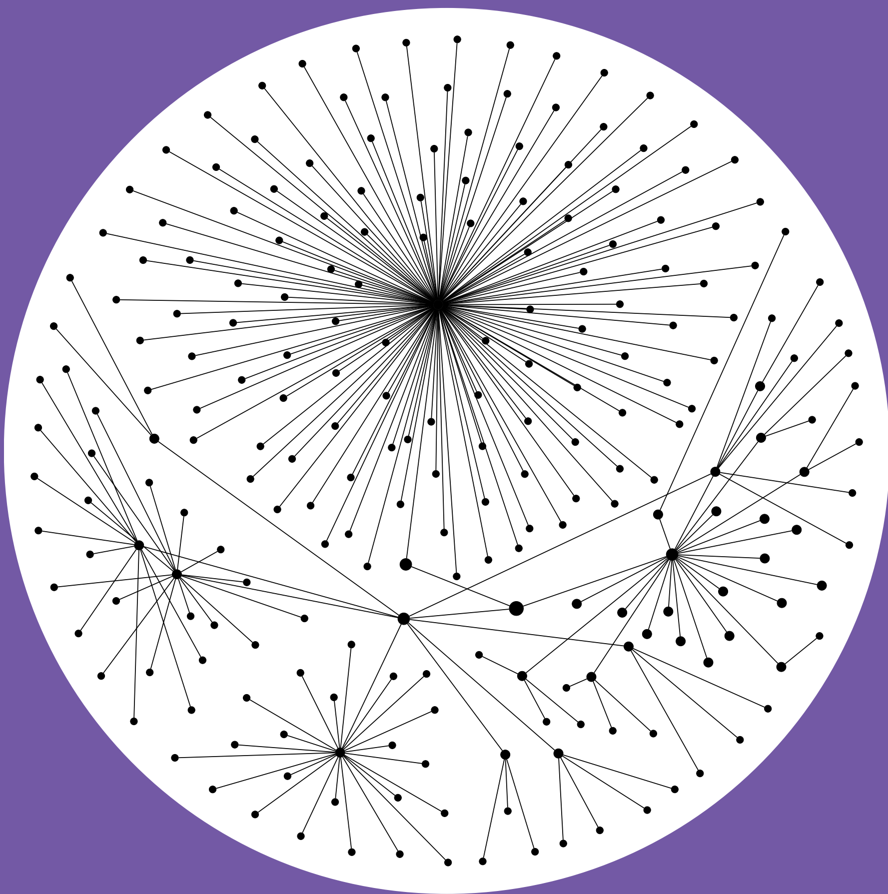
Department of Computer Science

Ontology Services for Knowledge Organization Systems

Jouni Tuominen



Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 101/2017

Ontology Services for Knowledge Organization Systems

Jouni Tuominen

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science, at a public examination held at the lecture hall TU1 of the school on 16 June 2017 at 12 noon.

Aalto University
School of Science
Department of Computer Science
Semantic Computing Research Group

Supervising professor
Eero Hyvönen

Thesis advisor
Eero Hyvönen

Preliminary examiners
Professor Vagan Terziyan, University of Jyväskylä, Finland
Professor Mathieu d'Aquin, National University of Ireland Galway, Ireland

Opponent
Professor Marcia Lei Zeng, Kent State University, United States

Aalto University publication series
DOCTORAL DISSERTATIONS 101/2017

ISBN 978-952-60-7456-6 (printed)
ISBN 978-952-60-7455-9 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-7455-9

Unigrafia Oy
Helsinki 2017

Finland

441 697
Printed matter

A? Aalto University

Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Author
Jouni Tuominen

Name of the doctoral dissertation
Ontology Services for Knowledge Organization Systems

Publisher School of Science

Unit Department of Computer Science

Series Aalto University publication series DOCTORAL DISSERTATIONS 101/2017

Field of research Knowledge Technology

Manuscript submitted 12 March 2017 **Date of the defence** 16 June 2017

Permission to publish granted (date) 25 April 2017 **Language** English

☐ **Monograph** ☒ **Article dissertation** ☐ **Essay dissertation**

Abstract

Ontologies and other knowledge organization systems, such as controlled vocabularies, can be used to enhance the findability of information. By describing the contents of documents using a shared, harmonized terminology, information systems can provide efficient search and browsing functionalities for the contents. Explicit descriptive metadata aims to solve some of the prevailing issues in full text search in many search engines, including the processing of synonyms and homonyms. The use of ontologies as domain models enables the machine-processability of contents, semantic reasoning, information integration, and other intelligent ways of processing the data.

The utilization of knowledge organization systems in content indexing and information retrieval can be facilitated by providing automated tools for their efficient use. This thesis studies and presents novel methods and systems for publishing and using knowledge organization systems as ontology services. The research is conducted by designing and evaluating prototype systems that support the use of ontologies in real-life use cases. The research follows the principles of the design science and action research methodologies.

The presented ONKI system provides user interface components and application programming interfaces that can be integrated into external applications to enable ontology-based workflows. The features of the system are based on analyzing the needs of the main user groups of ontologies. The common functionalities identified in ontology-based workflows include concept search, browsing, and selection.

The thesis presents the Linked Open Ontology cloud approach for managing and publishing a set of interlinked ontologies in an ontology service. The system enables the users to use multiple ontologies as a single, interoperable, cross-domain representation instead of individual ontologies. For facilitating the simultaneous use of ontologies published in different ontology repositories, the Normalized Ontology Repository approach is presented.

As a use case of managing and publishing a semantically rich knowledge organization system as an ontology, the thesis presents the Taxon Meta-Ontology model for biological nomenclatures and classifications. The model supports the representation of changes and differing opinions of taxonomic concepts.

The ONKI system and the ontologies developed using the methods presented in this thesis have been provided as a living lab service http://onki.fi, which has been run since 2008. The service provides tools and support for the users of ontologies, including content indexers, information searchers, ontology developers, and application developers.

Keywords semantic web, knowledge organization systems, metadata creation, ontology services

ISBN (printed) 978-952-60-7456-6 **ISBN (pdf)** 978-952-60-7455-9

ISSN-L 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2017

Pages 198 **urn** http://urn.fi/URN:ISBN:978-952-60-7455-9

Aalto-yliopisto

Aalto-yliopisto, PL 11000, 00076 Aalto www.aalto.fi

Tiivistelmä

Tekijä
Jouni Tuominen

Väitöskirjan nimi
Ontology Services for Knowledge Organization Systems

Julkaisija Perustieteiden korkeakoulu

Yksikkö Tietotekniikan laitos

Sarja Aalto University publication series DOCTORAL DISSERTATIONS 101/2017

Tutkimusala Tietämystekniikka

Käsikirjoituksen pvm 12.03.2017 **Väitöspäivä** 16.06.2017

Julkaisuluvan myöntämispäivä 25.04.2017 **Kieli** Englanti

☐ **Monografia** ☒ **Artikkeliväitöskirja** ☐ **Esseeväitöskirja**

Tiivistelmä

Ontologioita ja muita tietämyksen järjestämisen menetelmiä, kuten kontrolloituja sanastoja, voidaan käyttää tiedon löytämisen parantamiseksi. Kun dokumenttien sisällöt kuvaillaan käyttämällä jaettua, yhtenäistettyä terminologiaa, tietojärjestelmät voivat tarjota tehokkaita haku- ja selaustoiminnallisuuksia sisältöihin. Eksplisiittisesti esitetty, kuvaileva metatieto pyrkii ratkaisemaan monien hakukoneiden käyttämän kokotekstihaun ongelmia, kuten synonyymien ja homonyymien huomioimisen. Ontologioiden käyttäminen käsitemalleina mahdollistaa sisältöjen koneellisen käsittelyn, semanttisen päättelyn, tiedon integroinnin ja muita älykkäitä menetelmiä.

Tietämyksen järjestämisen menetelmien hyödyntämistä sisältöjen indeksoinnissa ja haussa voidaan helpottaa tarjoamalla käyttäjille automatisoituja työkaluja niiden tehokkaaseen käyttämiseen. Tässä väitöskirjassa tutkitaan ja esitetään uudenlaisia menetelmiä ja järjestelmiä tietämyksen järjestämisen menetelmien julkaisemiseksi ontologiapalveluina. Tutkimus on toteutettu suunnittelemalla ja arvioimalla prototyyppijärjestelmiä, jotka edistävät ontologioiden käyttämistä todellisissa käyttötapauksissa. Tutkimus nojautuu suunnittelutieteen ja toimintatutkimuksen metodologioiden periaatteisiin.

Työssä esitetty ONKI-järjestelmä tarjoaa käyttöliittymäkomponentteja ja ohjelmallisia rajapintoja, jotka voidaan integroida ulkoisiin sovelluksiin ontologiaperustaisten työnkulkujen mahdollistamiseksi. Järjestelmän ominaisuudet on toteutettu perustuen ontologioiden keskeisten käyttäjäryhmien tarpeiden selvittämiseen. Ontologiaperustaisista työnkuluista tunnistettuja yleisiä toiminnallisuuksia ovat käsitteen haku, selailu ja valinta.

Tässä työssä esitetään linkitetyn avoimen ontologiapilven menetelmä toisiinsa linkitettyjen ontologioiden ylläpitämiseen ja julkaisemiseen ontologiapalvelussa. Järjestelmän avulla käyttäjät voivat käyttää useita ontologioita yhtenä, yhteentoimivana, alat yhdistävänä kokonaisuutena erillisten ontologioiden sijaan. Eri ontologiapalveluissa julkaistujen ontologioiden samanaikaisen käytön helpottamiseksi esitetään normalisoidun ontologiapalvelun menetelmä.

Käyttötapauksena semanttisesti rikkaan tietämyksen järjestämisen menetelmän ylläpitämisestä ja julkaisemisesta työssä esitetään biologisten nimistöjen ja luokitusten taksonominen ontologiamalli. Malli mahdollistaa taksonomisten käsitteiden muutosten ja toisistaan poikkeavien näkemysten esittämisen.

ONKI-järjestelmä ja työssä esitetyillä menetelmillä kehitetyt ontologiat ovat olleet käytettävissä living lab -palvelussa http://onki.fi, joka on ollut toiminnassa vuodesta 2008 lähtien. Palvelu tarjoaa työkaluja ja tukea ontologioiden käyttäjille, kuten tiedon indeksoijille, hakijoille, ontologioiden kehittäjille ja sovelluskehittäjille.

Avainsanat semanttinen web, tietämyksen järjestämisen menetelmät, metadatan tuottaminen, ontologiapalvelut

ISBN (painettu) 978-952-60-7456-6 **ISBN (pdf)** 978-952-60-7455-9

ISSN-L 1799-4934 **ISSN (painettu)** 1799-4934 **ISSN (pdf)** 1799-4942

Julkaisupaikka Helsinki **Painopaikka** Helsinki **Vuosi** 2017

Sivumäärä 198 **urn** http://urn.fi/URN:ISBN:978-952-60-7455-9

Preface

The seeds for this dissertation were planted in 2007 when I started working in the Semanting Computing Research Group (SeCo), at the Department of Computer Science, Aalto University (formerly the Department of Media Technology, Helsinki University of Technology) and the Department of Computer Science, University of Helsinki, Finland.

I have had the opportunity to work with a lot of wonderful colleagues and collaborators during these years. I have been enjoying the guidance and support of professor Eero Hyvönen, who has made the work possible by fertilizing my imagination and encouraging me to continue my work.

I wish to thank the pre-examiners professor Vagan Terziyan and professor Mathieu d'Aquin for valuable feedback and suggestions for improving this thesis.

When I started in SeCo, I was handed to the gentle care of my colleagues Kim Viljanen and Eetu Mäkelä. Kim has been my trusted wingman in building the ONKI service and an elevating mentor—for that I will always be grateful. Eetu has given me valuable guidance and support on methodological and technological decisions.

During my doctoral studies, I have been collaborating widely with Matias Frosterus. All the work would have been much less fun without him. I would like to thank Nina Laurenne for introducing me to the world of biological taxonomies, and diving with me in the great ball pool of colorful organisms. I wish to thank Tomi Kauppinen for his fruitful ideas on ontology services and showing by example how to be a productive researcher, and Tuukka Ruotsalo for always emphasizing the importance of scientific rigor in our work.

The ontology services presented in the thesis would be of no use without the actual ontologies. I express my gratitude to the head ontologist Katri Seppälä and Tuomas Palonen, who have been working with the ontologies

of the KOKO cloud. I wish also thank Osma Suominen, Sini Pessala, Jussi Kurki, Reetta Sinkkilä, and Robin Lindroos for participating in building the tools for supporting the ontology infrastructure, and Mikko Salonoja, Rami Aamulehto, Alex Johansson, and Henri Ylikotila for their work on the user interfaces of the ONKI service. Mikko Koho and Hannu Saarenmaa have been valuable contributors on the biological taxonomies. I am grateful for the support and collaboration of Erkki Heino, Esko Ikkala, Petri Leskinen, Arttu Oksanen, Claire Tamper, and all my other colleagues in SeCo and collaborators I have been working with during the years.

The work presented in this thesis has been carried out as part of the National Semantic Web Ontology Project in Finland (FinnONTO, 2003–2012) and Linked Data Finland project (2012–2014), both funded by the Finnish Funding Agency for Innovation (Tekes). The work has also received funding from Lusto – Finnish Forest Museum (2008), EU project MedIEQ (2006–2008), EU FP7 project SMARTMUSEUM (2008–2010), and EU FI-PPP project ENVIROFI (2011–2013). I have received grants from the Finnish Cultural Foundation (2013) and KAUTE Foundation (2015) for the doctoral research.

During the thesis work, I have also worked in the National Gazetteer of Historical Places project (2015–2016), funded by the Finnish Cultural Foundation, the Semantic Finlex project (2015–2017), funded by the Ministry of Justice and the Ministry of Finance, the Linked Open Data Science Service project (2015–2016), funded by the Ministry of Education and Culture, and the Cultures of Knowledge project (2015–2017), funded by The Andrew W. Mellon Foundation. I have received funding for short term scientific mission from the European Cooperation in Science and Technology (COST, 2016) and travel grants from the Helsinki Doctoral Programme in Computer Science (2014, 2015).

I would like to thank my parents for their support and encouragement on the path I have chosen. Finally, I wish to express my deepest gratitude to Krista for her love and understanding during the long hours.

Helsinki, May 10, 2017,

Jouni Tuominen

Contents

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I Kim Viljanen, Jouni Tuominen, and Eero Hyvönen. Ontology Libraries for Production Use: The Finnish Ontology Library Service ONKI. In *The Semantic Web: Research and Applications: 6th European Semantic Web Conference, ESWC 2009, Heraklion, Crete, Greece, May 31–June 4, 2009, Proceedings*, Lora Aroyo, Paolo Traverso, Fabio Ciravegna, Philipp Cimiano, Tom Heath, Eero Hyvönen, Riichiro Mizoguchi, Eyal Oren, Marta Sabou, and Elena Simperl (editors), Lecture Notes in Computer Science, volume 5554, pages 781–795, ISBN 978-3-642-02120-6, Springer-Verlag, June 2009.

II Jouni Tuominen, Matias Frosterus, Kim Viljanen, and Eero Hyvönen. ONKI SKOS Server for Publishing and Utilizing SKOS Vocabularies and Ontologies as Services. In *The Semantic Web: Research and Applications: 6th European Semantic Web Conference, ESWC 2009, Heraklion, Crete, Greece, May 31–June 4, 2009, Proceedings*, Lora Aroyo, Paolo Traverso, Fabio Ciravegna, Philipp Cimiano, Tom Heath, Eero Hyvönen, Riichiro Mizoguchi, Eyal Oren, Marta Sabou, and Elena Simperl (editors), Lecture Notes in Computer Science, volume 5554, pages 781–795, ISBN 978-3-642-02120-6, Springer-Verlag, June 2009.

III Jouni Tuominen, Tomi Kauppinen, Kim Viljanen, and Eero Hyvönen. Ontology-Based Query Expansion Widget for Information Retrieval. In *Proceedings of the 5th International Workshop on Scripting and Develop-*

ment for the Semantic Web at ESWC 2009, Heraklion, Greece, May 31, Sören Auer, Chris Bizer, and Gunnar Aastrand Grimnes (editors), CEUR Workshop Proceedings, volume 449, pages 52–57, ISSN 1613-0073, online CEUR-WS.org/Vol-449/ShortPaper1.pdf, May 2009.

IV Matias Frosterus, Jouni Tuominen, Sini Pessala and Eero Hyvönen. Linked Open Ontology cloud: managing a system of interlinked cross-domain light-weight ontologies. *International Journal of Metadata, Semantics and Ontologies*, 10, 3, pages 189–201, DOI 10.1504/IJMSO.2015.073879, December 2015.

V Kim Viljanen, Jouni Tuominen, Eetu Mäkelä and Eero Hyvönen. Normalized Access to Ontology Repositories. In *ICSC 2012: 2012 IEEE Sixth International Conference on Semantic Computing*, Palermo, Italy, 19-21 September 2012, pages 109–116, ISBN 978-1-4673-4433-3, IEEE, September 2012.

VI Jouni Tuominen, Nina Laurenne, and Eero Hyvönen. Biological Names and Taxonomies on the Semantic Web – Managing the Change in Scientific Conception. In *The Semantic Web: Research and Applications: 8th Extended Semantic Web Conference, ESWC 2011, Heraklion, Crete, Greece, May 29 – June 2, 2011, Proceedings, Part II*, Grigoris Antoniou, Marko Grobelnik, Elena Simperl, Bijan Parsia, Dimitris Plexousakis, Pieter De Leenheer, and Jeff Pan (editors), Lecture Notes in Computer Science, volume 6644, pages 255–269, ISBN 978-3-642-21063-1, Springer-Verlag, June 2011.

VII Nina Laurenne, Jouni Tuominen, Hannu Saarenmaa and Eero Hyvönen. Making species checklists understandable to machines – a shift from relational databases to ontologies. *Journal of Biomedical Semantics*, 5, 40, DOI 10.1186/2041-1480-5-40, September 2014.

VIII Jouni Tuominen, Nina Laurenne and Eero Hyvönen. Publishing and Using Plant Names as an Ontology Service. In *Proceedings of the first international Workshop on Semantics for Biodiversity at ESWC 2013*, Montpellier, France, May 27, Pierre Larmande, Elizabeth Arnaud, Isabelle Mougenot, Clement Jonquet, Therese Libourel, Manuel Ruiz (editors), CEUR Workshop Proceedings, volume 979, ISSN 1613-0073,

online CEUR-WS.org/Vol-979/WS_s4biodiv2013_paper_2.pdf, May 2013.

Author's Contribution

Publication I: "Ontology Libraries for Production Use: The Finnish Ontology Library Service ONKI"

The author contributed significantly to the design and specification of the ONKI library system, and the general requirements of ontology library services. The author was one of the two primary developers of the ONKI SKOS server and the ONKI selector widget, and implemented the ONKI API in the ONKI SKOS server.

Publication II: "ONKI SKOS Server for Publishing and Utilizing SKOS Vocabularies and Ontologies as Services"

The author is the first author of the publication. The author was one of the two primary designers and developers of the ONKI SKOS server and the ONKI selector widget, and primary developer of the ONKI API and SKOS support for the system. The author implemented the demonstration search interface for using ontology-based query expansion in the Kantapuu system.

