27]http://www.ifla/

Archives (ICA)[28] has a similar annual congress series, International Conference of the Round Table on Archives (CITRA) for archivists.

The intersection of computing and the disciplines of the humanities are studied in the field of *Digital Humanities*, also called *Humanities Computing*. [105] The general goal here is to develop and apply computational methods in humanities research. Since 1990, the digital humanities community has been organizing the Digital Humanities conference series[29]. A major journal in the field is the *Digital Humanities Quarterly*[30].

---

[28]http://www.ica.org/
[29]http://digitalhumanities.org/conference
[30]http://www.digitalhumanities.org/dhq/

Bibliography

[1] J. Aitchison, A. Gilchrist, and D. Bawden. *Thesaurus construction and use: a practical manual*. Europa Publications, London, 2000. Cited on page(s) 60, 88

[2] David Allemang and Jim Hendler. *Semantic Web for the working Ontologiest*. Morgan Kaufmann, San Francisco, 2008. Cited on page(s) 77, 104

[3] David Allemang and Jim Hendler. *Speech and Language Processing, 2nd edition*. Prentice-Hall, Upper Saddle River, New Jersey, 2008. Cited on page(s) 110

[4] Grigoris Antoniu and Frank van Harmelen. *Semantic Web Primer*. The MIT Press, Cambridge, Massachusetts, 2008. Cited on page(s) 77, 83, 86

[5] L. Aroyo, R. Brussee, L. Rutledge, P. Gorgels, N. Stash, and Y. Wang. Personalized museum experience: The Rijksmuseum use case. In J. Trant and D. Bearman, editors, *Museums and the Web 2007: Proceedings*, pages 137–144, 2007. Cited on page(s) 38, 113, 118

[6] L. Aroyo, E. Hyvönen, and J. van Ossenbruggen, editors. *Cultural Heritage on the Semantic Web. Workshop Proceedings. The 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference, Busan, Korea*. ISWC + ASWC, 2007. http://www.cs.vu.nl/~laroyo/CH-SW.html. Cited on page(s) 20, 132

[7] L. Aroyo, N. Stash, Y. Wang, P. Gorgels, and L. Rutledge. CHIP demonstrator: Semantics-driven recommendations and museum tour generation. In *Proceedings of ISWC 2007 + ASWC 2007, Busan, Korea*, pages 879–886. Springer–Verlag, 2007. Cited on page(s) 20, 38, 113, 118, 119

[8] Murtha Baca, editor. *Introduction to metadata*. Getty Publications, Los Angeles, 2008. Cited on page(s) 35, 55, 126

[9] Murtha Baca and Patricia Harpring, editors. *Categories for description of works of art*. Getty Publications, Los Angeles, 2009. http://www.getty.edu/research/publications/electronic_publications/cdwa/index.html. Cited on page(s) 40

[10] Murtha Baca, Patricia Harpring, Elisa Lanzi, Linda McRae, and Ann Whiteside, editors. *Cataloging Cultural Objects. A guide to describing works and their images*. American Library Association, 2008. Cited on page(s) 40

124   BIBLIOGRAPHY

[11] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley, 1999. Cited on page(s) 119

[12] Chryssoula Bekiari, Martin Doerr, and Patrick Le Boeuf, editors. *FRBR object-oriented definition and mapping to FRBR(version 1.0)*. International Federation of Library Associations and Institutions (IFLA), 2009. `http://www.cidoc-crm.org/docs/frbr_oo/frbr_docs/FRBRoo_V1.0_2009_june_.pdf`. Cited on page(s) 49

[13] V. R. Benjamins, J. Contreras, M. Blázquez, J.M. Dodero, A. Garcia, E. Navas, F. Hernandez, and C. Wert. Cultural heritage and the semantic web. In *The Semantic Web: Research and Applications*, pages 433–444. Springer–Verlag, 2004. Cited on page(s) 20

[14] T. Berners-Lee, M. Fischetti, and M. Dertouzos. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web*. Harper Business, 2000. Cited on page(s) 9

[15] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, 284(5):34–43, May 2001. Cited on page(s) 10

[16] J. Bhogal, A. Macfarlane, and P. Smith. A review of ontology based query expansion. *Information Processing & Management*, 43(4):866–886, 2007. Cited on page(s) 73

[17] C. Bizer, T. Heath, and T. Berners-Lee. Linked data – the story so far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 2009. Cited on page(s) 33

[18] Christian Bizer and Richard Cyganiak. Quality-driven information filtering using the WIQA policy framework. *Journal of Semantic Web*, 7(1):1–10, 2009. Cited on page(s) 105

[19] C. Borgman and S. Siegfriend. Getty's synoname and its cousins: A survey of applications of personal name-matching algorithms. *Journal of the American Society for Information Science and Technology*, 43(7):459–476, 1992. Cited on page(s) 69