Publication III: "Ontology-Based Query Expansion Widget for Information Retrieval"

The author is the first author of the publication. The author designed and implemented the query expansion functionality of the ONKI selector widget, configured the ontologies used in the use case and implemented the demonstration search interface.

Publication IV: "Linked Open Ontology cloud: managing a system of interlinked cross-domain light-weight ontologies"

The author contributed significantly to the principles guiding the building of the ontology cloud and the management process of the cloud. The author designed the publication of the KOKO ontology cloud in the ontology service, and developed the functionalities in the ONKI service to support the processing of the ontology cloud and individual domain ontologies mapped to the general upper ontology.

Publication V: "Normalized Access to Ontology Repositories"

The author participated in the design of the Normalized Ontology Repository (NOR) approach and its API. The author was the primary developer of the HTTP API used to access the ontologies in the ONKI SKOS server. The author contributed to the design of the ONKI3 user interface and the ontology metadata specification.

Publication VI: "Biological Names and Taxonomies on the Semantic Web – Managing the Change in Scientific Conception"

The author shares the first authorship of the publication. The author was the one of the two primary designers of the Taxon Meta-Ontology (TaxMeOn), responsible for the computer science expertise, and was in charge of the technical implementation of the model. The author committed the technical details related to the use cases.

Publication VII: "Making species checklists understandable to machines – a shift from relational databases to ontologies"

The author contributed to the publication equally as the first author. The author was one of the two primary designers of the taxonomic meta-ontology, and responsible for the computer science expertise, the technical comparison of the LSID and HTTP URI identifiers, and evaluation of the models for managing taxonomic information.

Publication VIII: "Publishing and Using Plant Names as an Ontology Service"

The author is the first author of the publication. The author was one of the two primary designers of the ontology model of the vernacular names, responsible for the computer science expertise, published the ontology in the ONKI service, and designed the management process of the nomenclature based on the SAHA metadata editor.

In addition to these publications, the thesis contains references to related work by the author. They include the description of the first version of the ONKI widget [268], the FinnONTO ontology infrastructure [136], the ONKI2 user interface [258], the SPARQL-based ontology service ONKI Light [248], the deployment and further development of the ONKI service by the National Library of Finland [249], and the use of ontology services in the legal domain [87].

The author has also published other work related to ontology services, including using the ONKI service in cultural heritage [133], birdwatching [129], for publishing historical places as on ontology time series [135], and for visualizing automatically enriched ontology relationships based on co-occurrences of concepts in annotations [153]. Furthermore, the author has worked on publishing ontologies in a linked data service [134] and developing a federated ontology service for historical places [130].

1. Introduction

1.1 Background and Research Environment

As the amount of information in information systems grows, it is harder for the users to find relevant information for their needs [137]. Not only is the information hard to find, but it is not connected to other relevant information—thus getting an extensive understanding about a topic is challenging. These issues are intensified in the massive World Wide Web, which was estimated to contain 11.5 billion indexable web pages already in 2005 [102], and almost 50 billion of them in a more recent study [263].

The full text search employed by many search engines has several limitations. A simple search algorithm does not distinguish significant words from non-significant ones in the document, leading to the loss of precision of the search [46]. The issue can be compensated by ordering the search results based on their assessed relevancy for the search task [181], effectively displaying the most relevant results first. On the other hand, all significant terms regarding the information content of the document might not appear in the document, decreasing the recall of the search [114].

The Semantic Web[1] [28, 77] is an extension of the current World Wide Web, providing technologies for processing information based on their semantics. Managing textual contents on the conceptual level resolves the issues of purely lexical full text search, such as handling synonymy and homonymy. A practical subtopic of the Semantic Web is the Linked Data [27, 112] concept, which is a method for publishing data in an interlinked way. The semantic interoperability and interlinking of the information contents in the web changes the nature of the web from the web of documents to the web of data. These technologies enable building of

[1] `http://www.w3.org/2001/sw/`

services on top of the integrated datasets, for example, providing users novel search and recommendation interfaces, thus improving information findability.

Ontologies are at the core of the Semantic Web infrastructure, as they model the domains of interest in a formal, machine-understandable way. An ontology acts as a shared conceptualization of a domain [98, 35, 244], enabling different parties to use a common language when communicating about the domain [100]. When information objects are described with metadata [222, 17, 71, 46] referring to concepts of an ontology, machines can interpret their meanings. This semantic content annotation enables, e.g., the integration of heterogeneous data collections and automatic reasoning based on the properties of concepts [143, 239].

An ontology can be built by a domain expert, but also methods for automatic and semi-automatic ontology generation exist [176]. Also the content annotation can be done manually or (semi-)automatically [159, 234, 276]. In addition to ontologies, semantically lighter knowledge organization systems (KOS) originating from library and information science, such as subject headings, classifications, and thesauri can be used to harmonize the used terminology in content descriptions and search [124, 95]. KOSs can also be utilized in, e.g., query expansion, cross-language search, and as a navigation aid for accessing contents. Knowledge organization systems can be interlinked in order to facilitate the integration of data described using different KOSs, by utilizing the methods of ontology mapping [127] and matching [233].

An alternative method for using explicit ontology-based metadata for improving the information findability is to use automatic methods analyzing the contents of information objects. For example, natural language processing methods can be used to identify the meanings of words based on their context in a text document [53]. However, the strength of using explicit metadata is its applicability also to non-textual objects, such as images and videos, as otherwise text search would not be possible without reliable content analysis methods.

For facilitating the use of ontologies, specialized software systems—ontology servers, have been proposed for publishing ontologies and providing services for using them [79, 68, 9, 108, 60]. Most of the ontology server implementations introduced in the Semantic Web research have been designed for developing ontologies, and not for their actual usage, e.g., in content annotation or information retrieval, and therefore lack crucial functionalities needed in applications [9]. The common functionalities of ontology servers include user interfaces for visualization and browsing of ontologies, and searching for concepts in an ontology. Several implementations also provide application programming interfaces (API) for the programmatic use of the ontologies. In addition to APIs, ontology functionalities may be integrated into client systems with user interface components.

1.2 Objectives and Scope

The aim of this thesis is to provide methods and technological solutions for publishing knowledge organization systems in such a way that they can be utilized cost-effectively in external applications. As a solution, the notion of an ontology service is presented. An ontology service is a software system that can be used by ontology developers for publishing their ontologies, and by content indexers, information searchers, and application developers to use ontologies in their tasks. The user needs for knowledge organization systems are analyzed, and based on them a set of requirements for functionalities is formulated. As a proof-of-concept system, implementations of such an ontology service are provided and their application to real life cases is reported. The KOSs used in the cases include Finnish and international thesauri, lightweight ontologies covering general and domain-specific concepts, and semantically richer biological nomenclatures and classifications.

The objectives of the ontology services presented in this thesis are:

- **Ontology publication channel.** Provide a complete publication workflow for the ontology developers to publish an ontology, or a new version of it.
- **Heterogeneous ontologies.** Support for distinct ontology formats by using a harmonizing data model and configuration options.
- **Tools for metadata creation.** Provide means to ontology-based content indexing.
- **Support for distributed content creation.** Facilitate content creation in distributed workflows, where content is curated by independent parties and aggregated into one system, e.g., a web portal.
- **Facilitate search tasks.** Support the use of published ontologies in

information retrieval by offering functionalities, e.g., for query expansion.

- **Multiple ontologies and repositories.** The users should be able to access multiple ontologies, even originating from different ontology services, simultaneously in a coherent way.
- **Programmatic access.** Applications should be able to use the ontologies via application programming interfaces (API) for searching for concepts and getting their properties.
- **Evaluation by applying into practice.** The applicability of the services will be tested by building a proof-of-concept system, which is piloted in real life scenarios.
- **Promote complex KOSs.** The system should support not only simple, but also richer knowledge organization systems.

Based on these objectives that guide the design and implementation of the ontology services this thesis seeks to find solutions for the following research questions (RQ):

1. How can lightweight ontologies be published on the Semantic Web so that they can be utilized in content indexing and information retrieval tasks?

2. How can a collection of independent or interconnected ontologies—in different formats and repositories—be published and utilized using shared user interfaces and APIs?

3. How can richer knowledge organization systems, such as biological nomenclatures and classifications, be managed as an ontology and published using an ontology service?

The research questions are answered with the publications I–VIII. Table 1.1 shows which research questions the individual publications contribute to. The contributions of the publications are summarized in Chapter 3.

1.3 Research Process and Dissertation Structure

The research presented in this thesis has been conducted by applying the methodologies of design science [182, 116, 211] and action research [24, 42,

Research question	PI	PII	PIII	PIV	PV	PVI	PVII	PVIII
RQ1	x	x	x					
RQ2				x	x			
RQ3						x	x	x

Table 1.1. The relationship between the research questions and the publications.

62].

Design science is a technology-oriented paradigm in the information systems discipline that aims to create things that serve human purposes [182]. The significance of a research is determined by the value or utility it provides—does it work, is it an improvement? Instead of new theories, the outcomes of design science are innovative and useful artifacts, which include constructs, models, methods, and implementations. The process of design science includes two phases: building and evaluation. The nature of design science tends to be applied: it exploits knowledge created by basic research to develop new technologies. However, the created artifact and its working environment might not be well understood, and in such case the artifact itself presents new scientific questions. By designing, building, and applying an artifact, knowledge and understanding about a problem domain and its solution is achieved [116]. As opposed to routine design and systems building work, design science builds novel ways to solve important, unsolved problems or provides more effective or efficient ways to address previously solved problems. The solutions are generalizable and provide new knowledge for the application domain. The applicability of the artifact is evaluated in real-world scenarios by observational, analytical, experimental, testing, or descriptive methods.

Complementing the technological aspect of design science, action research emphasizes the social elements of information systems research. In an action research setting, scientists and the subjects of the study collaborate in order to study and solve problems in organizations [24]. The research involves two phases: the diagnostic stage to analyze the current situation and the therapeutic stage to carry out changes to improve the situation. In contrast to case studies, in action research the researcher is involved in the studied phenomenon and the research is carried out in a more rigorous way [23]. The rigor is ensured by following the action research cycle: diagnosis, action planning, action taking, evaluating, and specifying learning.

The design of the ontology services presented in this thesis is based on analyzing the user requirements of ontology users and existing ontology server implementations by a) conducting a literature and system review, b) formulating illustrative use scenarios, and c) running a prototype system as a living lab service, gathering feedback from the actual users of the system. Based on the cyclic nature of design science and action research, similar methods have been used to evaluate the purposefulness of the developed ontology services. The prototype system itself acts as a proof of concept, demonstrating the utility or suitability of the software artifact for the given requirements [210]. Furthermore, using the system in an action research setting in real use cases evaluates the effects of the system use in real-world situations. By basing the functionalities of the system on existing research and illustrative use scenarios, the utility of the system is ensured.

This thesis is organized as follows. The theoretical background of the work is presented in Chapter 2. Building on the theory and based on the publications included, the results of the thesis are summarized in Chapter 3. Finally, the implications of the results, the validity of the work, and further research are discussed in Chapter 4.

2. Theoretical Foundation

2.1 Modeling Knowledge Organization Systems

2.1.1 Knowledge Organization Systems

Knowledge organization systems (KOS) originate from library and information science, where they are used as schemes for organizing information and promoting knowledge management [124, 95]. Examples of different types of KOSs include classification schemes, subject headings, authority files, taxonomies, thesauri, and ontologies [124, 119, 238, 95]. They provide a controlled vocabulary in the given domain of interest, and harmonize the terminology used to describe the information items in information collections, e.g., in digital libraries or document databases.

In its simplest form, a controlled vocabulary is a list of terms, where each term corresponds to a concept of the domain. It can also include other information about the terms, such as synonyms, descriptions, and source information. A taxonomy arranges the terms in a controlled vocabulary into a hierarchy, aiding the users selecting a suitable term in, e.g., content description or information retrieval [91]. Extending taxonomies, thesauri may include richer information about the terms, such as associative relations between them [91]. Guidelines for creating, displaying, and managing thesauri are documented in international and national standards, such as ISO 25964 [6, 66] and SFS 5471 [2].

Thesauri and other controlled vocabularies are used primarily for improving information retrieval [11, 236]. This is accomplished by using the concepts or terms of a thesaurus in content indexing, content searching, or in both of them, thus simplifying the matching of query terms and the indexed resources (e.g., documents) when compared with using free, uncontrolled natural language. The relations of thesauri can be utilized in information retrieval, for example by expanding query terms to more specific terms based on the concept hierarchy. Multilingual thesauri can be used for cross-language search where the information contents or the related metadata are expressed in different language than the one used by the end user.

Knowledge Organization Systems, such as thesauri, are of great benefit for the Semantic Web [19, 282, 193, 106, 262, 278], enabling semantically disambiguated data exchange and integration of data from different sources, though not to the same extent as ontologies [240] where the semantics of concepts is defined in more refined and machine-understandable ways [232, 10]. Ontologies based on thesaurus-like structures can be called lightweight ontologies [82, 159, 93, 143].

The Simple Knowledge Organization System (SKOS) [188, 19, 189] developed within W3C is a data model and a syntax for expressing concept schemes such as thesauri, and is largely compatible with the ISO 25964 thesaurus standard [66, 139]. SKOS provides a standard way for creating vocabularies and migrating existing vocabularies to the Semantic Web. SKOS solves the problem of diverse, non-interoperable thesaurus representation formats by offering a standard convention for presentation. Existing thesauri can be transformed into SKOS format via conversion processes [261, 245, 193, 237, 282]. When a thesaurus is expressed as a SKOS vocabulary, it can be processed with standard RDF/SKOS tools in a uniform way.

There are also methods for converting thesauri into semantically richer OWL ontologies [262, 136, 44, 162]. Compared with SKOS conversion techniques, the OWL-based methods cannot be fully automated, as they require human effort for refining the semantic relations of the concepts [162]. Especially the *is-a* hierarchy of the ontology needs to be carefully constructed since the hierarchy of a thesaurus may have been built using a mix of different hierarchical relations [136, 162], which cannot be used as is for, e.g., subclass reasoning. The use of existing thesauri as the basis for ontologies enables the backwards compatibility with legacy data annotated with the thesauri, and facilitates the publication of the data as Linked Data.

This thesis seeks to develop publication methods for the cost-effective utilization of KOSs in, e.g., content indexing and information retrieval. In this context, SKOS is applied as a harmonizing model for representing

KOSs.

2.1.2 Interlinked Ontologies

In Linked Data paradigm, entities can be linked on many levels: data instances [112, 140, 103], metadata schema fields [267], and concepts of ontologies [286] can be interlinked to facilitate interoperability between datasets. Linking the data through ontologies allows additional interoperability due to the inferred knowledge gained through the shared ontology semantics [121]. When integrating datasets that use different ontologies (or KOSs), the ontologies need to be reconciled. Ontology reconciliation [104, 283] is a broad term, covering ontology merging, alignment, and integration. Most of the reconciliation methods are automatic or semi-automatic, which can lead to lower quality [32], especially if the ontologies were originally expert-made [73].

To faciliate the interoperability between different ontologies, there have been efforts to establish guidelines for the creation and management of ontologies, e.g., in the context of the OBO Foundry initiative [235]. The focus in OBO Foundry is on coordinating the development of different ontologies in the biomedical domain under shared principles. General, domain-independent ontology design principles have been proposed by several researchers [99, 260, 101, 191, 273, 89]. Linked Open Vocabularies (LOV) [267] is an effort on building a high quality catalogue of reusable vocabularies, and making their interconnections visible. Instead of KOSs, LOV focuses on metadata schemas.

Ontology linking methods are used also in ontology modularization [243, 7], where ontologies are divided into smaller interlinked parts to facilitate distributed development and re-use. There have been several efforts on building a general upper ontology [184] that can be used as a foundational basis for domain ontologies. Some of the upper ontologies have been developed from scratch, such as CYC [170], while, e.g., the Suggested Upper Merged Ontology SUMO [194] was created by merging existing ontologies.

For ensuring the consistency of interlinked ontologies, the changes of the ontologies have to be communicated to the dependent ontologies, e.g., by applying methods from the field of ontology evolution [281, 111, 169, 81, 128]. Methods include the detection of changes in an updated ontology by using logs [242, 157, 144] or comparing two versions of the ontology [164, 272]. To facilitate the processing of changes, different change types can be identified [246, 186], or more abstract change patterns can be constructed from atomic changes [160, 157, 144]. There has also been research on the nature of the change types the users are most interested in [37]. Also, the extra challenges of distributed ontology development [160, 241, 175, 146] have to be addressed when operating with interlinked ontologies.

This thesis aims to design methods and tools for managing and publishing an interlinked cloud of cross-domain ontologies in such a way that they can be utilized using shared user interfaces and APIs. The approach is based on modularizing the ontology development work into an upper ontology and domain ontologies extending it, and keeping track of their semantic dependencies.

2.1.3 Biological Nomenclatures and Taxonomies

Management of names and taxonomies of organisms in biology is an example use case for the need of a complex KOS that cannot be represented using simple, general KOS presentation languages, such as SKOS. In biology, taxonomy refers to the discipline that identifies, describes, and names groups of organisms (taxa) based on their shared characters [150]. The organism groups are organized into taxonomic hierarchies. Taxon names and classifications are important when integrating biological data from multiple sources [225, 201, 145, 253], and are therefore considered central resources, for example in biodiversity management [25, 200, 228, 215, 85, 86, 105, 206, 219]. The changing nature of the names poses challenges for their management [148, 166, 208, 229].

There are several issues that make the biological nomenclatures and classifications a suitable domain for the study of the modeling and publishing a rich KOS as an ontology. 1) Biological names are not stable or reliable identifiers for organisms as they or their meaning change in time. 2) The same name can be used by different authors to refer to different taxa, and a taxon can have more than one name. 3) Taxonomic knowledge is changing all the time and increases due to new research results. The number of new organism names in biology increases by 25,000 every year as new taxa to science are discovered [163]. At the same time, the rate of changes in existing names has accelerated by the implementation of molecular methods suggesting new positions to organisms in taxonomies. 4) The notion of 'species' in the general case is actually very hard to define precisely. For example, some authors discuss as many as 22 different ways of defining the concept of species [185].

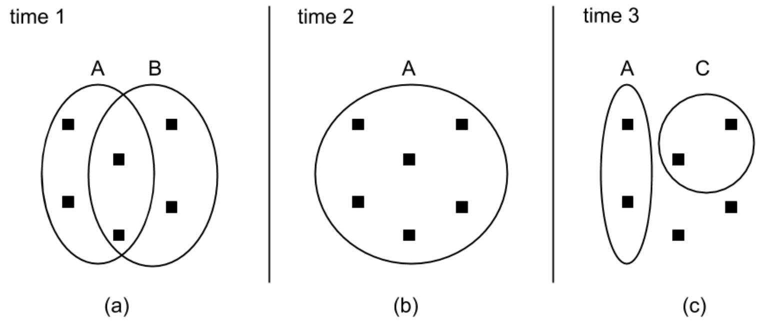


Figure 2.1. A hypothetical example of changes in taxonomic concepts and taxon names in time. First, two separate species A and B partly overlap (a). Then the two species are merged into a single species that has the name A, and the name B becomes a synonym to the name A (b). Finally, the species A is split into two separate species. The name A remains a valid name with a narrower meaning, and the name C is given to the new taxonomic concept. The black squares illustrate the biological characters of organisms, and the ellipses describe the limits of taxonomic concepts.

Although biological naming convention is regulated by nomenclature codes, e.g., International Code of Zoological Nomenclature (ICZN) [138] and International Code of Nomenclature for algae, fungi, and plants (ICN) [187], the names cannot be used as reliable identifiers when referring to taxa due to their ambiguity. Figure 2.1 depicts a typical series of changes in taxonomic concepts and their names. The border between the two species is unclear, and later the two species are merged into a single one, which is finally split into two (or more) species again. The taxon names may or may not change accordingly. Both the taxonomic concepts and their names change over time, and tracing the meaning of a name is impossible without a reference to a study. The reason for the continual changes is that every study has a different set of taxa, biological characters, and methods, and consequently their results are different.

A species checklist is a collection of names of organisms of a certain taxonomic group, compiled into a single taxonomic hierarchy by scientific experts. Checklists often present the species occurring in a particular geographical area. Comprehensive reference lists and catalogues of the names have been proposed as a solution to facilitate the access to the names and harmonize their usage [285, 105, 65, 208, 225, 173]. The need for such a list has been recognised, e.g., for vascular plants by the Convention on Biological Diversity (CBD) [5].

There are efforts on curating or aggregating taxon names from authoritative sources covering all species groups in the world, such as the Catalogue of Life (CoL) [223], the Encyclopedia of Life (EoL) [207], the Universal

Biological Indexer and Organizer (uBio) NameBank [225], WikiSpecies[1], the NCBI Taxonomy database [76], Open Tree Taxonomy [120], GBIF ChecklistBank [94], Index to Organism Names (ION)[2], and BioNames [203]. Other efforts are focusing on specific taxonomic groups or regions, such as ZooBank [215], the International Plant Names Index (IPNI) [55], World Register of Marine Species [51], Fauna Europaea [64], and the Atlas of Living Australia[3].

Berendsohn [25] introduced the concept of a ”potential taxon” to overcome the name ambiguity issues. A potential taxon is a combination of a taxon name and a literature reference to the taxonomic concept, that can be used, e.g., in databases for taxon references, enabling the interlinking of differing taxonomic views [26]. The importance of persistent identifiers for organism names has been further discussed by several researchers [200, 209, 225, 205, 219].

Pullan et al. [214] presented the Prometheus data model for managing taxonomic nomenclature and multiple related classifications separately, while Ytow et al. have developed the Nomencurator data model for representing and managing taxonomic nomenclature in a relational database [280]. Page [200] presented a simple data model for presenting taxon names and their relations, using using Life Science Identifiers (LSID) for identifying them. The use of LSIDs has been suggested also by organizations publishing taxonomic data [212, 56, 221], and piloted, for example in the Catalogue of Life database [148]. Further, Schulz et al. [228] presented an ontology model of biological taxa and its application to physical individuals. The model is based on a single unchangeable classification. Franz and Peet [85] formulated the use of semantics in relating taxa to each other, within a single taxonomic hierarchy and between two distinct hierarchies. Franz and Thau [86] evaluated the limitations of applying ontologies to the scientific names and concluded that ontologies should focus either on a nomenclatural point of view or on strategies for aligning multiple taxonomies.

As a use case for taxonomic ontologies, Lepage et al. [171] have implemented the Avibase database system for managing and organizing taxonomic concepts from major bird taxonomic checklists. There are also practical efforts on publishing taxonomic concepts as Linked Data, such

[1] http://species.wikimedia.org
[2] http://www.organismnames.com
[3] http://www.ala.org.au

as Taxonconcept.org[4] and Geospecies[5] that aim to provide Linked Open Data identifiers for species concepts and link them to related data from different sources. Chawuthai et al. [48] have presented an ontology model for managing the change of taxonomic concepts and publishing them as Linked Data.

In addition to data models that are focused on taxonomic information, there are metadata schemas for exchanging biodiversity data on broader scope, such as Darwin Core (DwC) [61, 277], the related Taxonomic Concept Transfer Schema (TCS) [251, 155], created by the Biodiversity Information Standards (TDWG), Access to Biological Collection Data (ABCD) [125], and Biological Collections Ontology (BCO) [271]. Darwin Core has a set of taxonomic extensions, Global Names Architecture (GNA) Profile [220], that introduces properties for richer nomenclatural details of taxa. Also, a semantically refined version of Darwin Core, Darwin-SW [22], has been proposed.

Methods of ontology versioning [161], evolution [195], and matching [233, 74] are relevant in management of taxonomic and nomenclatural information, as there exist multiple views on taxonomy and taxonomic knowledge changes as new research results are published. Thus, systems that support the usage of multiple versions of ontologies [128] and concepts [252] simultaneously are needed. There are approaches that are focused on life science ontologies, providing support for mapping ontologies [97, 110, 158], and systems for matching taxonomic concepts between databases [115, 202, 148, 51, 36, 218, 266, 84].

This thesis uses biological nomenclatures and classifications as an example use case of modeling and managing a richer KOS as an ontology. The focus is on practical management of the names and their changes.

2.2 Publishing and Using Knowledge Organization Systems

2.2.1 Ontology Servers

Once an ontology is modeled and serialized in some format, such as SKOS or OWL, it can be published for the wider community to be used as a shared domain model. In order to facilitate the usage of ontologies, on-

[4] http://www.taxonconcept.org
[5] http://lod.geospecies.org

tology servers [79, 68, 9, 108, 60] have been proposed for publishing ontologies and vocabularies on the web. Together with ontologies they have been considered a key resource for enabling the vision of the Semantic Web [136, 18, 108, 60]. The motivation for publishing ontologies using ontology servers instead of making them available as mere data files is to support the use of ontologies in applications. Ontology servers can provide ready-to-use services that can be integrated into information systems in a cost-effective way. Without such services, the user organizations have to implement the common functionalities for accessing ontologies in their own systems, leading to redundant work.

Parallel terms with ontology server are ontology library, ontology repository, and ontology service, each with a slightly different emphasis on the topic. In addition, the community of Networked Knowledge Organization Systems (NKOS)[6], that aims to develop web-based information services to support the description and retrieval of diverse information resources using KOS, uses the terms terminology registry and terminology services [95]. Common to these systems is that they are intended for publishing, managing, sharing, finding, and reusing ontologies and vocabularies for content indexing, information retrieval, content integration, and other purposes. Traditionally, the main focus in ontology server systems has been in supporting ontology development instead of the runtime usage of ontologies such as indexing and ontology-based end-user applications [68, 9]. The features of the systems vary greatly as they are designed for different purposes and based on specific user requirements [60].

An ontology server can support different phases in the ontology lifecycle, which can be defined as 1) design, 2) commit, and 3) runtime [9], or in a more fine-grained way as 1) acquisition, creation, and modification of vocabularies, 2) publication of vocabularies, 3) access, search, and discovery, 4) use, and 5) archiving and preservation of vocabularies [95]. The different lifecycle phases involve different user groups, which can be classified into ontology developers, ontology users, and application developers [60], or similarly in the NKOS community into KOS owners or creators, end users, and system developers [95]. The design phase covers tasks involved in ontology development, such as ontology engineering or editing, storing, versioning, mapping, and publishing. Once an ontology is published, the commit phase refers to the activity where a user is trying to find a suitable ontology for her needs, and needs support for discovering and evaluating

[6] `http://nkos.slis.kent.edu`

candidate ontologies. In the runtime phase, the user needs tooling for finding concepts for a given task, such as content indexing. Based on a survey of ontology servers, d'Aquin and Noy [60] have identified the key functions of an ontology server to be search, browsing, selecting and evaluating ontologies, and programmatic access to ontologies. Similar features have been recognized also by other researchers [108, 18, 95].

Most ontology servers are web-based systems that catalogue ontologies on a specific domain, such as biomedical sciences [235, 279, 196, 52], agriculture[7], oceanography [224, 167], government[8], by a specific organization, such as Library of Congress Linked Data Service[9], or with no such restriction. They provide access mechanisms to ontologies as user interfaces, APIs, or both of them [9, 60]. The user interfaces typically include search and browsing functionalities for finding ontologies and concepts in them. Search functionalities can be provided as string search based on the labels or other textual properties of the concepts, and utilizing the semantic relations of the concepts. Browsing interfaces typically visualize the structure of an ontology as a hierarchical tree or as a graph, where the concepts are presented as nodes and the relations as arcs between them. Ontology servers can provide users with listings of available ontologies, which are often classified based on different criteria. In some implementations, the set of ontologies can be further filtered and investigated using a faceted search.

Some ontology servers have been implemented for publishing a single, specific ontology, such as SUMO Browser [230] and GTAA Browser [40], whereas some systems focus on providing a directory-like listings of ontologies, such as DAML Ontology Library[10], Protégé Ontology Library[11], OBO Foundry [235], oeGOV, OntologyDesignPatterns.org [33], SHOE ontology library [113], Basel Register of Thesauri, Ontologies & Classifications (BARTOC) [168], and TaxoBank [122]. There are systems that focus on the ontology development and editing functionalities, such as iQvoc [20], PoolParty [226], VocBench [43], TemaTres[12], SKOS Shuttle[13], TopBraid Enterprise Vocabulary Net [255], SKOS editor [50],

[7] http://agroportal.lirmm.fr
[8] http://oegov.us
[9] http://id.loc.gov
[10] http://www.daml.org/ontologies/
[11] http://protegewiki.stanford.edu/wiki/Protege_Ontology_Library
[12] http://www.vocabularyserver.com
[13] http://skosshuttle.ch

PeriodO gazetteer [231], Neologism [21], Open Metadata Registry [213], WebOnto [69], Adapted Ontology Server [49], Ontolingua Server [75], and Medical Ontology Server [92].

Inter-ontology relations are considered important in several ontology servers, such as BioPortal [196], ACOS [172], MMI Ontology Registry and Repository [224], CATCH Vocabulary and alignment repository [264], ROMULUS [156], and Ontohub [190], as they support the creation and representation of concept mappings. There are systems that emphasize community-based aspects, such as uploading, rating and commenting on the ontologies. These ontology servers include, e.g., BioPortal, ACOS, and CupBoard [58].

There exist several search engines that crawl the web for RDF data and index them, such as Swoogle [67], Watson [59], and Sindice [198] for any data, whereas OntoSelect [41] and Falcons [217] are designed especially for ontologies. OntoSearch2 [254] is a similar ontology search engine, but it uses its own repository as opposed to crawling the public web. Such systems can be useful when searching for suitable ontologies to use in applications, and provide an overview of web-published ontologies in a specific domain.

Many ontology servers provide APIs for accessing the ontologies, typically for querying for ontologies and/or their concepts, and getting information about them. There are several specifications for APIs, such as Foundation for Intelligent Physical Agents (FIPA) Ontology Service [4] and Ontology Web Services (OWS) [57]. There are API implementations for processing ontologies in programming languages, e.g., the Java-based SKOS API [151], OWL API[14], and the APIs in DOGMA Server [141] and KAON Server [197]. For accessing ontologies on the web, there are several ontology servers that provide Web Service (SOAP) APIs, e.g., SKOS Web Service API [257], OCLC Terminology Services [269], Ontology Lookup Service (OLS) [52], Watson, CATCH Vocabulary and alignment repository, and NERC Vocabulary Server [167]. A more recent approach is to provide access to ontologies through a RESTful HTTP API, such as in the case of SISSVoc [54], OCLC Terminology Services, Otago Ontology Repository [204], BioPortal, iQvoc, PoolParty, NERC Vocabulary Server, HIVE [96], TemaTres, and SKOS Shuttle. SPARQL [107] is the standard way to provide an application interface to Semantic Web databases, and it can be used also to access ontology repositories. Similarly, general RDF

[14] `http://owlapi.sourceforge.net`

libraries can be used for processing ontologies, such as Apache Jena[15] and RDFLib[16]. Ontology servers such as OCLC Terminology Services and BioPortal provide widgets, user interface components that can be integrated into applications, to enable ontology-based search functionalities in applications.

To facilitate simultaneous access to several ontology servers, shared access mechanisms and protocols are needed. Open Ontology Repository (OOR) [18] is an initiative that aims to specify an architecture and interfaces for interoperability between ontology repositories. Also other researchers have emphasized the importance of interoperability between ontology repositories [60, 108]. Ontohub is an ontology repository engine that follows the ideas of the OOR initiative, and provides inter-repository access by defining a generalized federation API that needs to be implemented in the participating repository or as a wrapper around the legacy API of the repository. Common Ontology API Tasks (OntoCAT) [8] is a programming interface to query multiple ontology repositories seamlessly from an application. The system is based on wrappers that are implemented for each supported ontology repository. OLS2OWL [90] is a plugin for Protégé ontology editor enabling simultaneous queries to multiple ontology servers by using a similar wrapper approach as in OntoCAT. JSKOS-API [168] is a general HTTP API for accessing knowledge organization systems, with methods for concept search and lookup. By implementing the API in multiple ontology servers or using wrappers, it is possible to provide inter-repository search and browsing interfaces. There are also general protocols for accessing knowledge bases, such as the Open Knowledge Base Connectivity (OKBC) [47], and agent communications languages FIPA-ACL [3], the Knowledge Query and Manipulation Language (KQML) [80], and the Semantic Agent Programming Language (S-APL) [152].

Regarding the different aspects of ontology servers, the focus of this thesis is to provide publication channels for ontologies in different formats and support for their runtime use, e.g., in a network of distributed content creation. One of the design principles is the support for the use of multiple ontologies and ontology repositories simultaneously.

[15] `https://jena.apache.org`

[16] `https://rdflib.readthedocs.io/en/stable/`

2.2.2 Semantic Annotation and Information Retrieval

The typical use cases for ontologies and other knowledge organization systems are semantic annotation and information retrieval. Both these tasks can be facilitated with ontology services. In ontology-based manual annotation human cataloguers create content descriptions using ontological concepts [159, 199, 16]. The annotation process is usually guided by indexing guidelines and conventions that may be general [1, 15] or shared by particular disciplines or organizations. The created metadata can be based on a specific schema, which can be a simple collection of key-value-pairs, such as Dublin Core [63], or a semantically richer model with relations between the individual metadata fields [250, 227]. The annotation task can be defined as a process of analyzing the content to be annotated, finding relevant ontological concepts from the selected ontologies, and storing them in metadata fields. The process can be streamlined and made more effective with different kinds of automated tools [17].

Based on a user study, Hildebrand et al. [118] have identified the following use cases of a human annotator for a concept search:

- The user already knows the concept she would like to use as a descriptor for the content, and wants to find the concept from the used thesauri.
- The user does not know the most suitable concept for the content description beforehand and she needs to examine the thesauri to find one.
- The user suspects that the used thesauri do not contain a concept she would need for the task, and she needs to ensure this before she adds a new concept into one of the thesauri.

The human annotator's task to find the best matching concepts for her needs can be aided by providing concept search and browsing functionalities. These general functionalities can be provided by ontology servers as APIs and user interface components that can be used for integrating them into applications. The importance of such services for thesaurus use, sharing, and interoperability has been emphasized by several researchers [95, 183, 216, 284, 30, 119, 124]. Such an approach relates to the notion of service-oriented architecture (SOA) [231, 70, 154], where software components are provided as technology independent services to other applications to be used over a network through a communication protocol.

Regarding the ontology services discussed in the previous section, OCLC Terminology Services provides a widget for querying controlled vocabularies and displaying information about their terms in the sidebar of the Internet Explorer browser, and transferring selected terms into a web-based cataloging application [269]. BioPortal provides web widgets that can be integrated into web applications, for concept search, selection, and visualizing ontologies [196]. The use of SKOS Web Service API as web widgets has been demonstrated in the STAR project by providing functionalities for concept search, expansion, and presentation [31].

Hildebrand et al. [117] have implemented a configurable autocompletion search widget for RDF repositories, based on which Amin et al. [14] have conducted a user study on the organization strategies for the autocompletion suggestions to provide effective means of navigation and finding relevant terms. Further, Hildebrand et al. [118] performed a user study in which museum professionals used the widget for annotation. They reported on different strategies for matching, sorting, grouping, and displaying contextual information for the autocompletion suggestions. Malaisé et al. [179] conducted a user study on expert annotators using GTAA Browser, and concluded that the users mainly used the alphabetical search functionality to find relevant concepts, whereas the concept hierarchy browser was not used that much.

The ability of the information retrieval systems to produce relevant search results depends on the user's ability to represent her information needs in a query [270]. If the vocabularies used by the user and the system are not shared, or if the vocabulary is used in different levels of specificity, the search results are usually poor. Query expansion has been proposed to solve these issues and to improve information retrieval by expanding the query with terms related to the original query terms [45]. Query expansion can be based on a corpus, e.g., analyzing the co-occurrences of terms, or on knowledge models, such as thesauri [83, 274] or ontologies [270]. Methods based on knowledge models are especially useful in cases of short, incomplete query expressions with few terms found in the search index [270, 274].

Ontology-based query expansion can be used interactively to guide the user to formulate his query, for example by providing an autocompletion text search for disambiguating and selecting ontological concepts [132], and automatically by adding concepts to the initial query based on their ontological relations [34, 123, 29, 180, 149]. Typical relationships used in query expansion are the synonym, hyponym, hypernym, and associative relations [14, 126, 147, 142, 72]. When considering general associative relations, caution should be exercised as their use in query expansion can lead to an uncontrolled expansion of result sets, and thus to potential loss in precision [256, 126]. In an experiment by Navigli and Velardi [192], it was noted that extracting the query expansion terms from the sense definitions of the query terms from an ontology produced better results than using the taxonomic relations (e.g., synonym, hypernym). Ontology-based query expansion can also be used for cross-language information retrieval [45], and in addition to general-purpose ontologies, domain-specific ontologies can be utilized, e.g., for spatial query expansion [88]. A concrete example of a terminology service with query expansion support is the FACET Web demonstrator [30] that provides a web service and user interface for accessing thesauri.

This thesis aims to develop practical solutions and tools for ontology-based content annotation and information retrieval. The common user tasks of such workflows are catered by providing user interface components and APIs that can be integrated into external applications.

3. Results

In the following, the results to the research questions of this thesis are presented. Furthermore, the results are reflected against previous research in Chapter 4.

3.1 Ontology Services (RQ1)

The research question 1 concerns publishing thesauri in such a way that they can be easily and cost-effectively used in ontology-based workflows, especially in content indexing and information retrieval.

1. How can lightweight ontologies be published on the Semantic Web so that they can be utilized in content indexing and information retrieval tasks?

The publications I–III provide solutions for this question by presenting ontology services. The main idea is to publish the ontologies not only as data, but as services that can be used by humans and machines for integrating ontology-based functionalities into applications, provided as user interfaces and application programming interfaces (API). The functionalities of the developed ontology services were developed based on the analysis of the user groups of ontologies, acting in different phases of the life cycle of an ontology.

The main user groups of ontologies were identified as 1) ontology developers, 2) content creators, 3) information searchers, and 4) software developers. In Publication I, their needs for using ontologies are classified into the following tasks.

1. **Designing the ontologies.** The structure and modeling principles of an ontology are developed based on the analysis of the subject domain

and the use cases of the ontology. Ontology developers need tools for creating the ontology, collaborative editing, reuse, and alignment.

2. **Populating the ontologies.** An ontology may contain a high amount of instances, e.g., people, organizations, and places. To save efforts, they can be harvested from existing sources, or collected from the end users. Tools may provide support for content collection and updating processes.