[20] J. P. Bowen and S. Filippini-Fantoni. Personalization and the web from a museum perspective. In *Selected Papers from an International Conference Museums and the Web 2004 (MW2004), Arlington, Virginia, USA*. Archieves & Museum Informatics, 2004. `http://www.museumsandtheweb.com/mw2004/papers/bowen/bowen.html`. Cited on page(s) 119

[21] Ronald Brachman and Hector Levesque. *Knowledge representation and reasoning*. Morgan Kaufmann, San Francisco, 2004. Cited on page(s) 35

[22] Ivan Bratko. *Prolog programming for Artificial Intelligence (4th edition)*. Pearson, Harlow, U.K., 2012. Cited on page(s) 86

[23] P. Buitelaar and P. Cimiano, editors. *Ontology Learning from Text: Methods, Evaluation and Applications*. Series information for Frontiers in Artificial Intelligence and Applications. IOS Press, Amsterdam, The Netherlands, 2005. Cited on page(s) 104

[24] P. Buitelaar and P. Cimiano, editors. *Ontology Learning and Population: Bridging the Gap between Text and Knowledge*. Series information for Frontiers in Artificial Intelligence and Applications. IOS Press, Amsterdam, The Netherlands, 2008. Cited on page(s) 104

[25] R. Burke. Knowledge-based recommender systems. In A. Kent, editor, *Encyclopedia of Library and Information Systems*, volume 69. Marcel Dekker, New York, 2000. Cited on page(s) 113

[26] E. Clementine, P. De Felice, and P. van Oosterom. A small set of formal topological relationships suitable for end-user interaction. In A. Abel and B. Chin Ooi, editors, *Advances in databases*, pages 277–295. Springer–Verlag, 1993. Cited on page(s) 73

[27] Erin Coburn, Richard Light, Gordon McKenna, Regine Stein, and Axel Vitzthum, editors. *LIDO – Lightweight Information Describing Objects. Version 1.0.* ICOM-CIDOC Working Group Data Harvesting and Interchange, 2010. Cited on page(s) 49

[28] Nick Crofts, Martin Doerr, Tony Gill, Stephen Stead, and Matthew Stiff (Eds.), editors. *Definition of the CIDOC Conceptual Reference Model, Version 5.0.4.* ICOM/CIDOC Documentation Standards Group (CIDOC CRM Special Interest Group), 2011. `http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.4.pdf`. Cited on page(s) 44

[29] Aba-Sah Dadzie and Matthew Rowe. Approaches to visualising linked data: A survey. *Semantic Web – Interoperability, Usability, Applicability*, 1(1–2), 2011. Cited on page(s) 119

[30] Mathieu d'Aquin and Holger Lewen. Cupboard – a place to expose your ontologies to applications and the community. In *Proceedings of the ESWC 2009*, pages 913–918. Springer–Verlag, 2009. Cited on page(s) 77

[31] Mathieu dÁquin and Enrico Motta. Watson, more than a semantic web search engine. *Semantic Web – Interoperability, Usability, Applicability*, 2(1):55–63, 2011. Cited on page(s) 32

[32] M. Doerr. The CIDOC CRM—an ontological approach to semantic interoperability of metadata. *AI Magazine*, 24(3):75–92, 2003. Cited on page(s) 44

[33] Martin Doerr. Ontologies for cultural heritage. In Staab and Studer [139], pages 463–486. Cited on page(s) 44

[34] John Dominque, Dieter Fensel, and James A. Hendler, editors. *Handbook of Semantic Web*. Springer–Verlag, 2011. Cited on page(s) 5, 77

[35] Martin Dzbor, Enrico Motta, and Laurian Gridinoc. Browsing and navigating in semantically rich spaces: Experiences with magpie application. In Staab and Studer [139], pages 687–709. Cited on page(s) 119

[36] 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. Cited on page(s) 112

126   BIBLIOGRAPHY

[37] J. Euzenat and P. Shvaiko. *Ontology Matching*. Springer–Verlag, 2007. Cited on page(s) 93, 104

[38] C. Fellbaum, editor. *WordNet. An electronic lexical database*. The MIT Press, Cambridge, Massachusetts, 2001. Cited on page(s) 59

[39] Tim Finin, Li Ding, Rong Pan, Anupam Joshi, Pranam Kolari, Akshay Java, and Yun Peng. Swoogle: Searching for knowledge on the semantic web. In *In AAAI 05 (intelligent systems demo*, pages 1682–1683. The MIT Press, Cambridge, Massachusetts, 2005. Cited on page(s) 32

[40] D. J. Foskett. Thesaurus. In *Encyclopaedia of Library and Information Science*, volume 30, pages 416–462. Marcel Dekker, New York, 1980. Cited on page(s) 88

[41] J. French, A. Powell, and E. Schulman. Using clustering strategies for creating authority files. *Journal of the American Society for Information Science*, 51(8):774–786, jun 2000. Cited on page(s) 69

[42] Matias Frosterus, Eero Hyvönen, and Mika Wahlroos. Extending ontologies with free keywords in a collaborative annotation environment. In *Proceedings of the ISWC 2011 Workshop Ontologies Come of Age in the Semantic Web (OCAS)*. CEUR Workshop Proceedings, Vol 809, http://ceur-ws.org, 2011. Cited on page(s) 96

[43] C. Galvez and F Moya-Anegon. Approximate personal name-matching through finite-state graphs. *Journal of the American Society for Information Science and Technology*, 58(13):1960–1976, 2007. Cited on page(s) 69

[44] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider. Sweetening ontologies with DOLCE. In *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2002)*. Springer–Verlag, 2002. Cited on page(s) 67, 91, 104

[45] V. Geroimenko and C. Chen, editors. *Visualizing the Semantic Web: XML–based Internet and Information Visualization*. Springer–Verlag, 2002. Cited on page(s) 119

[46] A. J. Gilliland. Setting the stage. In Baca [8], pages 1–19. Cited on page(s) 35

[47] N. Guarino and C. Welty. Evaluating ontological decisions with OntoClean. *Communications of the ACM*, 45(2):61–65, 2002. Cited on page(s) 91

[48] Nicola Guarino, Daniel Oberle, and Steffen Staab. What is an ontology? In Staab and Studer [139], pages 1–17. Cited on page(s) 62

[49] P. Haase, M. Schmidt, and A. Schwarte. The information workbench as a self-service platform for linked data applications. In *Proceedings of the 2nd International Workshop on Consuming Linked Data (COLD 2011)*, 2011. CEUR Workshop Proceedings, Vol 782, http://ceur-ws.org. Cited on page(s) 104

[50] A. Hameed, A. Preese, and D. Sleeman. Ontology reconciliation. In S. Staab and R. Studer, editors, *Handbook on ontologies*, pages 231–250. Springer–Verlag, 2004. Cited on page(s) 93

[51] Bernhard Haslhofer and Antoine Isaac. data.europeana.eu – the Europeana linked open data pilot. In *International Conference on Dublin Core and Metadata Applications (DC 2011)*, 2011. http://dcevents.dublincore.org/index.php/IntConf/dc-2011/paper/view/55/14. Cited on page(s) 43

[52] 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. Cited on page(s) 119

[53] T. Heath and C. Bizer. *Linked Data: Evolving the Web into a Global Data Space (1st edition)*. Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool, 2011. Cited on page(s) 8, 10, 33

[54] Dominik Heckmann, Tim Schwartz, Boris Brandherm, Michael Schmitz, and Margeritta von Wilamowitz-Moellendorff. Gumo – the general user model ontology. In Liliana Ardissono, Paul Brna, and Antonija Mitrovic, editors, *User Modeling*, pages 428–432. Springer–Verlag, 2005. Cited on page(s) 118

[55] Riikka Henriksson, Tomi Kauppinen, and Eero Hyvönen. Core geographical concepts: Case finnish geo-ontology. In *Location and the Web (LocWeb) 2008 workshop, 17th International World Wide Web Conference WWW 2008*, volume 300 of *ACM International Conference Proceeding Series*, pages 57–60, 2008. Cited on page(s) 71

[56] J. H. Herlocker, J. A. Konstan, and J. Riedl. Explaining collaborative filtering recommendations. In *Computer Supported Cooperative Work*, pages 241–250. ACM, 2000. Cited on page(s) 113

[57] Pascal Hitzler, Markus Krötzsch, and Sebastian Rudolph. *Foundations of Semantic Web technologies*. Springer–Verlag, 2010. Cited on page(s) 77, 81, 83, 86

[58] Johannes Hoffart, Fabian M. Suchanek, Klaus Berberich, and Gerhard Weikum. Yago: A spatially and temporally enhanced knowledge base from Wikipedia. *Artificial Intelligence*, 2012. Forth-coming. Cited on page(s) 70

[59] Aidan Hogan, Andreas Harth, Jürgen Umrich, and Stefan Decker. Towards a scalable search and query engine for the web. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 1301–1302. Association of Computing Machinery, New York, 2007. Cited on page(s) 32

128   BIBLIOGRAPHY

[60] M. Holi and E. Hyvönen. Fuzzy view-based semantic search. In *Proceedings of the 1st Asian Semantic Web Conference (ASWC2006), Beijing, China*. Springer–Verlag, 2006. Cited on page(s) 119

[61] Laura Hollink. *Semantic annotation for retrieval of visual resources*. PhD thesis, Free Univerisity of Amsterdam, 2006. SIKS Dissertation Series, No. 2006-24. Cited on page(s) 108

[62] E. Hyvönen and E. Mäkelä. Semantic autocompletion. In *Proceedings of the first Asia Semantic Web Conference (ASWC 2006), Beijing*. Springer–Verlag, 2006. Cited on page(s) 111

[63] E. Hyvönen, E. Mäkela, 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. Cited on page(s) 14, 20, 65, 86, 111

[64] E. Hyvönen, T. Ruotsalo, T. Häggströ, M. Salminen, M. Junnila, M. Virkkilä, 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*. Finnish AI Society, Espoo, Finland, 2006. Also in: Klaus Robering (Ed.), Information Technology for the Virtual Museum. LIT Verlag, 2008. Cited on page(s) 86, 114