3. **Publishing the ontologies.** To promote the use and reuse of an ontology, the ontology owners can make it accessible to the users. Prior to publishing the ontology, its quality can be ensured with manual and automatic methods.

4. **Finding, comparing, and committing to ontologies.** When there is a need to use an ontology in an information system, the information architect of the system needs support for selecting a suitable ontology for her needs.

5. **Ontology-based semantic application creation.** Application developers need to learn, evaluate, and apply methods for integrating ontologies into applications, e.g., by using user interface components and APIs.

6. **Ontology-based semantic content creation.** The work of content indexers can be facilitated by providing tools, e.g., for browsing the ontologies, searching and selecting concepts, and (semi-)automatic indexing.

7. **Ontology-based end-user applications.** Application developers can utilize an ontology for building end-user applications, such as semantic portals, to facilitate information findability. The ontology can function as an educational resource for end users to learn about its domain.

The common functionalities for utilizing ontologies in ontology-based applications of different kinds were identified as concept search, browsing, and selection. The developed user interfaces, widgets, and APIs are designed to support these basic tasks. The proposed ONKI ontology service is based on several implementations of ontology servers suited for ontologies of different kinds due to distinct needs for accessing them. For example, a natural way to display a thesaurus is a tree-like visualization, whereas the users of geographical ontologies may prefer map-based user interfaces. Publication II presents an ontology server for thesaurus-like, lightweight ontologies, the ONKI SKOS system. The system supports publishing of syntactically, semantically, and structurally differing thesaurus formats, with the requirement that the thesaurus has to be in RDF format and have some basic structure including concepts, their labels, and possible inter-concept relations.

Existing thesauri in legacy formats can be converted into W3C's SKOS data model, which acts as a harmonizing model for expressing knowledge organization systems. Such thesauri can be published in the ONKI SKOS server, which then provides functionalities for the users of the ontology. Thus, organizations developing thesauri do not have to implement their own thesaurus-specific publication systems and the users can use different thesauri with shared tools. The real-life benefits of using the ONKI SKOS server have been demonstrated by applying the system to use cases of a) content indexing in health promotion and cultural heritage context, among others, and b) information retrieval in the collections of the Finnish museums of forestry. The ONKI service enables content creation in a distributed network of organizations, where each organization uses shared ontologies and ontology services for accessing them, thus harmonizing the created metadata and facilitating information integration.

Publication III gives a detailed view on how the query expansion facilities of the ontology service are used in practice in information retrieval scenarios. The query expansion widget uses the semantic relations of the ontology to refine the query with additional query terms to increase the recall of the search. For example, if the user is searching for ”animals”, the query can be expanded to include also ”cats” and ”dogs” based on the concept hierarchy.

As a proof of concept for the process of converting a legacy thesauri into the SKOS data model and publishing it in the ONKI SKOS service, the case of Finnish General Thesaurus YSA, is reported. YSA has been developed in the National Library of Finland since 1987 and is widely used in libraries, museums, and archives in Finland. In addition to YSA, over 80 national and international vocabularies, taxonomies, and ontologies are published in the ONKI SKOS system.

The user interface of the ONKI service, including the ontology directory, search view, and ontology browser have been developed in an iterative process where different versions of the system have been published and made accessible online as ONKI1 (Publication I), ONKI2 [258], ONKI3 [12], and ONKI Light [248]. The ONKI system has been run as a living lab ontology service since the official announcement in the autumn of 2008, and before the successor service, Finto[1] of the National Library of Finland, was publicly released in January 2014, the ONKI service had over 10 000 monthly users (excluding the widget and API users), with 400 registered user domains for the widget and API use.

3.2 Publishing Multiple Ontologies (RQ2)

The research question 2 extends the research question 1 by introducing the dimension of multiple ontologies and ontology repositories used simultaneously.

2. How can a collection of independent or interconnected ontologies—in different formats and repositories—be published and utilized using shared user interfaces and APIs?

Such a multi-ontology scenario is prevalent in many cases, for example when indexing content with more than one ontology, or when trying to find and choose a relevant ontology for a specific use case. The publications IV and V present methods for solving issues in such use cases.

Publication IV discusses the publication of an ontology cloud consisting of individual, interconnected ontologies. The system is based on an upper ontology and domain-specific ontologies extending it. The publication process involves merging the component ontologies into a single, coherent representation, and using the ONKI service to publish it to end users. The structure of the ontology cloud appears as a single whole to the users, without emphasizing the ontology boundaries. Thus, the users should be able to use it in a straightforward way, for example in content indexing, focusing only on the main task of choosing relevant concepts, not ontologies.

The motivation for building such an ontology cloud is to facilitate the data integration of heterogeneous datasets from different domains, using domain-specific ontologies. The approach complements other existing techniques in data integration on the Semantic Web—data entity linking in Linked Open Data (LOD) and metadata schema linking in Linked Open Vocabularies (LOV). The formation of an ontology cloud can be more efficient since the mappings between ontologies can be reused for different datasets. Using an upper ontology as the base for the cloud instead of mapping individual ontologies on a one-to-one basis eliminates redundant

[1] http://finto.fi

mapping work.

For building such an ontology cloud for end users, ontology development and management processes need to take into account the semantic dependencies between the individual ontologies. This includes the identification of conceptually overlapping parts of the ontologies to avoid redundant development work, and the communication of the changes of the upper ontology to the domain ontologies extending it. The feasibility of the approach was demonstrated in practice by building the ontology cloud KOKO of more than 47,000 concepts, based on the Finnish General Finnish Ontology YSO and 15 domain ontologies. The ontologies were developed by a network of domain experts, where each organization took responsibility for managing a single domain. Based on the experiences in building KOKO, a set of seven principles guiding the building and management process of the cloud, was formed. The principles are general enough to be applied to other ontology clouds as well. The principles aim to streamline the ontology development process and ensure the semantic integrity of the resulting cloud, especially concerning the transitive subclass hierarchies of concepts. The developed approach of the proactive linking of ontologies as part of their development phase instead of mapping them afterwards aims to minimize redundant work and maximize interoperability.

Publication V discusses an environment consisting of several ontology repositories, where users need to access the repositories concurrently. The proposed Normalized Ontology Repository (NOR) approach allows accessing ontology repositories using shared tools and user interfaces based on a) harmonizing the representation of concepts in ontologies by using the SKOS data model, and b) providing a uniform API that encompasses the general ontology access needs, such as functionalities for getting the metadata of ontologies in the repository, searching for concepts, and querying their properties.

Based on the approach, it is possible to give the user an overview of the ontologies available in a set of ontology repositories. This helps the user to choose a suitable ontology repository or a specific ontology for her needs. The user can browse the ontologies using a uniform browser view, eliminating the need for learning to use ontology-specific user interfaces. The NOR API and concept representation is an extra layer on an ontology repository, meaning the functionalities of the repository are not restricted in any way. When the user has found a suitable ontology, she can move from the uniform NOR browser to the possible own, specialized user interface provided by the underlying ontology repository. This mechanism is motivated because different ontologies might benefit from access mechanisms of different kinds, such as user interfaces and APIs. The system also allows the simultaneous usage of public ontology repositories and private repositories of organizations.

To demonstrate the feasibility of NOR, it has been applied in real-life use scenarios. The ONKI ontology repository itself uses such an approach for providing a uniform user interface for searching and browsing over 80 vocabularies and ontologies published in the ONKI SKOS backends. The backends encompass ontologies in different RDF-based formats, such as SKOS and RDFS, which are accessible via an HTTP API that provides a concept search functionality and normalizes the representation of the concepts. The ONKI frontend provides users with a possibility to perform a global search to all the ontologies at the same time. The system also includes a directory listing of all the ontologies available and offers a faceted search view to them, so the user can find such ontologies, for example, whose subject domain is ”business” and are published by ”FinnONTO Consortium”. The directory listing is built using a uniform metadata representation of the ontologies in the system. ONKI Widget uses the NOR approach in an even more heterogeneous environment, by providing users with access to not only ONKI SKOS backends, but also to the ONKI Geo ontology server [131], offering access to the geographical ontology of Finnish contemporary place names. The approach has also been tested by implementing a metasearch prototype for accessing the ontologies of ONKI and BioPortal repositories simultaneously, and even for accessing more general data repositories than ontology repositories, the CultureSampo portal [178] and SAHA metadata editor [165].

The relationship between the ontology cloud and NOR methodologies is a complementary one. The ontology cloud enables interoperability on the ontology level, whereas the NOR approach is focused on compatibility on the ontology service level. Building an ontology cloud requires mapping effort between the domain ontologies and the general upper ontology, and thus makes evident the relations between datasets described using different domain ontologies. On the other hand, in the NOR approach the ontologies are presented using a harmonized data model, but there is no need to map the ontologies to each other on the concept level. Thus, the system can be used for simultaneously accessing and processing even a set of mutually unrelated ontologies with shared tools.

3.3 Complex Knowledge Organization Systems (RQ3)

The research question 3 concerns the applicability of the Semantic Web technologies to managing richer KOSs as ontologies and publishing them using ontology services.

3. How can richer knowledge organization systems, such as biological nomenclatures and classifications, be managed as an ontology and published using an ontology service?

As a case study, the publications VI–VIII present the modeling and management of biological names and classifications. The proposed solution is an ontology model for taxonomic concepts and their scientific and vernacular names. Publication VI presents the Taxon Meta-Ontology (TaxMeOn) model which is aimed for the following data: 1) species checklists and mappings between them, 2) vernacular name collections, and 3) the changes of scientific names and classifications, and differing opinions of taxonomic concepts based on biological research results. The model makes a distinction between the taxonomic concept and its name. Thus, they can be managed separately, and the nomenclatural changes and re-classifications of a concept can be tracked and managed.

The model is flexible in a sense that it is designed to be suitable for data with different levels of details. The simplest use case is to express a static list of taxon names, but the model also supports more complex needs of representing the changes of names and taxonomic concepts. As the model is based on Linked Data, it offers possibilities to link divergent data serving divergent purposes and detailed information with more general information.

An advantage of the model is its practicality and applicability to real-life use cases. This is demonstrated by applying it into three use cases: 1) publishing biological species checklists in an ontology service (27 lists, over 80,000 names), 2) collaborative management of vernacular names (ca. 26,000 taxa), and 3) management of individual scientific name changes resulting from biological research results (9 genera).

Publication VII presents the use case of applying TaxMeOn into modeling and publishing species checklists of scientific names, and compares the ontology model with storing the checklists in a legacy relational database. The model allows mapping of taxa between different checklists (e.g., based on their congruency) and representing and managing the changes in taxonomic concepts, their classifications, and names in individual checklists. The main advantages of the ontology model as opposed to a traditional database are the linkability to other datasets, extendability of the data model, (re)usability of the data via standard publication mechanisms, and possibility to edit the data with standard RDF tools. This means that the ontologies can be utilized in applications using general ontology services, without the need to implement domain-specific access mechanisms for biological name collections.

The species checklists were published in the ONKI ontology service, facilitating their reuse via user interfaces and APIs. The ONKI browser interface can be used for searching and browsing taxa, finding currently valid names, and tracing the temporal changes in the scientific names. The ONKI autocompletion widget provides a way for integrating an access mechanism to checklists into user applications, e.g., a content management system. Furthermore, HTTP and SOAP APIs are available for programmatic access and a SPARQL endpoint for querying the ontologies.

Publication VIII gives a more thorough presentation on how the TaxMeOn model can be applied into management of vernacular names. The model provides a solution for managing the approval process of common names, supporting the temporal tracking of their changes via statuses and their time stamps. The system is used by the Finnish Biology Society Vanamo[2] to manage the Finnish names of vascular plants in a collaborative way. In the typical workflow, a new common name is first proposed for a plant, after which it can be accepted to be the recommended name, and finally it can be made an alternative if another recommended name is introduced later.

We present the complete workflow for managing a vernacular name ontology from a collaborative development of the ontology to publishing it as Linked Open Data and in an ontology service which makes it accessible to the general public. The ontology is available in machine-processable RDF format, with explicit semantics, e.g., the hierarchical relations are set between the plant URIs, facilitating data integration and information retrieval in cases where data is combined from heterogeneous sources. The plant name ontology helps harmonizing the terminology, which in turn enhances communication between various users. Application developers can utilize the ontology by using the plant name URIs for unambiguous referencing to plant species.

[2] http://www.vanamo.fi

The applicability of the TaxMeOn model for its most complex use case, the management of individual scientific name changes based on biological research, has been demonstrated with a test dataset representing changes in the taxonomic classification of Afro-tropical beetle family Eucnemidae in Publication VI. The family has gone through numerous taxonomic treatments. For example, the position of the species *Pterotarsus historio* in the taxonomic classification has changed 22 times and at least eight taxonomic concepts are associated to the genus *Pterotarsus*. The TaxMeOn representation of the dataset encompasses the different conceptions of a taxon (e.g., *Pterotarsus*), the temporal order of the changes, and the references to scientific publications whose results justify these changes. Such a detailed information source provides a unified view of a complex taxon, which can be beneficial even to the researchers of biology, as the details of taxa have traditionally been scattered across the original publications, and piecing them together can be difficult and time-consuming. The detailed data can be further linked to other datasets with less taxonomic information, such as species checklists, which provides their users with more precise information.

3.4 Summary

To summarize the results, the research questions are re-visited in this section.

1. How can lightweight ontologies be published on the Semantic Web so that they can be utilized in content indexing and information retrieval tasks?

To faciliate the use of ontologies, they should be published as ontology services to fulfill the needs of different user groups, supporting their workflows during the phases of the life cycle of an ontology, e.g., in content indexing and information retrieval. For the cost-efficient reuse of the ontologies, user interface components and APIs can be used to integrate ontology-based functionalities into applications. The common functionalities for ontology use include concept search, browsing, and selection. W3C's SKOS data model can be used for harmonizing different legacy thesauri, which facilitates their publication via shared mechanisms in ontology services. Ontologies of different kinds might benefit from differing user interfaces, e.g., a thesaurus can be visualized as a tree, whereas a geographical ontology might employ a map view.

2. How can a collection of independent or interconnected ontologies—in different formats and repositories—be published and utilized using shared user interfaces and APIs?

A set of interconnected ontologies can be combined into an ontology cloud that can be made accessible to end users via an ontology service as a single representation encompassing the different domains of the member ontologies. For building such an ontology cloud, ontology development and management processes need to take into account the semantic dependencies between the individual ontologies. The ontology development can be streamlined by formulating practical ontology design principles that guide the work and ensure the consistency of the cloud. The use of multiple ontology repositories simultaneously can be accomplished by using a shared upper data model for the ontologies, e.g., SKOS, and providing a shared API for accessing the ontologies. The approach facilitates, e.g., the building of aggregated ontology directories and global search on a network of ontology repositories.

3. How can richer knowledge organization systems, such as biological nomenclatures and classifications, be managed as an ontology and published using an ontology service?

Designing an ontology model for the specific needs of the domain, and mapping it to a harmonizing data model, e.g. SKOS, allows the publication of the ontology in shared ontology services. This way, the ontology can be accessed with general ontology user interfaces and APIs, without losing the detailed representation of the information. In the case study of biological names and classifications, the main modeling solutions in the TaxMeOn model are the separation of the taxonomic concepts and names, representation of the changes and their temporal order, and mappings between the different conceptions of the taxonomic concepts. The management and publication workflow of the ontology can be implemented using a general RDF editor and ontology service.

4. Discussion

Traditional, comparative evaluation of the models, processes, and tools developed in this thesis is difficult, as the systems provide novel solutions for the problems described in Chapter 1. In particular, evaluation of Semantic Web applications is difficult as the usefulness and usability of the systems depend on multiple factors: the quality of the heterogeneous source data used, the underlying search and inference software, and the user interface [265]. In more mature fields of computer science, such as in relational databases or text-based information retrieval, established data models and search algorithms are often available to be used as building blocks for creating new methods. In Semantic Web, such existing components are rare, and they usually have to be designed for every application. The state adds complexity to the development process and requires a level of maturity from the system to be properly evaluated. Unless all the components are of high quality, the system is not useful for the user.

As the evaluation method in this thesis, the extensive application to practice has been used as a proof of concept with the developed artifacts being adjusted based on real-life experiences in accordance with the principles of action research. Burstein and Gregor [42] have proposed criteria for evaluating systems development research, covering the following aspects: 1) theoretical and practical significance, 2) internal validity, 3) external validity, 4) objectivity, and 5) reliability. In the following, the research of this thesis is evaluated according to these criteria.

4.1 Theoretical Implications

4.1.1 Ontology Services (RQ1)

In comparison to the earlier ontology server research presented in Chapter 2, the main contribution of the ONKI ontology service model is the tight integration and support for the runtime use of ontologies, focusing on the content indexing and information retrieval use cases. Many previous efforts have been focusing on development and editing capabilities [20, 226, 43, 50, 231, 21, 213, 69, 49, 75, 92] of ontology servers, or merely publishing ontologies via ontology directories [235, 33, 113, 168, 122] or ontology browsers [40, 230] without providing modular components or web services that can be integrated into external applications. The developed system is domain-agnostic, general solution for publishing and using ontologies, and thus is not restricted to a single KOS or domain, as opposed to several other ontology servers. Compared with general semantic web search engines [67, 59, 198], the ONKI system provides focused support for ontology-based tasks, e.g., by displaying concept hierarchies. Ontology search engines [41, 217, 254], on the other hand, are tools suited especially for finding suitable ontologies, but not for using individual ontologies, e.g., in content indexing.

ONKI widget for integrating concept search and selection functionalities into external applications is similar to the widget models and implementations of the OCLC Terminology Services and BioPortal. The ONKI approach is more general than the OCLC widget, as ONKI widget can be integrated directly into the user interface of the application, and is not used as a separate browser toolbar. ONKI widget aims to provide a streamlined user experience, where the ontology-based functionalities are served with as little interference with the original user interface as possible. ONKI widget [268] was published earlier than the BioPortal widget [275], to the best of the author's knowledge[1]. The autocompletion component by Hildebrand et al. for general RDF repositories is based on similar ideas as ONKI widget. However, ONKI widget is focused on ontology-based interactions and is packaged as a ready-to-use service. One of the novelties of ONKI widget is the combination of the autocompletion

[1] The history of the NCBO Widgets (BioPortal) documentation page `https://www.bioontology.org/wiki/index.php/NCBO_Widgets` dates to 12 May 2009.

search and ontology browsing mechanism. Thus, the user can start the concept search task by typing in a query term, select a matching concept, and further refine the selection by browsing the ontology, e.g., to find more specific concepts based on the concept hierarchy. The system also supports semantic query expansion in similar vein as piloted in the FACET Web demonstrator and STAR project widgets, but has been made publicly available as a widget that can be integrated into external applications.

The HTTP and SOAP API of the ONKI service provide high level abstraction access to ontologies with a compact API specification, focusing on supporting concrete use cases of content indexing and information retrieval. The APIs are web-based, meaning they can be utilized in distributed systems, promoting loosely coupled services and complying with the resource- [78] and service-oriented [70] architectures. The use of the APIs is not tied to a single ontology modeling or programming language, as opposed to implementations such as SKOS API and OWL API, or general RDF libraries, such as Jena and RDFLib. Compared with the general RDF query language SPARQL, ONKI is focused on providing ontology-based functionalities on a more abstract level.