[65] E. Hyvönen, S. Saarela, and K. 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)*. Springer–Verlag, 2004. Cited on page(s) 112, 119

[66] E. Hyvönen, M. Salminen, and M. Junnila. Annotation of heterogeneous database content for the semantic web. In *Proceedings of SemAnnot2004, Hiroshima, Japan*, Nov 2004. http://www.seco.tkk.fi/publications/2004/hyvonen-salminen-et-al-annotation-of-heterogeneous-2004.pdf. Cited on page(s) 100

[67] E. Hyvönen, K. Viljanen, E. Mäkelä, T. Kauppinen, T. Ruotsalo, O. Valkeapää, K. Seppälä, O. Suominen, O. Alm, R. Lindroos, T. Känsälä, 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*, pages 216–223. IEEE Press, Sept 2007. Cited on page(s) 10

[68] E. Hyvönen, K. Viljanen, J. Tuominen, and K. Seppälä. Building a national semantic web ontology and ontology service infrastructure—the FinnONTO approach. In *Proceedings of the 5th European Semantic Web Conference (ESWC 2008)*. Springer–Verlag, 2008. Cited on page(s) 90, 94, 104

[69] Eero Hyvönen. Preventing interoperability problems instead of solving them. *Semantic Web – Interoperability, Usability, Applicability*, 1(1–2):33–37, 2010. Cited on page(s) 122

[70] Eero Hyvönen, Olli Alm, and Heini Kuittinen. Using an ontology of historical events in semantic portals for cultural heritage. In *Proceedings of the Cultural Heritage on the Semantic Web Workshop at the 6th International Semantic Web Conference (ISWC 2007)*, 2007. `http://www.cs.vu.nl/~laroyo/CH-SW.html`. Cited on page(s) 75

[71] Eero Hyvönen, Thea Lindquist, Juha Törnroos, and Eetu Mäkelä. History on the semantic web as linked data – an event gazetteer and timeline for the World War I. In *Proceeedings of CIDOC 2012 – Enriching Cultural Heritage, Helsinki, Finland*. CIDOC, June 2012. `http://www.cidoc2012.fi/en/cidoc2012/programme`. Cited on page(s) 75

[72] Eero Hyvönen, Eetu Mäkelä, Tomi Kauppinen, Olli Alm, Jussi Kurki, Tuukka Ruotsalo, Katri Seppälä, Joeli Takala, Kimmo Puputti, Heini Kuittinen, Kim Viljanen, Jouni Tuominen, Tuomas Palonen, Matias Frosterus, Reetta Sinkkilä, Panu Paakkarinen, Joonas Laitio, and Katariina Nyberg. CultureSampo – Finnish culture on the Semantic Web 2.0. Thematic perspectives for the end-user. In *Museums and the Web 2009, Proceedings*. Archives and Museum Informatics, Toronto, 2009. Cited on page(s) 20, 33, 116

[73] Eero Hyvönen, Jouni Tuominen, Tomi Kauppinen, and Jari Väätäinen. Representing and utilizing changing historical places as an ontology time series. In Naveen Ashish and Amit Sheth, editors, *Geospatial Semantics and Semantic Web: Foundations, Algorithms, and Applications*. Springer–Verlag, 2011. Cited on page(s) 73, 95

[74] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. *Recommender Systems. An introduction*. Cambridge University Press, Cambridge, UK, 2011. Cited on page(s) 113

[75] K. Järvelin, J. Kekäläinen, and T. Niemi. ExpansionTool: Concept-based query expansion and construction. *Information Retrieval*, 4(3/4):231–255, 2001. Cited on page(s) 73

[76] M. Jensen. Vizualising complex semantic timelines. NewsBlip Research Papers, Report NBTR2003-001, 2003. `http://www.newsblip.com/tr/`. Cited on page(s) 75

[77] M. Junnila, E. Hyvönen, and M. Salminen. Describing and linking cultural semantic content by using situations and actions. In Klaus Robering, editor, *Information Technology for the Virtual Museum*. LIT Verlag, Berlin, 2008. Cited on page(s) 75, 113

[78] T. Känsälä and E. Hyvönen. A semantic view-based portal utilizing Learning Object Metadata. In *Semantic Web Applications and Tools Workshop, the 1st Asian Semantic Web Conference (ASWC2006)*, 2006. `http://www.seco.hut.fi/publications/2006/kansala-hyvonen-2006-semantic-portal-lom.pdf`. Cited on page(s) 118

[79] Akrivi Katifori, Constantis Halatsis, George Lepouras, Costas Vassilakis, and Eugeniua Giannopoulou. Ontology visualization methods – a survey. *ACM Computing Surveys*, 39(4), 2007. Article 10. Cited on page(s) 115, 119

[80] T. Kauppinen, C. Deichstetter, and E. 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. Cited on page(s) 116

[81] T. Kauppinen, R. Henriksson, J. Väätäinen, C. Deichstetter, and E. Hyvönen. Ontology-based modeling and visualization of cultural spatio-temporal knowledge. In *Semantic Web at Work— Proceedings of STeP 2006*. Finnish AI Society, Espoo, Finland, Nov 2006. Cited on page(s) 116

[82] Tomi Kauppinen, Glauco Mantegari, Panu Paakkarinen, Heini Kuittinen, Eero Hyvönen, and Stefania Bandini. Determining relevance of imprecise temporal intervals for cultural heritage information retrieval. *International Journal of Human-Computer Studies*, 86(9):549–560, September 2010. Cited on page(s) 74

[83] Tomi Kauppinen, Kimmo Puputti, Panu Paakkarinen, Heini Kuittinen, Jari Väätäinen, and Eero Hyvönen. Learning and visualizing cultural heritage connections between places on the semantic web. In *Proceedings of the Workshop on Inductive Reasoning and Machine Learning on the Semantic Web (IRMLeS2009), The 6th Annual European Semantic Web Conference (ESWC2009)*, 2009. Cited on page(s) 117

[84] Tomi Kauppinen, Jari Väätäinen, and Eero Hyvönen. Creating and using geospatial ontology time series in a semantic cultural heritage portal. In *Proceedings of the 5th European Semantic Web Conference (ESWC 2008)*, pages 110–123. Springer–Verlag, 2008. Cited on page(s) 95

[85] Suvi Kettula and Eero Hyvönen. Process-centric cataloguing of intangible cultural heritage. In *Proceeedings of CIDOC 2012 – Enriching Cultural Heritage, Helsinki, Finland*, 2012. http://www.cidoc2012.fi/en/cidoc2012/programme. Cited on page(s) 55, 113

[86] Barbara Ann Kipfer, editor. *Roget's International Thesaurus, 7th Edition*. Harper Collins Publishers, New York, 2011. Cited on page(s) 58

[87] R. Kishore, R. Ramesh, and R. Sharman, editors. *Ontologies: A Handbook of Principles, Concepts and Applications in Information Systems*. Springer–Verlag, 2006. Cited on page(s) 71

[88] Jussi Kurki and Eero Hyvönen. Authority control of people and organizations on the semantic web. In *Proceedings of the International Conferences on Digital Libraries and the Semantic Web 2009 (ICSD2009), Trento, Italy*, September 2009. http://www.seco.tkk.fi/publications/2009/kurki-hyvonen-onki-people-2009.pdf. Cited on page(s) 68

[89] Jussi Kurki and Eero Hyvönen. Collaborative metadata editor integrated with ontology services and faceted portals. In *Workshop on Ontology Repositories and Editors for the Semantic Web (ORES 2010), the Extended Semantic Web Conference ESWC 2010, Heraklion, Greece*, 2010. CEUR Workshop Proceedings, Vol 596, http://ceur-ws.org/. Cited on page(s) 104

[90] 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. Cited on page(s) 108, 119

[91] Thea Lindquist, Eero Hyvönen, Juha Törnroos, and Eetu Mäkelä. Leveraging linked data to enhance subject access - a case study of the University of Colorado Boulder's World War I collection online. In *World Library and Information Congress: 78th IFLA General Conference and Assembly, Helsinki*. International Federation of Library Associations and Institutions (IFLA), 2012. http://conference.ifla.org/ifla78. Cited on page(s) 75

[92] John LLoyd. *Foundations of Logic Programming (2nd edition)*. Springer–Verlag, 1987. Cited on page(s) 86

[93] Olivia Madison, John Byrum, Suzanne Jouguelet, Dorothy McGarry, Nancy Williamson, and Maria Witt, editors. *Functional requirements for bibliographic records. Final Report*. International Federation of Library Associations and Institutions (IFLA), 2009. http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf. Cited on page(s) 46, 47

[94] 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. Cited on page(s) 20

[95] E. Mäkelä, K. Viljanen, O. Alm, J. Tuominen, O. Valkeapää, T. Kauppinen, J. Kurki, R. Sinkkilä, T. Känsälä, R. Lindroos, O. Suominen, T. Ruotsalo, and E. Hyvönen. Enabling the semantic web with ready-to-use semantic widgets. In L. Nixon, R. Cuel, and C. Bergamini, editors, *First Industrial Results of Semantic Technologies, proceedings, co-located with ISWC 2007 + ASWC 2007, Busan, Korea*, 2007. CEUR Workshop Proceedings, Vol 293, http://ceur-ws.org. Cited on page(s) 119

[96] Eetu Mäkelä, Kaisa Hypén, and Eero Hyvönen. BookSampo—lessons learned in creating a semantic portal for fiction literature. In *Proceedings of ISWC-2011, Bonn, Germany*. Springer–Verlag, 2011. Cited on page(s) 33

[97] Eetu Mäkelä, Kaisa Hypén, and Eero Hyvönen. Fiction literature as linked open data—the BookSampo dataset. *Semantic Web – Interoperability, Usability, Applicability*, 2012. Accepted for publication. Cited on page(s) 33