The common functionalities of an ontology server identified in this study based on the analysis of the user requirements of KOS—the concept search, browsing, and selection—are supported by previous literature [60, 108, 18, 95]. Also, the decision of using the SKOS vocabulary as the harmonizing model for publishing KOSs is reaffirmed by existing research [261, 237, 54, 50].

4.1.2 Publishing Multiple Ontologies (RQ2)

The ontology cloud model presented in this thesis is based on the idea of expressing multiple interlinked ontologies as a single coherent system, published in an ontology service. Previous research on the inter-ontology support in ontology servers has been focused on the creation and representation of concept mappings between individual ontologies [196, 172, 224, 264, 156, 190], and not providing them as a shared, easy-to-use, cross-domain ontology for use cases such as content indexing. The model is supported by previous research on upper ontologies [184, 170, 194] and ontology modularization [243, 7], where the ontology content is divided into subsets based on the generality and domain of the concepts.

The presented principles and methods for the creation and management of the ontology cloud complement the previous guidelines for ontology interoperability [235] and general ontology design principles [99, 260, 101, 191, 273, 89], by emphasizing the importance of the consistency of the concept hierarchy. The proposed methods and tools for identifying the overlappings of the participating ontologies and propagating the changes of the upper ontology to the domain ontologies aim to ensure that the ontology cloud is valid, up-to-date, and easy to use. The cloud model is based on utilizing existing legacy thesauri, and as such the system maintains the backward compatibility with existing annotations, providing a cost-efficient way to publish legacy data as Linked Data.

The NOR approach of accessing multiple ontology repositories simultaneously is based on a distributed architecture, whereas many previous ontology server models are centralized services. The system is not tied to a single ontology server implementation, but is based on a shared API instead. The NOR API is comprised of access methods to ontologies and concepts on a high abstraction level, with a focus on the concept search and representation of concept information and ontology metadata. The API is not based on a specific ontology language, as opposed to lower level APIs, such as SKOS API and OWL API. On other hand, in contrast to general knowledge and agent communication languages, such as OKBC, FIPA-ACL, KQML, and S-APL, NOR API is focused on practical use cases of content indexing and information retrieval. To avoid the building of extraneous wrappers, which are used in many federated ontology access systems [8, 90], NOR is designed to be lightweight and simple in order to be easy to implement in ontology servers. The approach of using a highly abstracted API and harmonizing metadata model is similar to the more recent approaches of Ontohub and JSKOS-API, of which the latter uses the same SKOS data model and basic methods of entity search and lookup as NOR API. The ontology metadata used in NOR utilizes the existing metadata models VoID [13], Dublin Core, and FOAF [38], complements them by adding information of the NOR endpoint address, and can be extended with other ontology and dataset description vocabularies, such as Ontology Metadata Vocabulary (OMV) [109] and Data Catalog Vocabulary (DCAT) [174].

4.1.3 Complex Knowledge Organization Systems (RQ3)

The developed TaxMeOn model for managing biological nomenclatures and classifications is an example of how a rich KOS can be modeled and maintained as an ontology and published in an ontology service. Compared with traditional species databases that aim to aggregate taxon names from various sources and harmonize their usage [223, 207, 225, 76, 94, 215, 64], TaxMeOn provides an explicit data model that is a general solution for managing heterogeneous biological name collections, and not tied to a single database system. As the model is based on Linked Data technologies, the data model can be extended in a flexible way and integrated with other data sources. The use of URIs as global identifiers in TaxMeOn is supported by the previous research which has emphasized the need for persistent identifiers when referring to taxa [25, 200, 209, 225, 205, 219], either in the form of a taxon name combined with a literature reference or a technical string.

Compared with previous ontology models on taxonomic information [228, 85, 86, 48], TaxMeOn is focused on practical name management of species checklists, research results, and other nomenclatural collections. The model supports the management of parallel classifications and nomenclatural conceptions of taxa in a semantically rich way, whereas some of the previous research have concentrated on modeling a single, unchangeable classification [228, 76]. Many Linked Data publishing projects of biological data and biodiversity data models [61, 251, 125, 220] are not focused on semantic rigor and therefore do not promote the machine processability of the contents optimally.

4.1.4 Summary

The theoretical implications of the methods and tools presented in this thesis are summarized in Table 4.1, by re-visiting the objectives of the ontology services defined in Section 1.2.

4.2 Practical Implications

4.2.1 Ontology Services (RQ1)

ONKI service provides out-of-the-box support for publishing and utilizing SKOS vocabularies and other lightweight ontologies in, e.g., content indexing, without needing to implement application specific user interfaces for end users. The system caters for many common, sharable tasks in ontology-based applications related to, e.g., concept finding, browsing, selecting, and query expansion. Lots of work and costs can be saved by implementing

Objective	Methodological and technological solutions
Ontology publication channel	General publication channel for various ontologies, created by different parties, including user-uploaded ontologies and ones fetched from external sources by automated processes.
Heterogeneous ontologies	Support for ontologies in different RDF-based formats, not tied to a single domain, modeling language, or a publisher.
Tools for metadata creation	Widgets and APIs that can be integrated into applications to support ontology-based cataloging practices.
Support for distributed content creation	Public living lab ontology service that can be used by a heterogeneous network of memory organisations for creating interoperable metadata based on shared ontologies.
Facilitate search tasks	The widgets and APIs support information retrieval by providing query expansion and cross-language search facilities.
Multiple ontologies and repositories	Publication of interlinked ontologies as a coherent, cross-domain cloud, and ability to perform federated search and uniform access to multiple ontology repositories based on a harmonized data model and shared API.
Programmatic access	HTTP and SOAP APIs for common ontology-based tasks, including concept search and lookup.
Evaluation by applying into practice	The feasibility of the developed ontology services has been demonstrated by extensive piloting in diverse use cases.
Promote complex KOSs	Rich knowledge organization systems can be mapped to a harmonizing data model, such as SKOS, and published in shared ontology services. The management of such a KOS can be realized using a general RDF editor.

Table 4.1. The objectives of the ontology services and the corresponding solutions presented in this thesis.

such functionalities in standard ways and providing them for production use as ready-to-use services. In this way, the use patterns of utilizing vocabularies in the user interface can be harmonized, which makes the systems easier to learn and use, as there is no need for vocabulary-specific access mechanisms. The ONKI service provides simple, yet powerful APIs and an autocompletion widget for content indexing and query expansion. The services can be used not only in ontology-based applications, but in legacy systems, which can utilize the ontologies in a similar way as traditional thesauri.

ONKI has been run as a living lab service since 2008 and has acted with the KOKO ontology cloud as the backbone of the Finnish national ontology infrastructure [136], which aims to enhance the interoperability of the collections of museums, libraries, archives, companies, and other organizations. Also, several international ontologies, such as Iconclass[2], Medical Subject Headings (MeSH)[3], the United Nations Standard Products and Services Code (UNSPSC)[4], and Integrated Public Sector Vocabulary

[2] http://www.iconclass.nl
[3] http://www.nlm.nih.gov/mesh/
[4] http://www.unspsc.org

(IPSV)[5], have been published in ONKI to facilitate their use in applications. The ONKI service has been used in the distributed ontology-based content creation approach of several semantic portals, such as CultureSampo, HealthFinland [247], and BookSampo [177]. The ontology infrastructure has gained maturity in Finland, and through technology transfer, ONKI's production level successor, Finto service has been run by the National Library of Finland since 2014 [249].

4.2.2 Publishing Multiple Ontologies (RQ2)

The design principles and tools presented in this thesis for building and managing a cloud of interlinked ontologies have been used to build the cross-domain KOKO cloud of Finnish ontologies. The ontologies are based on existing, established thesauri that have been used in various organizations for describing heterogeneous contents. The users of the legacy thesauri have been shifting to use the KOKO cloud, which facilitates the cross-domain interoperability of their datasets, as the novelties of KOKO include the mappings between the individual domain ontologies and the semantic consistency of the concept hierarchies. The maintenance and further development of the KOKO cloud have been transferred to the National Library of Finland, where it is managed by a network of domain ontology developers using the ontology cloud design principles and tools that are being further developed by the library.

The NOR approach of accessing multiple ontology repositories has been used as the internal architecture of the ONKI service, where multiple ontology backend servers are accessed and their information is aggregated by the frontend server. The system enables global search on the distributed ontology network, facilitating the comparison of different ontologies and using multiple ontologies simultaneously, e.g., in content indexing. By publishing ontologies using a shared API and metadata format, the users can access the ontologies using common tools and interfaces, making it easier for them to start using new ontologies and ontology repositories, and integrating them into external applications. The workflow of the users is streamlined as they do not have to access the ontology repositories separately and be familiar with the repository-specific user interfaces, modeling solutions, and other features. For the ontology publishers, NOR aims to increase the visibility of the ontologies as it is easier to incorporate

[5] `http://id.esd.org.uk/IPSV`

them into services that aggregate ontologies from different sources, such as ontology directories. The approach is applicable also to other kinds of data sources in addition to ontology repositories.

4.2.3 Complex Knowledge Organization Systems (RQ3)

The developed TaxMeOn model is a practical solution for managing various kinds of biological name collections. Its design principles include using terminology that is established in biology, focusing only on taxonomic information, and supporting data of various levels of granularity and of alternative views, in order to be simple, yet flexible to use. Accompanying the data model, the research presented in this thesis has contributed by providing a collaborative ontology maintenance and publication workflow utilizing existing tools, such as SAHA metadata editor and ONKI ontology service. By following the approach, it is straightforward and cost-efficient to develop and publish new ontologies for public use. The use of HTTP URIs as identifiers instead of LSIDs that are used in many previous taxonomic databases [148, 51, 55, 225, 215, 64] simplifies the publishing process and use of the taxonomic nomenclature. The system relies on standard resolving and locating mechanisms of the web infrastructure, without the need to implement a specialized LSID resolver.

As a proof of concept of the ontology model, several species checklists of the Finnish Museum of Natural History have been converted into TaxMeOn ontologies and published in the ONKI service. Different stakeholders, such as environmental authorities or biodiversity researchers, can use the system for cataloging, finding, and integrating information from heterogeneous sources, enabling the use of unambiguous taxon references. The ability to link scientific and vernacular names together is useful especially in the citizen science context and information retrieval by laymen as non-professionals might not be familiar with scientific nomenclature.

4.2.4 Summary

The practical implications of the methods and tools presented in this thesis are summarized in Table 4.2, by discussing how the ontology services and models have been applied in the case studies.

Feature	Application in a case study
Creation of interoperable metadata for memory organizations and semantic portals	
Shared ontologies	Semantic interoperability of heterogeneous data sources by creating ontology-based metadata.
Support for legacy data	By basing the ontologies on existing thesauri backwards compatibility to previously created contents is achieved.
Tools for content indexing	Support for common cataloging workflows in a cost-effective way, minimizing manual work.
Building the Finnish national ontology infrastructure	
National level ontology services	Free public services support the creation of standardized metadata in a shared way by governmental agencies, companies, and other organizations.
Formation of the KOKO cloud	A single interlinked representation of a set of domain ontologies makes the cross-domain data integration possible, acting as semantic glue. The redundant work of ontology developers can be eliminated as the overlapping parts of individual ontologies are identified.
Management process of the KOKO cloud	The domain ontologies are kept up-to-date regarding the changes of the general upper ontology, ensuring the validity and consistency of the ontology cloud.
Harmonized data model and API	Multiple ontologies, possibly originating from different repositories can be used simultaneously with shared tools in ontology-based workflows, e.g., in content indexing.
Management and publication of biological name collections	
Support for diverse name resources	The data model can be used for various kinds of name collections, facilitating their interoperability and linking.
Temporal management	The changes and different interpretations of the names and classifications can be tracked and controlled, supporting the scientific processes of taxonomy.
Extendability and external linking	The data model can be expanded to support new use cases and linked to external data sources providing additional information.
Tools for complete workflow	The processes for creating, managing, publishing, and using name collections are supported by general RDF editors, ontology services, and the web infrastructure.

Table 4.2. The features of the ontology services and models presented in this thesis and their application in the case studies.

4.3 Reliability and Validity

The reliability refers to the consistency of the research process and its stability over time and across researchers and methods. The research questions and objectives of this study are stated in Chapter 1, and the research questions are revisited when the results of the study are presented in Chapter 3. The developed models and prototype systems are presented explicitly in Chapter 3, and discussed in more details in the individual publications that are included in the thesis. The objectivity of the research is ensured by describing the research methods, the ontologies used in the study, and the organizations involved. The author of the thesis has no competing interests regarding the research presented and does not recognize personal biases that might affect the process.

The internal validity concerns the achievement of the stated objectives of the research and requirements of the developed systems, the alternative methods, and the limitations of the research. The models and systems developed meet the objectives presented in Chapter 1 as discussed in Chapter 3 and Sections 4.1 and 4.2, as evidenced by applying them in real-world use cases. The ONKI service has been used for publishing several ontologies, integrating them into external applications for creating interoperable metadata of distributed collections of museums and other organizations, and run as a living lab for supporting the users, including content indexers, information searchers, ontology developers, and application developers. The developed ontology models and principles have been used for the creation and management of the ontology cloud KOKO and several biological nomenclatures. The system demonstrates its applicability to the simultaneous usage of multiple ontologies and ontology repositories. The models and software implementations have been compared with relevant related work.

The usability of the user interface of the ONKI system was evaluated and improved by Rami Alatalo [12]. Based on the conducted user survey and interviews focusing on professional content indexers, and heuristic evaluation of the ONKI2 user interface, a new version of the user interface was built. The resulting user interface ONKI3 gained a mean usability score of 48/100 in the System Usability Scale (SUS) [39], based on a user survey. Considering the varying use cases and needs of the target users, the score can be viewed as decent.

In September 2011, a user inquiry was conducted for the users of the

ONKI service using an online questionnaire, consisting of Likert scale and open-ended questions. The inquiry received 107 responses from Finnish libraries, museums, archives, research institutes, and government agencies. The preliminary results of the inquiry were presented in a FinnONTO project board meeting [259]. Below, key findings regarding the functionalities of ONKI are summarized, based on the respondents who answered that they use ONKI daily or weekly.

- 58 % of the users regard the functionalities of ONKI as good or excellent for their needs, whereas 40 % of them think that the functionalities are poor or satisfactory.
- 57 % of the users who use the concept search continuously or occasionally regard the functionalities of ONKI as good or excellent for their needs, whereas 38 % of them think that the functionalities are poor or satisfactory.
- 45 % of the users who the use the ontology browsing continuously or occasionally regard the functionalities of ONKI as good or excellent for their needs, whereas 36 % of them think that the functionalities are poor or satisfactory.
- 55 % of the users think that ONKI is as good or better than the VESA Web Thesaurus Service[6], as opposed to 28 % who think that ONKI is worse than VESA.

When comparing ONKI with VESA, it should be noted that the results give insight on the usefulness of the two systems in relation to each other, not on the absolute quality of the systems. Based on the responses, the most used functionality of ONKI is the concept search. Overall, the ONKI service was regarded as important.

While the ONKI service can be considered an abstract concept, the concrete implementation introduces some limitations. For example, the KOSs that are to be published in the system need to be serialized in the RDF format. RDF is a non-proprietary format, and converting data into it is straightforward in many cases, but some complex data models might require careful and laborious design work. The proposed content

[6] A majority of the Finnish libraries, museums, and other memory organizations have used VESA previously to access established Finnish thesauri maintained by the National Library of Finland. VESA has been since replaced by the Finto service.

indexing methods are mainly manual, meaning they require expensive human efforts, though the methods streamline existing manual indexing processes and thus aim to save costs. In addition to manual indexing, the ontologies published in the ONKI system can be used via APIs, which can be used as components in automatic annotation systems.

The external validity refers to the generalizability of the research results and congruency with prior theory. The applicability of the developed methods to other settings has been ensured by demonstrating representative, diverse use cases in the proof-of-concept systems. The ONKI service has been used as a publishing platform for national and international thesauri and ontologies, covering different domains, ontology modeling languages, owners, and users. In addition to lightweight ontologies, the ONKI APIs and autocompletion widget have been used for accessing more complex geographical ontologies and biological classifications, with use cases in content indexing and query expansion. The tool that was developed for detecting overlappings between ontologies has been used not only in building the KOKO cloud, but also for generating the first version of a combined ontology of legal concepts, based on three vocabularies in the field of the Finnish legislation [87]. The NOR approach has been tested in three different use cases, including a demonstration involving an external ontology repository. The TaxMeOn model for biological nomenclatures has been applied into three distinct types of name collections, each with specific requirements.

The systems presented in the thesis have been designed by taking the previous research into account. The key functionalities of the ONKI service and NOR API are supported by the previous work on ontology servers, as the user requirements and common tasks in ontology-based workflows identified in this thesis are similar to findings by other researchers. The principles of building the ontology cloud complement existing, general ontology design patterns. The TaxMeOn model aims to be a simple, yet semantically rich solution, taking inspiration both from more theoretical modeling approaches and practical publishing efforts of the previous work.

4.4 Recommendations for Further Research

The research presented in this thesis paves way for future work on several areas. Deeper understanding about the roles and requirements of different ontology user groups and the suitability of the developed interfaces and tools for them would require more thorough user evaluation. The work includes testing of the user interfaces of the ONKI service and the suitability of the APIs for various use cases. The use of ontology services in information retrieval needs to be further studied, as more information is required for the optimal selection of the relations used for expanding the queries, and for improving the user interactions in the ONKI widget. Such insight can be gained by conducting a formal evaluation covering alternative options.

Concerning the publication of ontologies, it would be beneficial to make the different versions of an ontology available for the users, not only the latest version as in the ONKI service. By doing so, users would be able to compare the different versions of ontologies, analyze their changes, change frequencies, etc. Access to historical versions of ontologies is needed especially in situations where the user has content that has been indexed with an older version of the ontology and wishes to make it compatible with the current version. Implementing such functionalities would require further research on ontology versioning and visualization methods and their application to concrete use cases.

Regarding the management of a cloud of interlinked ontologies, the processes and tooling for tracking and communicating the changes of the upper ontology to domain ontologies need to be refined. In addition, a more formal process for the development and overall coordination of the ontology cloud would further guide the ontology developers and streamline their work. Research on methods for validating the consistency of the ontology cloud is needed in order to ensure the high quality of the cloud, e.g., to avoid logical errors when reasoning over the class hierarchies. The work of developing the management processes of the ontology cloud is currently underway in the National Library of Finland.

The NOR API that is used for accessing multiple ontology repositories simultaneously could be extended to allow more functionalities. In order to keep the basic API as simple and cost-effective to implement as possible, the extensions could be defined as optional modules that are not required to be implemented in every ontology repository participating in the NOR network. The possibilities for the extensions include more fine-grained ways to restrict the concept search, support for mappings between ontologies, and ranking of the search results, e.g., based on the ontology or repository they are included in.

To facilitate the development and maintenance of biological nomenclatures using the TaxMeOn model, user-friendly tools are needed, as the existing, generic RDF editors do not support efficient management of such complex ontologies. The user interface of the tool should hide the complexities of the data model and present the data in an intuitive and established way to biologists. To this end, research on developing a configurable RDF editor that can be adjusted to different data models would be beneficial. The TaxMeOn model could be extended with new structures to enable a more fine-grained modeling of hybrid taxa and taking into account the distinct features of zoological and botanical nomenclature. The value of the species checklists and name collections published in the ONKI service could be added by developing or using mapping tools to generate links from taxon names to complementing datasets, such as DBpedia.