[98] Eetu Mäkelä, Aleksi Lindblad, Jari Väätäinen, Rami Alatalo, Osma Suominen, and Eero Hyvönen. Discovering places of interest through direct and indirect associations in heterogeneous sources—the TravelSampo system. In *Terra Cognita 2011: Foundations, Technologies and Applications of the Geospatial Web*, 2011. CEUR Workshop Proceedings, Vol 798, http://ceur-ws.org/. Cited on page(s) 33

[99] Eetu Mäkelä, Tuukka Ruotsalo, and Hyvönen. How to deal with massively heterogeneous cultural heritage data—lessons learned in CultureSampo. *Semantic Web – Interoperability, Usability, Applicability*, 3(1), 2012. Cited on page(s) 20, 33

[100] Eetu Mäkelä, Osma Suominen, and Eero Hyvönen. Automatic exhibition generation based on semantic cultural content. In Aroyo et al. [6]. `http://www.cs.vu.nl/~laroyo/CH-SW.html`. Cited on page(s) 117

[101] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, Cambridge, UK, 2008. Cited on page(s) 119

[102] Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, Massachusetts, 1999. Cited on page(s) 110

[103] A. Maple. Faceted access: A review of the literature. Technical report, Working Group on Faceted Access to Music, Music Library Association, 1995. Cited on page(s) 119

[104] Suvodeep Mazumdar, Daniela Petrelli, and Fabio Ciravegna. Exploring user and system requirements of linked data visualization through a visual dashboard approach. *Semantic Web – Interoperability, Usability, Applicability*, 2012. In press. Cited on page(s) 117

[105] Willard McCarty. *Humanities Computing*. Palgrave, London, 2005. Cited on page(s) 11

[106] Pablo N. Mendes, Hannes Mühleise, and Christian Bizer. Sieve: Linked data quality assessment and fusion. In *EDBT/ICDT 2012 Joint Conference, Electronic Conference Proceedings, March 26–30, 2012, Berlin, Germany*, 2012. `http://www.edbt.org/Proceedings/2012-Berlin/workshops_toc.html`. Cited on page(s) 105

[107] Stuart E. Middleton, David De Roure, and Nigel R. Shadbolt. Ontology-based recommender systems. In Staab and Studer [139], pages 779–796. Cited on page(s) 119

[108] Peter Mika and Mike Potter. Metadata statistics for a large web corpus. In *Proceedings of Linked Data on the Web Workshop (LDOW 2012), at the 21th International World Wide Web Conference (WWW 2012)*, 2009. Cited on page(s) 33

[109] Gabor Nagypal, Richard Deswarte, and Jan Oosthoek. Applying the semantic web: The VICODI experience in creating visual contextualization for history. *Lit Linguist Computing*, 20(3):327–349, 2005. Cited on page(s) 75

[110] N. F. Noy. Ontology mapping. In Staab and Studer [139], pages 573–590. Cited on page(s) 93

[111] Natalya F. Noy, Nigam H. Shah, Patricia L. Whetzel, Benjamin Dai, Michael Dorf, Nicholas Griffith, Clement Jonquet, Daniel L. Rubin, Margaret-Anne Storey, Christopher G. Chute, and Mark A. Musen. BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research*, 37(Web Server issue):170–173, 2009. Cited on page(s) 77

[112] Natasha F. Noy and Mathieu d'Aquin. Where to publish and find ontologies? A survey of ontology libraries. *Web Semantics: Science, Services and Agents on the World Wide Web*, 11(0), 2011. Cited on page(s) 32, 77

[113] C. K. Ogden and I. A. Richards. *The meaning of meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*. Magdalene College, University of Cambridge, 1923. Cited on page(s) 60

[114] J. Park and S. Hunting. *XML Topic Maps: Creating and Using Topic Maps for the Web*. Addison-Wesley, New York, 2003. Cited on page(s) 10

[115] Glenn E. Patton, editor. *Functional Requirements for Authority Data – A Conceptual Model*. K. G. Saur, München, 2009. Cited on page(s) 47

[116] Daniel V. Pitti. Encoded archival description. An introduction and overview. *D-Lib Magazine*, 5(11), 1999. http://www.dlib.org/dlib/november99/11pitti.html. Cited on page(s) 41

[117] 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. Cited on page(s) 119

[118] Yves Raimond and Samer Abdallah. The event ontology, 2007. http://motools.sourceforge.net/event/event.html. Cited on page(s) 75

[119] E. Relph. *Place and placeness*. Pilon, London, U.K., 1976. Cited on page(s) 71

[120] D. Reynolds, P. Shabajee, and S. 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. Cited on page(s) 19, 20

[121] Bruce G. Robertson. Fawcett: A toolkit to begin an historical semantic web. *Digital Studies / Le champ numerique*, 1(2), 2009. Cited on page(s) 75

[122] 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. Cited on page(s) 108

[123] Anisa Rula. DC proposal: Towards linked data assessment and linking temporal facts. In *Proceedings of the 10th International Semantic Web Conference (ISWC 2011)*, pages 110–123. Springer–Verlag, 2010. Cited on page(s) 105

[124] T. Ruotsalo and E. Hyvönen. An event-based method for making heterogeneous metadata schemas and annotations semantically interoperable. In *Proceedings of ISWC 2007 + ASWC 2007, Busan, Korea*, pages 409–422. Springer–Verlag, 2007. Cited on page(s) 75, 86

[125] T. Ruotsalo and E. 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–Verlag, 2007. Cited on page(s) 113

[126] S. Russell and P. Norvig. *Artificial Intelligence. A Modern Approach.* Prentice-Hall, Upper Saddle River, New Jersey, 2010. Cited on page(s) 10

[127] G. M. Sacco. Dynamic taxonomies: guided interactive diagnostic assistance. In N. Wickramasinghe, editor, *Encyclopedia of Healthcare Information Systems*. Idea Group, 2005. Cited on page(s) 119

[128] Satya S. Sahoo, Wolfgang Halb, Sebastian Hellmann, Kingsley Idehen, Ted Thibodeau, Sören Auer, Juan Sequeda, and Ahmed Ezzat. A survey of current approaches for mapping of relational databases to RDF, 2009. `http://www.w3.org/2005/Incubator/rdb2rdf/RDB2RDF_SurveyReport.pdf`. Cited on page(s) 104

[129] G. Salton, A. Wong, and C. S. Yang. A vector space model for automatic indexing. *Communications of the ACM*, 18(11):613–620, 1975. Cited on page(s) 108

[130] Ansgar Scherp, Carsten Saathoff, and Thomas Franz. Event-Model-F, 2010. `http://www.uni-koblenz-landau.de/koblenz/fb4/AGStaab/Research/ontologies/events`. Cited on page(s) 75

[131] A. Schoulltz, A. Matteni, R. Isele, C. Bizer, and C. Becker. LDIF – linked data integration framework. In *Proceedings of the 2nd International Workshop on Consuming Linked Data (COLD 2011)*, 2011. CEUR Workshop Proceedings, Vol 782, `http://ceur-ws.org`. Cited on page(s) 86, 104

[132] Guus Schreiber, Alia Amin, Mark van Assem, Viktor de Boer, Lynda Hardman, Michiel Hildebrand, Laura Hollink, Zhisheng Huang, Janneke van Kersen, Marco de Niet, Borys Omelayenko, Jacco van Ossenbruggen, Ronny Siebes, Jos Taekema, Jan Wielemaker, and Bob Wielinga. Multimedian e-culture demonstrator. In *Proceedings of the 5th International Semantic Web Conference (ISWC 2006)*, pages 951–958, 2006. Cited on page(s) 20, 65, 86, 89, 114

[133] Ryan Shaw. LODE: An ontology for linking open descriptions of events, 2010. `http://linkedevents.org/ontology/`. Cited on page(s) 75

[134] A. Sheth, B. Aleman-Meza, I. B. Arpinar, C. Bertram, Y. Warke, C. Ramakrishnan, C. Halaschek, K. Anyanwu, D. Avant, F. S. Arpinar, and K. Kochut. Semantic association identification and knowledge discovery for national security applications. *Journal of Database Man-*

agement on Database Technology, 16(1):33–53, Jan–March 2005. Cited on page(s) 86, 114, 119

[135] S. Sirmakessis, editor. *Adaptive and Personalized Semantic Web*. Springer–Verlag, 2006. Cited on page(s) 117

[136] H. Southall, R. Mostern, and M. L. Berman. On historical gazetteers. *International Journal of Humanities and Arts Computing*, 5:127–145, 2011. Cited on page(s) 72

[137] J. Sowa. *Knowledge Representation. Logical, Philosophical, and Computational Foundations*. Brooks/Cole, 2000. Cited on page(s) 35, 58, 86

[138] 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*. Elsevier, Amsterdam, 2000. Cited on page(s) 19

[139] S. Staab and R. Studer, editors. *Handbook on ontologies (2nd Edition)*. Springer–Verlag, 2009. Cited on page(s) 77, 104, 125, 126, 132

[140] J. Suliman. Facilitating access: Empowering small museums. In *Museums and the Web 2007, Proceedings*. Archives and Museum Informatics, Toronto, 2007. http://www.museumsandtheweb.com/mw2007/papers/suliman/suliman.html. Cited on page(s) 14

[141] Osma Suominen and Eero Hyvönen. Improving the quality of SKOS vocabularies with Skosify. In *Proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2012)*. Springer–Verlag, 2012. Cited on page(s) 105

[142] Osma Suominen, Kim Viljanen, and Eero Hyvönen. User-centric faceted search for semantic portals. In *Proceedings of the 4th European Semantic Web Conference (ESWC 2007)*, pages 356–370. Springer–Verlag, 2007. Cited on page(s) 119

[143] Heidi Suonuuti. *A Guide to Terminology*. Finnish Centre for Technical Terminology/Nordterm, Helsinki, Finland, 2001. ISBN 952-9794-14-2. Cited on page(s) 60, 66