Bibliography

[1] ISO 5963: Documentation – methods for examining documents, determining their subjects, and selecting indexing terms. ISO standard, International Organization for Standardization, 1985.

[2] SFS 5471: Guidelines for the establishment and maintenance of Finnish language thesauri. SFS standard, Finnish Standards Association, 1988.

[3] FIPA agent communication language specifications. Tech. rep., Foundation for Intelligent Physical Agents, 2001. `http://www.fipa.org/repository/aclspecs.html`.

[4] FIPA ontology service specification. Experimental specification, Foundation for Intelligent Physical Agents, 2001. `http://www.fipa.org/specs/fipa00086/XC00086D.html`.

[5] Global Strategy for Plant Conservation: Technical Rationale, Justification for Updating and Suggested Milestones and Indicators. Meeting report UNEP/CBD/COP/10/19, Convention on Biological Diversity, 2010.

[6] ISO 25964-1: Information and documentation – thesauri and interoperability with other vocabularies – part 1: Thesauri for information retrieval. ISO standard, International Organization for Standardization, 2011.

[7] ABBÈS, S. B., SCHEUERMANN, A., MEILENDER, T., AND D'AQUIN, M. Characterizing modular ontologies. In *Proceedings of the 6th International Workshop on Modular Ontologies, the 7th International Conference on Formal Ontology in I nformation Systems (FOIS 2012), Graz, Austria, July 24* (2012), T. Schneider and D. Walther, Eds., CEUR Workshop Proceedings.

[8] ADAMUSIAK, T., BURDETT, T., KURBATOVA, N., VAN DER VELDE, K. J., ABEYGUNAWARDENA, N., ANTONAKAKI, D., KAPUSHESKY, M., PARKINSON, H., AND SWERTZ, M. A. OntoCAT – simple ontology search and integration in Java, R and REST/JavaScript. *BMC Bioinformatics 12* (2011).

[9] AHMAD, M. N., AND COLOMB, R. M. Managing ontologies: A comparative study of ontology servers. In *Database Technologies 2007: Proceedings of the Eighteenth Australasian Database Conference (ADC2007), Ballarat, Victoria, Australia, January 29 - February 2* (2007), J. Bailey and A. Fekete, Eds., Australian Computer Society, pp. 13–22.

[10] AITCHISON, J., AND DEXTRE CLARKE, S. The Thesaurus: A Historical Viewpoint, with a Look to the Future. *Cataloging and Classification Quarterly 37*, 3/4 (2004), 5–21.

[11] AITCHISON, J., GILCHRIST, A., AND BAWDEN, D. *Thesaurus Construction and Use: A Practical Manual*, 4 ed. Europa Publications, London, 2000.

[12] ALATALO, R. Ontologiapalvelun käyttöliittymän jatkokehitys (Further development of an ontology service user interface), Dec. 2010. Bachelor's thesis. Aalto University, Espoo, Finland.

[13] ALEXANDER, K., CYGANIAK, R., HAUSENBLAS, M., AND ZHAO, J. Describing linked datasets with the VoID vocabulary. W3C interest group note, World Wide Web Consortium, 03 March 2011. `http://www.w3.org/TR/2011/NOTE-void-20110303/`.

[14] AMIN, A., HILDEBRAND, M., VAN OSSENBRUGGEN, J., EVERS, V., AND HARDMAN, L. List, group or menu: organizing suggestions in autocompletion interfaces. Tech. Rep. INS-E0901, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, Netherlands, 2009.

[15] ANDERSON, J. D. Guidelines for indexes and related information retrieval devices. Tech. rep., National Information Standards Organization (NISO), 1997.

[16] ANDREWS, P., ZAIHRAYEU, I., AND PANE, J. A Classification of Semantic Annotation Systems. *Semantic Web 3*, 3 (2012), 223–248.

[17] BACA, M., Ed. *Introduction to Metadata*, 3 ed. Getty Publications, Los Angeles, California, 2016. `http://www.getty.edu/publications/intrometadata`.

[18] BACLAWSKI, K., AND SCHNEIDER, T. The open ontology repository initiative: Requirements and research challenges. In *Proceedings of the Workshop on Collaborative Construction, Management and Linking of Structured Knowledge (CK2009), collocated with the 8th International Semantic Web Conference (ISWC-2009), Washington D.C., USA, October 25* (2009), T. Tudorache, G. Correndo, N. Noy, H. Alani, and M. Greaves, Eds., CEUR Workshop Proceedings.

[19] BAKER, T., BECHHOFER, S., ISAAC, A., MILES, A., SCHREIBER, G., AND SUMMERS, E. Key choices in the design of Simple Knowledge Organization System (SKOS). *Journal of Web Semantics 20* (2013), 35–49.

[20] BANDHOLTZ, T., SCHULTE-COERNE, T., GLASER, R., FOCK, J., AND KELLER, T. iQvoc – Open Source SKOS(XL) Maintenance and Publishing Tool. In *Proceedings of the Sixth Workshop on Scripting and Development for the Semantic Web, co-located with the European Semantic Web Conference 2010 (ESWC 2010), Crete, Greece, May 31* (2010), G. A. Grimnes, S. Auer, and G. T. Williams, Eds., CEUR Workshop Proceedings.

[21] BASCA, C., CORLOSQUET, S., CYGANIAK, R., FERNÁNDEZ, S., AND SCHANDL, T. Neologism: Easy vocabulary publishing. In *Proceedings of the 4th Workshop on Scripting for the Semantic Web, Tenerife, Spain, June 02* (2008), C. Bizer, S. Auer, G. A. Grimnes, and T. Heath, Eds., CEUR Workshop Proceedings.

[22] BASKAUF, S. J., AND WEBB, C. O. Darwin-SW: Darwin Core-based terms for expressing biodiversity data as RDF. *Semantic Web 7*, 6 (2016), 645–667.

[23] BASKERVILLE, R. L. Distinguishing action research from participative case studies. *Journal of Systems and Information Technology 1*, 1 (1997), 25–45.

[24] BASKERVILLE, R. L., AND WOOD-HARPER, A. T. A critical perspective on action research as a method for information systems research. *Journal of Information Technology 11*, 3 (1996), 235–246.

[25] BERENDSOHN, W. G. The concept of "potential taxa" in databases. *Taxon 44* (1995), 207–212.

[26] BERENDSOHN, W. G. A taxonomic information model for botanical databases: The IOPI model. *Taxon 46*, 2 (1997), 283–309.

[27] BERNERS-LEE, T. Linked data, W3C design issues. `http://www.w3.org/DesignIssues/LinkedData.html`, 2006.

[28] BERNERS-LEE, T., HENDLER, J., AND LASSILA, O. The semantic web. *Scientific American 284*, 5 (2001), 34–43.

[29] BHOGAL, J., MACFARLANE, A., AND SMITH, P. A review of ontology based query expansion. *Information Processing and Management 43*, 4 (2007), 866–886.

[30] BINDING, C., AND TUDHOPE, D. KOS at your Service: Programmatic Access to Knowledge Organisation Systems. *Journal of Digital Information 4*, 4 (2004).

[31] BINDING, C., AND TUDHOPE, D. Terminology Web Services. *Knowledge Organization 37*, 4 (2010), 287–298.

[32] BINDING, C., AND TUDHOPE, D. Improving interoperability using vocabulary linked data. *International Journal on Digital Libraries 17*, 1 (2016), 5–21.

[33] BLOMQVIST, E., GANGEMI, A., AND PRESUTTI, V. Experiments on pattern-based ontology design. In *Proceedings of the fifth international conference on Knowledge capture (K-CAP '09), Redondo Beach, California, USA, September 1-4* (2009), ACM, pp. 41–48.

[34] BODNER, R. C., AND SONG, F. Knowledge-Based Approaches to Query Expansion in Information Retrieval. In *Advances in Artifical Intelligence: 11th Biennial Conference of the Canadian Society for Computational Studies of Intelligence, AI '96 Toronto, Ontario, Canada, May 21–24, 1996 Proceedings* (1996), G. McCalla, Ed., Springer-Verlag, pp. 146–158.

[35] BORST, W. N. *Construction of engineering ontologies for knowledge sharing and reuse*. PhD thesis, University of Twente, Netherlands, Sept. 1997.

[36] BOYLE, B., HOPKINS, N., LU, Z., RAYGOZA GARAY, J. A., MOZZHERIN, D., REES, T., MATASCI, N., NARRO, M. L., PIEL, W. H., MCKAY, S. J., LOWRY, S., FREELAND, C., PEET, R. K., AND ENQUIST, B. J. The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinformatics 14* (2013).

[37] BRAUN, S., SCHMIDT, A., WALTER, A., AND ZACHARIAS, V. The Ontology Maturing Approach for Collaborative and Work Integrated Ontology Development: Evaluation Results and Future Directions. In *Proceedings of the First International Workshop on Emergent Semantics and Ontology Evolution (ESOE 2007), co-located with ISWC 2007 + ASWC 2007, Busan, Korea, November 12* (2007), L. L. Chen, P. Cudré-Mauroux, P. Haase, A. Hotho, and E. Ong, Eds., CEUR Workshop Proceedings, pp. 5–18.

[38] BRICKLEY, D., AND MILLER, L. FOAF vocabulary specification 0.99. Namespace document, 14 January 2014. `http://xmlns.com/foaf/spec/20140114.html`.

[39] BROOKE, J. SUS - A quick and dirty usability scale. In *Usability evaluation in industry*, P. W. Jordan, B. Thomas, I. L. McClelland, and B. Weerdmeester, Eds. Taylor & Francis, 1996.

[40] BRUGMAN, H., MALAISÉ, V., AND GAZENDAM, L. A Web Based General Thesaurus Browser to Support Indexing of Television and Radio Programs. In *Proceedings of the fifth international conference on Language Resources and Evaluation (LREC 2006), Genoa, Italy, May 22-28* (2006).

[41] BUITELAAR, P., EIGNER, T., AND DECLERCK, T. OntoSelect: A Dynamic Ontology Library with Support for Ontology Selection. In *Proceedings of the 3rd International Semantic Web Conference (ISWC2004), Demo Papers, Hiroshima, Japan, November 7-11* (2004).

[42] BURSTEIN, F., AND GREGOR, S. The Systems Development or Engineering Approach to Research in Information Systems: An Action Research Perspective. In *Proceedings of 10th Australasian Conference on Information Systems, Wellington, New Zealand, December 1-3* (1999), B. Hope and P. Yoong, Eds., pp. 122–134.

[43] CARACCIOLO, C., STELLATO, A., RAJBAHNDARI, S., MORSHED, A., JOHANSSEN, G., JAQUES, Y., AND KEIZER, J. Thesaurus maintenance, alignment and publication as linked data: the AGROVOC use case. *International Journal of Metadata, Semantics and Ontologies 7*, 1 (2012), 65–75.

[44] CARDILLO, E., FOLINO, A., TRUNFIO, R., AND GUARASCI, R. Towards the reuse of standardized thesauri into ontologies. In *Proceedings of the 5th International Conference on Ontology and Semantic Web Patterns (WOP'14), Riva del Garda, Italy, October 19* (2014), V. de Boer, A. Gangemi, K. Janowicz, and A. Lawrynowicz, Eds., CEUR Workshop Proceedings, pp. 26–37.

[45] CARPINETO, C., AND ROMANO, G. A Survey of Automatic Query Expansion in Information Retrieval. *ACM Computing Surveys 44*, 1 (2012).

[46] CATHRO, W. Metadata: An Overview. In *Proceedings of the Standards Australia Seminar: Matching Discovery and Recovery, Sydney and Melbourne, Australia, August 14-15* (1997).

[47] CHAUDHRI, V. K., FARQUHAR, A., FIKES, R., KARP, P. D., AND RICE, J. P. Open knowledge base connectivity 2.0.3. Proposed specification, OKBC working group, April 9, 1998. `http://www.ai.sri.com/~okbc/spec/okbc2/okbc2.html`.

[48] CHAWUTHAI, R., TAKEDA, H., WUWONGSE, V., AND JINBO, U. Presenting and Preserving the Change in Taxonomic Knowledge for Linked Data. *Semantic Web 7*, 6 (2016), 589–616.

[49] CHUNG, H., CHOI, J.-M., YI, J.-H., HAN, J., AND LEE, E.-S. A Construction of the Adapted Ontology Server in EC. In *Intelligent Data Engineering and Automated Learning — IDEAL 2000. Data Mining, Financial Engineering, and Intelligent Agents: Second International Conference Shatin, N.T., Hong Kong, China, December 13–15, 2000 Proceedings* (2002), K. S. Leung, L.-W. Chan, and H. Meng, Eds., Springer-Verlag, pp. 355–360.

[50] CONWAY, M., KHOJOYAN, A., FANA, F., SCUBA, W., CASTINE, M., MOWERY, D., CHAPMAN, W., AND JUPP, S. Developing a web-based SKOS editor. *Journal of Biomedical Semantics 7*, 5 (2016).

[51] COSTELLO, M. J., BOUCHET, P., BOXSHALL, G., FAUCHALD, K., GORDON, D., HOEKSEMA, B. W., POORE, G. C. B., VAN SOEST, R. W. M., STÖHR, S., WALTER, T. C., VANHOORNE, B., DECOCK, W., AND APPELTANS, W. Global Coordination and Standardisation in Marine Biodiversity through the World Register of Marine Species (WoRMS) and Related Databases. *PLoS ONE 8*, 1 (2013).

[52] CÔTÉ, R. G., JONES, P., APWEILER, R., AND HERMJAKOB, H. The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. *BMC Bioinformatics 7* (2006).

[53] COWIE, J., AND LEHNERT, W. Information extraction. *Communications of the ACM 39*, 1 (1996), 80–91.

[54] COX, S. J. D., YU, J., AND RANKINE, T. SISSVoc: A Linked Data API for access to SKOS vocabularies. *Semantic Web 7*, 1 (2016), 9–24.

[55] CROFT, J., CROSS, N., HINCHCLIFFE, S., LUGHADHA, E. N., STEVENS, P. F., WEST, J. G., AND WHITBREAD, G. Plant Names for the 21st Century: The International Plant Names Index, a Distributed Data Source of General Accessibility. *Taxon 48*, 2 (1999), 317–324.

[56] CRYER, P., HYAM, R., MILLER, C., NICOLSON, N., TUAMA, E. O., PAGE, R., REES, J., RICCARDI, G., RICHARDS, K., AND WHITE, R. Adoption of persistent identifiers for biodiversity informatics: Recommendations of the GBIF LSID GUID task group, 6. November 2009. Tech. rep., Global Biodiversity Information Facility (GBIF), Copenhagen, Denmark, 2010. version 1.1, last updated 21 Jan 2010.

[57] DAMERON, O., NOY, N. F., KNUBLAUCH, H., AND MUSEN, M. A. Accessing and manipulating ontologies using web services. In *Proceedings of the ISWC 2004 Workshop on Semantic Web Services: Preparing to Meet the World of Business Applications, Hiroshima, Japan, November 8* (2004), D. Martin, R. Lara, and T. Yamaguchi, Eds., CEUR Workshop Proceedings.

[58] D'AQUIN, M., AND LEWEN, H. Cupboard – a place to expose your ontologies to applications and the community. In *The Semantic Web: Research and Applications: 6th European Semantic Web Conference, ESWC 2009 Heraklion, Crete, Greece, May 31–June 4, 2009 Proceedings* (2009), L. Aroyo, P. Traverso, F. Ciravegna, P. Cimiano, T. Heath, E. Hyvönen, R. Mizoguchi, E. Oren, M. Sabou, and E. Simperl, Eds., Springer-Verlag, pp. 913–918.

[59] D'AQUIN, M., MOTTA, E., SABOU, M., ANGELETOU, S., GRIDINOC, L., LOPEZ, V., AND GUIDI, D. Toward a new generation of semantic web applications. *IEEE Intelligent Systems 23*, 3 (2008), 20–28.

[60] D'AQUIN, M., AND NOY, N. F. Where to publish and find ontologies? a survey of ontology libraries. *Journal of Web Semantics 11* (2012), 96–111.

[61] DARWIN CORE TASK GROUP. Darwin Core. TDWG current standard, Biodiversity Information Standards (TDWG), 2009. http://www.tdwg.org/standards/450.

[62] DAVISON, R. M., MARTINSONS, M. G., AND KOCK, N. Principles of Canonical Action Research. *Information Systems Journal 14* (2004), 65–86.

[63] DCMI USAGE BOARD. DCMI metadata terms. DCMI recommendation, Dublin Core Metadata Initiative, 14.6.2012. http://dublincore.org/documents/2012/06/14/dcmi-terms/.

[64] DE JONG, Y., VERBEEK, M., MICHELSEN, V., BJØRN, P. D. P., LOS, W., STEEMAN, F., BAILLY, N., BASIRE, C., CHYLARECKI, P., STLOUKAL, E., HAGEDORN, G., WETZEL, F. T., GLÖCKLER, F., KROUPA, A., KORB, G., HOFFMANN, A., HÄUSER, C., KOHLBECKER, A., MÜLLER, A., GÜNTSCH, A., STOEV, P., AND PENEV, L. Fauna Europaea - all European animal species on the web. *Biodiversity Data Journal 2* (2014).

[65] DENGLER, J., BERENDSOHN, W. G., BERGMEIER, E., CHYTRÝ, M., DANIHELKA, J., JANSEN, F., KUSBER, W.-H., LANDUCCI, F., MÜLLER, A., PANFILI, E., SCHAMINÉE, J., VENANZONI, R., AND VON RAAB-STRAUBE, E. The need for and the requirements of EuroSL, an electronic taxonomic reference list of all European plants. *Biodiversity & Ecology 4* (2012), 15–24.

[66] DEXTRE CLARKE, S. G., AND ZENG, M. L. From ISO 2788 to ISO 25964: The evolution of thesaurus standards towards interoperability and data modeling. *Information Standards Quarterly 24*, 1 (2012), 20–26.

[67] DING, L., FININ, T., JOSHI, A., PAN, R., COST, R. S., PENG, Y., REDDIVARI, P., DOSHI, V., AND SACHS, J. Swoogle: a search and metadata engine for the semantic web. In *Proceedings of the thirteenth ACM international conference on Information and knowledge management (CIKM '04), Washington, D.C., USA, November 8-13* (2004), ACM, pp. 652–659.

[68] DING, Y., AND FENSEL, D. Ontology Library Systems: The key to successful Ontology Re-use. In *Proceedings of the 1st Semantic Web Working Symposium (SWWS'01), Stanford University, California, USA, July 30 - August 1* (2001), I. F. Cruz, S. Decker, J. Euzenat, and D. L. McGuinness, Eds., pp. 93–112.

[69] DOMINGUE, J. Tadzebao and WebOnto: Discussing, browsing, and editing ontologies on the web. In *Proceedings of the 11th Workshop on Knowledge Acquisition, Modeling and Management, Banff, Alberta, Canada, April 18-23* (1998).

[70] DRAHEIM, D. The Service-Oriented Metaphor Deciphered. *Journal of Computing Science and Engineering 4*, 4 (2010), 253–275.

[71] DUVAL, E. Metadata Standards: What, Who & Why. *Journal of Universal Computer Science 7*, 7 (2001), 591–601.