[144] A. Taylor. *Introduction to cataloging and classification*. Library and Information Science Text Series. Libraries Unlimited, 2006. Cited on page(s) 68, 69, 95

[145] B. Tillett. Authority control: State of the art and new perspectives. In *International Conference on Authority Control*. Haworth Press, Binghamton, NY, 2004. Cited on page(s) 68

[146] M. Tuffield, D. Millard, and N. Shadbolt. Ontological approaches to modelling narrative. In *2nd AKT DTA Symposium*, Jan. 2006. Cited on page(s) 75

[147] Jouni Tuominen, Nina Laurenne, and Eero Hyvönen. Biological names and taxonomies on the semantic web – managing the change in scientific conception. In *Proceedings of the 8th Extended Semantic Web Conference (ESWC 2011)*. Springer–Verlag, 2011. Cited on page(s) 76

[148] M. van Assem, V. Malaise, A. Miles, and G. Schreiber. A method to convert thesauri to SKOS. In *Proceedings of the 3rd European Semantic Web Conference (ESWC 2006)*. Springer–Verlag, 2006. Cited on page(s) 89

[149] M. van Assem, M. R. Menken, G. Schreiber, J. Wielemaker, and B. Wielinga. A method for converting thesauri to RDF/OWL. In *Proceedings of 3rd International Semantic Web Conference (ISWC 2004), Hiroshima, Japan*. Springer–Verlag, 2004. Cited on page(s) 89

[150] Mark van Assem. *Converting and Integrating Vocabularies for the Semantic Web*. PhD thesis, VU University, 2010. Cited on page(s) 89, 104

[151] Willem Robert van Hage, Véronique Malaisé, Roxane Segers, Laura Hollink, and Guus Schreiber. Design and use of the simple event model (SEM). *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(2):128–136, 2011. Cited on page(s) 75

[152] 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. Teesing. Searching and Annotating Virtual Heritage Collections with Semantic-Web Techniques. In *Proceedings of Museums and the Web 2007*, San Francisco, California, March 2007. Archives and Museum Informatics, Toronto. Cited on page(s) 20, 89

[153] K. Viljanen, T. Känsälä, 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. Cited on page(s) 113

[154] K. Viljanen, J. Tuominen, T. Känsälä, and E. Hyvönen. Distributed semantic content creation and publication for cultural heritage legacy systems. In *Proceedings of the 2008 IEEE International Conference on Distibuted Human–Machine Systems, Athens, Greece*. IEEE Press, 2008. Cited on page(s) 77, 118

[155] Ubbo Visser. *Intelligent information integration for the Semantic Web*. Springer–Verlag, 2004. Cited on page(s) 71

[156] Manolis Wallace, Marios C. Angelides, and Phivos Mylonas, editors. *Advances in Semantic Media Adaptation and Personalization*. Springer–Verlag, 2006. Cited on page(s) 119

[157] W. Wong, W. Liu, and M. Bennamoun. Ontology learning from text: A look back and into the future. *ACM Computing Surveys*, 44(4):1–36, 2012. article 20. Cited on page(s) 104

[158] David Wood, editor. *Linking Government Data*. Springer–Verlag, 2011. Cited on page(s) 10

[159] G. P. Zarri. Semantic annotations and semantic web using nkrl (narrative knowledge representation language). In *Proceedings of the 5th International Conference on Enterprise Information Systems, Angers, France (ICEIS 2003)*, pages 387–394, 2003. Cited on page(s) 75

[160] Marcia Lei Zeng, Maja Zumer, and Athena Salaba, editors. *Functional Requirements for Subject Authority Data (FRSAD). A Conceptual Model*. International Federation of Library Associations and Institutions (IFLA), 2010. `http://www.ifla.org/files/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-model.pdf`. Cited on page(s) 48

Author's Biography

EERO HYVÖNEN

Eero Hyvönen[1] is a professor of semantic media technology at the Aalto University, Department of Media Technology, and an adjunct professor of computer science at the University of Helsinki, Department of Computer Science. He directs the Semantic Computing Research Group SeCo[2] specializing in Semantic Web technologies and applications. A major theme in his research during the last years has been the development of a semantic web content infrastructure on a national scale in Finland and its applications in areas such as Cultural Heritage. Eero Hyvönen has published over 300 articles, papers, and books. With his SeCo group, he has received several international and national awards, including the Semantic Web Challenge Award (in 2004 and 2008), World Summit Award (WSA) (2010), and Apps4Finland – Doing Good with Open Data (2010). He acts on the editorial boards of *Semantic Web – Interoperability, Usability, Applicability, Semantic Computing, and International Journal of Metadata, Semantics, and Ontologies,* and has co-chaired and acted on the program committees of tens of major international conferences and workshops, such as ESWC, ISWC, IJCAI, WWW, ICSC, etc.

[1] http://www.seco.tkk.fi/u/eahyvone/
[2] http://www.seco.tkk.fi/

Index