[72] EFTHIMIADIS, E. N. Interactive query expansion: A user-based evaluation in a relevance feedback environment. *Journal of the American Society for Information Science and Technology 51*, 11 (2000), 989–1003.

[73] EL JERROUDI, Z., AND ZIEGLER, J. iMERGE: interactive ontology merging. In *EKAW 2008: 16th International Conference on Knowledge Engineering and Knowledge Management: Knowledge Patterns, Acitrezza, Italy, September 29 - October 2, Poster and Demo Proceedings* (2008), pp. 52–56.

[74] EUZENAT, J., AND SHVAIKO, P. *Ontology Matching*. Springer-Verlag, Berlin Heidelberg, 2007.

[75] FARQUHAR, A., FIKES, R., AND RICE, J. The Ontolingua Server: a tool for collaborative ontology construction. *International Journal of Human-Computer Studies 46*, 6 (1997), 707–727.

[76] FEDERHEN, S. The NCBI Taxonomy database. *Nucleic Acids Research 40*, D1 (2012), 136–143.

[77] FEIGENBAUM, L., HERMAN, I., HONGSERMEIER, T., NEUMANN, E., AND STEPHENS, S. The semantic web in action. *Scientific American 297*, 6 (2007), 90–97.

[78] FIELDING, R. T. *Architectural Styles and the Design of Network-based Software Architectures*. PhD thesis, University of California, Irvine, USA, 2000.

[79] FIKES, R., AND FARQUHAR, A. Distributed Repositories of Highly Expressive Reusable Ontologies. *IEEE Intelligent Systems 14*, 2 (1999), 73–79.

[80] FININ, T., WEBER, J., WIEDERHOLD, G., GENESERETH, M., FRITZSON, R., MCKAY, D., MCGUIRE, J., PELAVIN, R., SHAPIRO, S., AND BECK, C. Specification of the KQML agent-communication language – plus example agent policies and architectures. Draft specification, DARPA Knowledge Sharing Initiative External Interfaces Working Group, June 15, 1993. `http://www.csee.umbc.edu/csee/research/kqml/papers/kqmlspec.pdf`.

[81] FLOURIS, G., MANAKANATAS, D., KONDYLAKIS, H., PLEXOUSAKIS, D., AND ANTONIOU, G. Ontology change: classification and survey. *The Knowledge Engineering Review 23*, 2 (2008), 117–152.

[82] FLUIT, C., SABOU, M., AND VAN HARMELEN, F. Supporting User Tasks through Visualisation of Light-weight Ontologies. *Handbook on Ontologies in Information Systems* (2003), 1–20.

[83] FOX, E. A. Lexical Relations: Enhancing Effectiveness of Information Retrieval Systems. *ACM SIGIR Forum 15*, 3 (1980), 5–36.

[84] FRANZ, N. M., CHEN, M., KIANMAJD, P., YU, S., BOWERS, S., WEAKLEY, A. S., AND LUDÄSCHER, B. Names Are Not Good Enough: Reasoning over Taxonomic Change in the Andropogon Complex. *Semantic Web 7*, 6 (2016), 645–667.

[85] FRANZ, N. M., AND PEET, R. K. Towards a language for mapping relationships among taxonomic concepts. *Systematics and Biodiversity 7*, 1 (2009), 5–20.

[86] FRANZ, N. M., AND THAU, D. Biological taxonomy and ontology development: scope and limitations. *Biodiversity Informatics 7* (2010), 45–66.

[87] FROSTERUS, M., TUOMINEN, J., AND HYVÖNEN, E. Facilitating Re-use of Legal Data in Applications—Finnish Law as a Linked Open Data Service. In *Legal Knowledge and Information Systems: JURIX 2014: The Twenty-Seventh Annual Conference* (2014), R. Hoekstra, Ed., IOS Press, pp. 115–124.

[88] FU, G., JONES, C. B., AND ABDELMOTY, A. I. Ontology Based Spatial Query Expansion in Information Retrieval. In *On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2005, Agia Napa, Cyprus, October 31 - November 4, 2005, Proceedings Part II* (2005), R. Meersman and Z. Tari, Eds., Springer-Verlag, pp. 1466–1482.

[89] GANGEMI, A., AND PRESUTTI, V. Ontology Design Patterns. In *Handbook on Ontologies*, S. Staab and R. Studer, Eds., 2 ed. Springer-Verlag, 2009.

[90] GARCÍA-CASTRO, A., GARCÍA-CASTRO, L., VILLAVECES, J. M., CALDERÓN, G., AND HEPP, M. OLS2OWL. A repository management facility. In *Proceedings of the 11th International Protégé Conference, Amsterdam, Netherlands, June 23-26* (2009).

[91] GARSHOL, L. M. Metadata? Thesauri? Taxonomies? Topic Maps! Making Sense of it all. *Journal of Information Science 30*, 4 (2004), 378–391.

[92] GENNARI, J. H., OLIVER, D. E., PRATT, W., RICE, J., AND MUSEN, M. A. A Web-Based Architecture for a Medical Vocabulary Server. In *Proceedings of the 19th Annual Symposium on Computer Applications in Medical Care, New Orleans, Lousiana, USA, October 28 - November 1* (1995), R. M. Gardner, Ed., American Medical Informatics Association, pp. 275–279.

[93] GIUNCHIGLIA, F., AND ZAIHRAYEU, I. Lightweight ontologies. In *Encyclopedia of Database Systems*, L. Liu and M. T. Özsu, Eds. Springer-Verlag, 2009, pp. 1613–1619.

[94] GLOBAL BIODIVERSITY INFORMATION FACILITY (GBIF). ChecklistBank. `http://github.com/gbif/checklistbank`. Referenced 12 March 2017.

[95] GOLUB, K., TUDHOPE, D., ZENG, M. L., AND ŽUMER, M. Terminology Registries for Knowledge Organization Systems: Functionality, Use, and Attributes. *Journal of the Association for Information Science and Technology 65*, 9 (2014), 1901–1916.

[96] GREENBERG, J., LOSEE, R., PÉREZ AGÜERA, J. R., SCHERLE, R., WHITE, H., AND WILLIS, C. HIVE: Helping Interdisciplinary Vocabulary Engineering. *Bulletin of the American Society for Information Science & Technology 37*, 4 (2011), 23–26.

[97] GROSS, A., PRUSKI, C., AND RAHM, E. Evolution of biomedical ontologies and mappings: Overview of recent approaches. *Computational and Structural Biotechnology Journal 14* (2016), 333–340.

[98] GRUBER, T. R. A translation approach to portable ontology specification. *Knowledge Acquisition 5*, 2 (1993), 199–220.

[99] GRUBER, T. R. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies 43*, 5-6 (1995), 907–928.

[100] GUARINO, N., OBERLE, D., AND STAAB, S. What Is an Ontology? In *Handbook on ontologies*, S. Staab and R. Studer, Eds., 2 ed. Springer-Verlag, 2009, pp. 1–17.

[101] GUARINO, N., AND WELTY, C. Evaluating Ontological Decisions with Ontoclean. *Communications of the ACM 45*, 2 (2002), 61–65.

[102] GULLI, A., AND SIGNORINI, A. The Indexable Web is More than 11.5 Billion Pages. In *Proceedings of the WWW '05 Special interest tracks and posters of the 14th international conference on World Wide Web, Chiba, Japan, May 10-14* (2005), ACM, pp. 902–903.

[103] HALPIN, H., HAYES, P. J., MCCUSKER, J. P., MCGUINNESS, D. L., AND THOMPSON, H. S. When owl:sameAs isn't the Same: An Analysis of Identity in Linked Data. In *The Semantic Web – ISWC 2010: 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7-11, 2010, Revised Selected Papers, Part I* (2010), P. S. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika, L. Zhang, J. Z. Pan, I. Horrocks, and B. Glimm, Eds., Springer-Verlag, pp. 305–320.

[104] HAMEED, A., PREECE, A., AND SLEEMAN, D. Ontology reconciliation. In *Handbook on ontologies*, S. Staab and R. Studer, Eds. Springer-Verlag, 2004, pp. 231–250.

[105] HARDISTY, A., ROBERTS, D., AND BIODIVERSITY INFORMATICS COMMUNITY. A decadal view of biodiversity informatics: challenges and priorities. *BMC Ecology 13*, 16 (2013).

[106] HARPER, C. A., AND TILLETT, B. B. Library of Congress Controlled Vocabularies and Their Application to the Semantic Web. *Cataloging & Classification Quarterly 43*, 3-4 (2007), 47–68.

[107] HARRIS, S., AND SEABORNE, A. SPARQL 1.1 query language. W3C recommendation, World Wide Web Consortium, 21 March 2013. `http://www.w3.org/TR/2013/REC-sparql11-query-20130321/`.

[108] HARTMANN, J., PALMA, R., AND GÓMEZ-PÉREZ, A. Ontology Repositories. In *Handbook on Ontologies*, S. Staab and R. Studer, Eds., 2 ed. Springer-Verlag, 2009, pp. 551–571.

[109] HARTMANN, J., PALMA, R., SURE, Y., SUÁREZ-FIGUEROA, M. C., HAASE, P., GÓMEZ-PÉREZ, A., AND STUDER, R. Ontology metadata vocabulary and applications. In *On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops: OTM Confederated Internationl Workshops and Posters, AWeSOMe, CAMS, GADA, MIOS+INTEROP, ORM, PhDS, SeBGIS, SWWS, and WOSE 2005, Agia Napa, Cyprus, October 31 - November 4, 2005. Proceedi* (2005), R. Meersman, Z. Tari, and P. Herrero, Eds., Springer-Verlag, pp. 906–915.

[110] HARTUNG, M., GROSS, A., AND RAHM, E. COnto-Diff: Generation of complex evolution mappings for life science ontologies. *Journal of Biomedical Informatics 46*, 1 (2013), 15–32.

[111] HARTUNG, M., TERWILLIGER, J., AND RAHM, E. Recent Advances in Schema and Ontology Evolution. *Evolution 141*, 3 (2011), 149–190.

[112] HEATH, T., AND BIZER, C. *Linked Data: Evolving the Web into a Global Data Space*. Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool, Palo Alto, California, 2011.

[113] HEFLIN, J., AND HENDLER, J. Dynamic Ontologies on the Web. In *Proceedings of the 17th National Conference on Artificial Intelligence, Austin, Texas, USA, July 30 - August 3* (2000), H. A. Kautz and B. W. Porter, Eds., AAAI Press / MIT Press, pp. 443–449.

[114] HEFLIN, J. D. *Towards the Semantic Web: Knowledge Representation in a Dynamic, Distributed Environment*. PhD thesis, University of Maryland, USA, 2001.

[115] HERBERT, K. G., GEHANI, N. H., PIEL, W. H., WANG, J. T. L., AND WU, C. H. BIO-AJAX: an extensible framework for biological data cleaning. *ACM SIGMOD Record 33*, 2 (2004), 51–57.

[116] HEVNER, A. R., MARCH, S. T., PARK, J., AND RAM, S. Design Science in Information Systems Research. *MIS Quarterly 28*, 1 (2004), 75–105.

[117] HILDEBRAND, M., VAN OSSENBRUGGEN, J., AMIN, A., AROYO, L., WIELEMAKER, J., AND HARDMAN, L. The design space of a configurable autocompletion component. Tech. Rep. INS-E0708, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, Netherlands, 2007.

[118] HILDEBRAND, M., VAN OSSENBRUGGEN, J., HARDMAN, L., AND JACOBS, G. Supporting subject matter annotation using heterogeneous thesauri, a user study in Web data reuse. Tech. Rep. INS-E0902, Centrum Wiskunde & Informatica, 2009.

[119] HILL, L., BUCHEL, O., JANÉE, G., AND ZENG, M. L. Integration of Knowledge Organization Systems into Digital Library Architectures. In *13th ASIST SIGICR Workshop, Reconceptualizing Classification Research* (2002), pp. 46–52.

[120] HINCHLIFF, C. E., SMITH, S. A., ALLMAN, J. F., BURLEIGH, J. G., CHAUDHARY, R., COGHILL, L. M., CRANDALL, K. A., DENG, J., DREW, B. T., GAZIS, R., GUDE, K., HIBBETT, D. S., KATZ, L. A., LAUGHINGHOUSE, H. D., MCTAVISH, E. J., MIDFORD, P. E., OWEN, C. L., REE, R. H., REES, J. A., SOLTIS, D. E., WILLIAMS, T., AND CRANSTON, K. A. Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proceedings of the National Academy of Sciences 112*, 41 (2015), 12764–12769.

[121] HITZLER, P., AND VAN HARMELEN, F. A Reasonable Semantic Web. *Semantic Web 1*, 1-2 (2010), 39–44.

[122] HLAVA, M. Developing an eclectic terminology registry. *Bulletin of the Association for Information Science and Technology 37*, 4 (2011), 19–22.

[123] HOANG, H. H., AND TJOA, A. M. The State of the Art of Ontology-based Query Systems: A Comparison of Existing Approaches. In *Proceedings of the International Conference on Computing and Informatics, Kuala Lumpur, Malaysia, June 6-8* (2006), IEEE.

[124] HODGE, G. Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files. Tech. rep., The Digital Library Federation, Council on Library and Information Resources, 2000.

[125] HOLETSCHEK, J., DRÖGE, G., GÜNTSCH, A., AND BERENDSOHN, W. G. The ABCD of primary biodiversity data access. *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology 146*, 4 (2012), 771–779.

[126] HOLLINK, L., SCHREIBER, G., AND WIELINGA, B. Patterns of semantic relations to improve image content search. *Journal of Web Semantics 5*, 3 (2007), 195–203.

[127] HOOI, Y. K., HASSAN, M. F., AND SHARIFF, A. M. A Survey on Ontology Mapping Techniques. In *Advances in Computer Science and its Applications: CSA 2013* (2014), H. Y. Jeong, M. S. Obaidat, N. Y. Yen, and J. J. J. H. Park, Eds., Springer-Verlag, pp. 829–836.

[128] HUANG, Z., AND STUCKENSCHMIDT, H. Reasoning with multi-version ontologies: A temporal logic approach. In *The Semantic Web – ISWC 2005: 4th International Semantic Web Conference, ISWC 2005, Galway, Ireland, November 6-10, 2005, Proceedings* (2005), Y. Gil, E. Motta, V. R. Benjamins, and M. A. Musen, Eds., Springer-Verlag, pp. 398–412.

[129] HYVÖNEN, E., ALONEN, M., KOHO, M., AND TUOMINEN, J. BirdWatch — Supporting Citizen Scientists for Better Linked Data Quality for Biodiversity Management. In *Proceedings of the first international Workshop on Semantics for Biodiversity Montpellier, France, May 27* (2013), P. Larmande, E. Arnaud, I. Mougenot, C. Jonquet, T. Libourel, and M. Ruiz, Eds., CEUR Workshop Proceedings.

[130] HYVÖNEN, E., IKKALA, E., AND TUOMINEN, J. Linked data brokering service for historical places and maps. In *Proceedings of the 1st Workshop on Humanities in the Semantic Web co-located with 13th ESWC Conference 2016 (ESWC 2016) Anissaras, Greece, May 29* (2016), A. Adamou, E. Daga, and L. Isaksen, Eds., CEUR Workshop Proceedings, pp. 39–52.

[131] HYVÖNEN, E., LINDROOS, R., KAUPPINEN, T., AND HENRIKSSON, R. An Ontology Service for Geographical Content. In *Poster Proceedings of the International Semantic Web Conference (ISWC 2007), Busan, Korea, November 11-15* (2007).

[132] HYVÖNEN, E., AND MÄKELÄ, E. Semantic Autocompletion. In *The Semantic Web – ASWC 2006: First Asian Semantic Web Conference, Beijing, China, September 3-7, 2006, Proceedings* (2006), R. Mizoguchi, Z. Shi, and F. Giunchiglia, Eds., Springer-Verlag, pp. 739–751.

[133] HYVÖNEN, E., MÄKELÄ, E., KAUPPINEN, T., ALM, O., KURKI, J., RUOTSALO, T., SEPPÄLÄ, K., TAKALA, J., PUPUTTI, K., KUITTINEN, H., VILJANEN, K., TUOMINEN, J., PALONEN, T., FROSTERUS, M., SINKKILÄ, R., PAAKKARINEN, P., LAITIO, J., AND NYBERG, K. CultureSampo – A National Publication System of Cultural Heritage on the Semantic Web 2.0. In *The Semantic Web: Research and Applications: 6th European Semantic Web Conference, ESWC 2009 Heraklion, Crete, Greece, May 31–June 4, 2009 Proceedings* (2009), L. Aroyo, P. Traverso, F. Ciravegna, P. Cimiano,

T. Heath, E. Hyvönen, R. Mizoguchi, E. Oren, M. Sabou, and E. Simperl, Eds., Springer-Verlag, pp. 851–856.

[134] HYVÖNEN, E., TUOMINEN, J., ALONEN, M., AND MÄKELÄ, E. Linked Data Finland: A 7-star model and platform for publishing and re-using linked datasets. In *The Semantic Web: ESWC 2014 Satellite Events: ESWC 2014 Satellite Events, Anissaras, Crete, Greece, May 25-29, 2014, Revised Selected Papers* (2014), V. Presutti, E. Blomqvist, R. Troncy, H. Sack, I. Papadakis, and A. Tordai, Eds., Springer-Verlag, pp. 226–230.

[135] HYVÖNEN, E., TUOMINEN, J., KAUPPINEN, T., AND VÄÄTÄINEN, J. Representing and utilizing changing historical places as an ontology time series. In *Geospatial Semantics and the Semantic Web: Foundations, Algorithms, and Applications*, N. Ashish and A. P. Sheth, Eds. Springer-Verlag, 2011, pp. 1–25.

[136] HYVÖNEN, E., VILJANEN, K., TUOMINEN, J., AND SEPPÄLÄ, K. Building a National Semantic Web Ontology and Ontology Service Infrastructure—The FinnONTO Approach. In *The Semantic Web: Research and Applications: 5th European Semantic Web Conference, ESWC 2008, Tenerife, Canary Islands, Spain, June 1-5, 2008 Proceedings* (2008), S. Bechhofer, M. Hauswirth, J. Hoffmann, and M. Koubarakis, Eds., Springer-Verlag, pp. 95–109.

[137] HÖLSCHER, C., AND STRUBE, G. Web search behaviour of internet experts and newbies. *Computer networks 33*, 1–6 (2000), 337–346.

[138] ICZN - INTERNATIONAL COMMISSION ON ZOOLOGICAL NOMENCLATURE. *International Code of Zoological Nomenclature*. International Trust for Zoological Nomenclature, 1999.

[139] ISO TC46/SC9/WG8, AND ISAAC, A. Correspondence between ISO 25964 and SKOS/SKOS-XL Models. Tech. rep., National Information Standards Organization (NISO), 2013.

[140] JAIN, P., HITZLER, P., KUNAL VERMA, YEH, P. Z., AND SHETH, A. Moving beyond sameAs with PLATO : Partonomy detection for Linked Data. In *Proceedings of 23rd ACM conference on Hypertext and social media, Milwaukee, USA, June 25-28* (2012), ACM, pp. 33–42.

[141] JARRAR, M., AND MEERSMAN, R. Formal Ontology Engineering in the DOGMA Approach. In *On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE: Confederated International Conferences CoopIS, DOA, and ODBASE 2002 Proceedings* (2002), R. Meersman and Z. Tari, Eds., Springer-Verlag, pp. 1238–1254.

[142] JÄRVELIN, K., KEKÄLÄINEN, J., AND NIEMI, T. ExpansionTool: Concept-Based Query Expansion and Construction. *Information Retrieval 4*, 3-4 (2001), 231–255.

[143] JASPER, R., AND USCHOLD, M. A Framework for Understanding and Classifying Ontology Applications. In *Proceedings of the Twelfth Workshop on Knowledge Acquisition, Modeling and Management, Banff, Alberta, Canada, October 16-22* (1999).

[144] JAVED, M., ABGAZ, Y. M., AND PAHL, C. Ontology Change Management and Identification of Change Patterns. *Journal on Data Semantics 2*, 2 (2013), 119–143.

[145] JETZ, W., MCPHERSON, J. M., AND GURALNICK, R. P. Integrating biodiversity distribution knowledge: Toward a global map of life. *Trends in Ecology and Evolution 27*, 3 (2012), 151–159.

[146] JIMÉNEZ RUIZ, E., CUENCA GRAU, B., HORROCKS, I., AND BERLANGA, R. Supporting concurrent ontology development: Framework, algorithms and tool. *Data & Knowledge Engineering 70*, 1 (2011), 146–164.

[147] JOHO, H., COVERSON, C., SANDERSON, M., AND BEAULIEU, M. Hierarchical presentation of expansion terms. In *Proceedings of the 2002 ACM symposium on Applied computing, Madrid, Spain, March 11-14* (2002), ACM, pp. 645–649.

[148] JONES, A. C., WHITE, R. J., AND ORME, E. R. Identifying and relating biological concepts in the Catalogue of Life. *Journal of Biomedical Semantics 2*, 7 (2011).

[149] JOTHILAKSHMI, R., SHANTHI, N., AND BABISARASWATHI, R. A survey on semantic query expansion. *Journal of Theoretical and Applied Information Technology 57*, 1 (2013), 128–138.

[150] JUDD, W. S., CAMPBELL, C. S., KELLOGG, E. A., STEVENS, P. F., AND DONOGHUE, M. J. *Plant Systematics: A Phylogenetic Approach*, 4 ed. Sinauer Associates, Sunderland, Massachusetts, 2016.

[151] JUPP, S., BECHHOFER, S., AND STEVENS, R. A flexible API and editor for SKOS. In *The Semantic Web: Research and Applications: 6th European Semantic Web Conference, ESWC 2009 Heraklion, Crete, Greece, May 31–June 4, 2009 Proceedings* (2009), L. Aroyo, P. Traverso, F. Ciravegna, P. Cimiano, T. Heath, E. Hyvönen, R. Mizoguchi, E. Oren, M. Sabou, and E. Simperl, Eds., Springer-Verlag, pp. 506–520.

[152] KATASONOV, A., AND TERZIYAN, V. Semantic Agent Programming Language (S-APL): A Middleware Platform for the Semantic Web. In *Proceedings of the IEEE International Conference on Semantic Computing 2008 (ICSC 2008), Santa Clara, California, USA, August 4-7* (2008), IEEE, pp. 504–511.

[153] KAUPPINEN, T., KUITTINEN, H., TUOMINEN, J., SEPPÄLÄ, K., AND HYVÖNEN, E. Extending an Ontology by Analyzing Annotation Co-occurrences in a Semantic Cultural Heritage Portal. In *Proceedings of the ASWC 2008 Workshop on Collective Intelligence (ASWC-CI 2008) organized as a part of the 3rd Asian Semantic Web Conference (ASWC 2008), Bangkok, Thailand, February 2-5* (2009), pp. 1–6.

[154] KEEN, M., ACHARYA, A., BISHOP, S., HOPKINS, A., MILINSKI, S., NOTT, C., ROBINSON, R., ADAMS, J., AND VERSCHUEREN, P. *Patterns: Implementing an SOA using an enterprise service bus*. IBM, 2004.

[155] KENNEDY, J. B., KUKLA, R., AND PATERSON, T. Scientific names are ambiguous as identifiers for biological taxa: their context and definition are required for accurate data integration. In *Data Integration in the Life*

Sciences: Second International Workshop, DILS 2005, San Diego, CA, USA, July 20-22, 2005, Proceedings (2005), B. Ludäscher and L. Raschid, Eds., Springer-Verlag, pp. 80–95.

[156] KHAN, Z. C., AND KEET, C. M. The foundational ontology library ROMULUS. In *Model and Data Engineering: Third International Conference, MEDI 2013, Amantea, Italy, September 25-27, 2013, Proceedings* (2013), A. Cuzzocrea and S. Maabout, Eds., Springer-Verlag, pp. 200–211.

[157] KHATTAK, A. M., LATIF, K., AND LEE, S. Change management in evolving web ontologies. *Knowledge-Based Systems 37* (2013), 1–18.

[158] KIRSTEN, T., GROSS, A., HARTUNG, M., AND RAHM, E. GOMMA: a component-based infrastructure for managing and analyzing life science ontologies and their evolution. *Journal of Biomedical Semantics 2*, 6 (2011).

[159] KIRYAKOV, A., POPOV, B., TERZIEV, I., MANOV, D., AND OGNYANOFF, D. Semantic annotation, indexing, and retrieval. *Journal of Web Semantics 2*, 1 (2004), 49–79.

[160] KLEIN, M. *Change Management for Distributed Ontologies*. PhD thesis, Vrije Universiteit Amsterdam, Netherlands, Aug. 2004.

[161] KLEIN, M., AND FENSEL, D. Ontology versioning on the Semantic Web. In *Proceedings of the 1st Semantic Web Working Symposium (SWWS'01), Stanford University, California, USA, July 30 - August 1* (2001), I. F. Cruz, S. Decker, J. Euzenat, and D. L. McGuinness, Eds., pp. 75–91.

[162] KLESS, D., JANSEN, L., AND MILTON, S. A content-focused method for re-engineering thesauri into semantically adequate ontologies using OWL. *Semantic Web 7*, 5 (2016), 543–576.

[163] KNAPP, P., POLASZEK, A., AND WATSON, M. Spreading the word. *Nature 446* (2007), 261–262.

[164] KOZAKI, K., SUNAGAWA, E., KITAMURA, Y., AND MIZOGUCHI, R. A framework for cooperative ontology construction based on dependency management of modules. In *Proceedings of the First International Workshop on Emergent Semantics and Ontology Evolution (ESOE 2007), co-located with ISWC 2007 + ASWC 2007, Busan, Korea, November 12* (2007), L. L. Chen, P. Cudré-Mauroux, P. Haase, A. Hotho, and E. Ong, Eds., CEUR Workshop Proceedings, pp. 33–44.

[165] KURKI, J., AND HYVÖNEN, E. Collaborative metadata editor integrated with ontology services and faceted portals. In *Proceedings of the 1st Workshop on Ontology Repositories and Editors for the Semantic Web, Hersonissos, Crete, Greece, May 31* (2010), M. d'Aquin, A. García Castro, C. Lange, and K. Viljanen, Eds., CEUR Workshop Proceedings, pp. 7–11.

[166] LAPP, H., MORRIS, R. A., CATAPANO, T., HOBERN, D., AND MORRISON, N. Organizing our knowledge of biodiversity. *Bulletin of the American Society for Information Science and Technology 37*, 4 (2011), 38–42.

[167] LEADBETTER, A. The NERC Vocabulary Server version 2.0. Tech. rep., British Oceanographic Data Centre, 2012.

[168] LEDL, A., AND VOSS, J. Describing Knowledge Organization Systems in BARTOC and JSKOS. In *Proceedings of the 12th International Conference on Terminology and Knowledge Engineering, Copenhagen, Denmark, June 22-24* (2016), H. Erdman Thomsen, A. Pareja-Lora, and B. Nistrup Madsen, Eds., pp. 168–178.

[169] LEENHEER, P. D., AND MENS, T. ONTOLOGY EVOLUTION: State of the Art & Future Directions. *Ontology Theory Management and Design-Advanced Tools and Models 2*, 1 (2010), 1–47.

[170] LENAT, D. B. CYC: A Large-Scale Investment in Knowledge Infrastructure. *Communications of the ACM 38*, 11 (1995), 33–38.

[171] LEPAGE, D., VAIDYA, G., AND GURALNICK, R. Avibase - A database system for managing and organizing taxonomic concepts. *ZooKeys 135*, 420 (2014), 117–135.

[172] LI, Y., THOMPSON, S., TAN, Z., GILES, N., AND GHARIB, H. Beyond ontology construction; ontology services as online knowledge sharing communities. In *Semantic Web - ISWC 2003: Second International Semantic Web Conference, Sanibel Island, FL, USA, October 20-23, 2003, Proceedings* (2003), D. Fensel, K. Sycara, and J. Mylopoulos, Eds., Springer-Verlag, pp. 469–483.

[173] LUGHADHA, E. N. Towards a working list of all known plant species. *Philosophical transactions of the Royal Society of London B: Biological sciences 359*, 1444 (2004), 681–687.

[174] MAALI, F., AND ERICKSON, J. Data catalog vocabulary (DCAT). W3C recommendation, World Wide Web Consortium, 16 January 2014. `http://www.w3.org/TR/2014/REC-vocab-dcat-20140116/`.

[175] MAEDCHE, A., MOTIK, B., AND STOJANOVIC, L. Managing multiple and distributed ontologies on the Semantic Web. *The International Journal on Very Large Data Bases (The VLDB Journal) 12*, 4 (2003), 286–302.

[176] MAEDCHE, A., AND STAAB, S. Ontology learning in the semantic web. *IEEE Intelligent Systems 16*, 2 (2001), 72–79.

[177] MÄKELÄ, E., HYPÉN, K., AND HYVÖNEN, E. BookSampo—Lessons Learned in Creating a Semantic Portal for Fiction Literature. In *The Semantic Web – ISWC 2011: 10th International Semantic Web Conference, Bonn, Germany, October 23-27, 2011, Proceedings, Part II* (2011), L. Aroyo, C. Welty, H. Alani, J. Taylor, A. Bernstein, L. Kagal, N. Noy, and E. Blomqvist, Eds., Springer-Verlag, pp. 173–188.

[178] MÄKELÄ, E., RUOTSALO, T., AND HYVÖNEN, E. How to deal with massively heterogeneous cultural heritage data – lessons learned in CultureSampo. *Semantic Web 3*, 1 (2012), 85–109.

[179] MALAISÉ, V., AROYO, L., BRUGMAN, H., GAZENDAM, L., DE JONG, A., NEGRU, C., AND SCHREIBER, G. Evaluating a thesaurus browser for an audio-visual archive. In *Managing Knowledge in a World of Networks: 15th International Conference, EKAW 2006, Poděbrady, Czech Republic, October 2-6, 2006, Proceedings* (2006), S. Staab and V. Svátek, Eds., Springer-Verlag, pp. 272–286.

[180] MANGOLD, C. A survey and classification of semantic search approaches. *International Journal of Metadata, Semantics and Ontologies 2*, 1 (2007), 23–34.

[181] MANNING, C. D., RAGHAVAN, P., AND SCHÜTZE, H. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[182] MARCH, S. T., AND SMITH, G. F. Design and natural science research on information technology. *Decision Support Systems 15*, 4 (1995), 251–266.

[183] MARTÍNEZ-GONZÁLEZ, M. M., AND ALVITE-DÍEZ, M.-L. On the evaluation of thesaurus tools compatible with the Semantic Web. *Journal of Information Science 40*, 6 (2014), 711–722.

[184] MASCARDI, V., CORDÌ, V., AND ROSSO, P. A Comparison of Upper Ontologies. In *Proceedings of the 8th Workshop Dagli Oggetti agli Agenti: Agenti e Industria: Applicazioni tecnologiche degli agenti software (WOA 2007), Genova, Italy, September 24-25* (2007), pp. 55–64.

[185] MAYDEN, R. L. A hierarchy of species concepts: the denouement in the saga of the species problem. In *Species: The Units of Biodiversity*, M. F. Oaridge, H. A. Dawah, and M. R. Wilson., Eds. Chapman & Hall, 1997, pp. 381–424.

[186] MAYNARD, D., PETERS, W., D'AQUIN, M., AND SABOU, M. Change management for metadata evolution. In *Proceedings of the International Workshop on Ontology Dynamics (IWOD-07), the 4th European Semantic Web Conference (ESWC-07), Innsbruck, Austria, June 7* (2007), G. Flouris and M. d'Aquin, Eds., pp. 27–40.

[187] MCNEILL, J., BARRIE, F. R., BUCK, W. R., DEMOULIN, V., GREUTER, W., HAWKSWORTH, D. L., HERENDEEN, P. S., KNAPP, S., MARHOLD, K., PRADO, J., PRUD'HOMME VAN REINE, W. F., SMITH, G. F., WIERSEMA, J. H., AND TURLAND, N. J. *International Code of Nomenclature for algae, fungi, and plants (Melbourne Code) adopted by the Eighteenth International Botanical Congress Melbourne, Australia, July 2011*. Koeltz Scientific Books, Oberreifenberg, 2012.

[188] MILES, A., AND BECHHOFER, S. SKOS simple knowledge organization system reference. W3C recommendation, World Wide Web Consortium, 18 August 2009. `http://www.w3.org/TR/2009/REC-skos-reference-20090818/`.

[189] MILES, A., MATTHEWS, B., WILSON, M., AND BRICKLEY, D. SKOS Core: Simple Knowledge Organisation for the Web. In *Proceedings of the International Conference on Dublin Core and Metadata Applications 2008, Madrid, Spain, September 12-15* (2005), Dublin Core Metadata Initiative, pp. 3–10.

[190] MOSSAKOWSKI, T., KUTZ, O., AND CODESCU, M. Ontohub: A semantic repository for heterogeneous ontologies. In *Proceedings of the Theory Day in Computer Science (DACS 2014), satellite workshop of ICTAC 2014, Bucharest, Romania, September 17-19* (2014).

[191] NAGYPÁL, G., AND MÜLLER, W. Ontology Design. In *Semantics At Work, Ontology Management – Tools and Techniques*, R. Volz, Ed. Lulu Press, 2008.

[192] NAVIGLI, R., AND VELARDI, P. An Analysis of Ontology-based Query Expansion Strategies. *Information Retrieval* (2002), 42–49.

[193] NEUBERT, J. Bringing the "thesaurus for economics" on to the web of linked data. In *Proceedings of the Linked Data on the Web Workshop (LDOW2009), Madrid, Spain, April 20* (2009), C. Bizer, T. Heath, T. Berners-Lee, and K. Idehen, Eds., CEUR Workshop Proceedings.

[194] NILES, I., AND PEASE, A. Towards a Standard Upper Ontology. In *Proceedings of the international conference on Formal Ontology in Information Systems (FOIS-2001), Ogunquit, Maine, USA, October 17 - 19* (2001), ACM, pp. 2–9.

[195] NOY, N. F., AND KLEIN, M. Ontology evolution: Not the same as schema evolution. *Knowledge and Information Systems 6*, 4 (2004), 428–440.

[196] NOY, N. F., SHAH, N. H., WHETZEL, P. L., DAI, B., DORF, M., GRIFFITH, N., JONQUET, C., RUBIN, D. L., STOREY, M.-A., CHUTE, C. G., AND MUSEN, M. A. BioPortal: Ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research 37*, Web Server issue (2009), W170–W173.

[197] OBERLE, D., STAAB, S., STUDER, R., AND VOLZ, R. Supporting application development in the semantic web. *ACM Transactions on Internet Technology 5*, 2 (2005), 328–358.

[198] OREN, E., DELBRU, R., CATASTA, M., CYGANIAK, R., STENZHORN, H., AND TUMMARELLO, G. Sindice.com: A Document-oriented Lookup Index for Open Linked Data. *International Journal of Metadata, Semantics and Ontologies 3*, 1 (2008), 37–52.

[199] OREN, E., MÖLLER, K. H., SCERRI, S., HANDSCHUH, S., AND SINTEK, M. What are Semantic Annotations? Tech. rep., DERI Galway, 2006.

[200] PAGE, R. D. M. Taxonomic names, metadata, and the Semantic Web. *Biodiversity Informatics 3* (2006), 1–15.

[201] PAGE, R. D. M. Biodiversity informatics: The challenge of linking data and the role of shared identifiers. *Briefings in Bioinformatics 9*, 5 (2008), 345–354.

[202] PAGE, R. D. M. Linking NCBI to Wikipedia: A wiki-based approach. *PLoS Currents* (2011), 1–14.

[203] PAGE, R. D. M. BioNames: linking taxonomy, texts, and trees. *PeerJ 1* (2013).

[204] PAN, J., CRANEFIELD, S., AND CARTER, D. A lightweight ontology repository. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems - AAMAS '03* (2003), p. 632.

[205] PARR, C. S., GURALNICK, R., CELLINESE, N., AND PAGE, R. D. M. Evolutionary informatics: Unifying knowledge about the diversity of life. *Trends in Ecology and Evolution 27*, 2 (2012), 94–103.

[206] PARR, C. S., SCHULZ, K. S., HAMMOCK, J., WILSON, N., LEARY, P., RICE, J., AND CORRIGAN, JR., R. J. TraitBank: Practical semantics for organism attribute data. *Semantic Web 7*, 6 (2016), 577–588.

[207] PARR, C. S., WILSON, N., LEARY, P., SCHULZ, K. S., LANS, K., WALLEY, L., HAMMOCK, J. A., GODDARD, A., RICE, J., STUDER, M., HOLMES, J. T. G., AND CORRIGAN, JR., R. J. The Encyclopedia of Life v2: Providing Global Access to Knowledge About Life on Earth. *Biodiversity Data Journal 2* (2014).

[208] PATTERSON, D. J., COOPER, J., KIRK, P. M., PYLE, R. L., AND REMSEN, D. P. Names are key to the big new biology. *Trends in Ecology and Evolution 25*, 12 (2010), 686–691.

[209] PATTERSON, D. J., REMSEN, D., MARINO, W. A., AND NORTON, C. Taxonomic Indexing—Extending the Role of Taxonomy. *Systematic Biology 55*, 3 (2006), 367–373.

[210] PEFFERS, K., ROTHENBERGER, M., TUUNANEN, T., AND VAEZI, R. Design Science Research Evaluation. In *Design Science Research in Information Systems. Advances in Theory and Practice: 7th International Conference, DESRIST 2012, Las Vegas, NV, USA, May 14-15, 2012. Proceedings* (2012), K. Peffers, M. Rothenberger, and B. Kuechler, Eds., Springer-Verlag, pp. 398–410.

[211] PEFFERS, K., TUUNANEN, T., ROTHENBERGER, M. A., AND CHATTERJEE, S. A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems 24*, 3 (2007), 45–77.

[212] PEREIRA, R., RICHARDS, K., HOBERN, D., HYAM, R., BELBIN, L., AND BLUM, S. TDWG life sciences identifiers (LSID) applicability statement. TDWG draft standard, Biodiversity Information Standards (TDWG), 2009. `http://www.tdwg.org/standards/150`.

[213] PHIPPS, J., AND HILLMANN, D. I. The Open Metadata Registry: An Update. *Bulletin of the Association for Information Science and Technology 37*, 4 (2011), 35–37.

[214] PULLAN, M. R., WATSON, M. F., KENNEDY, J. B., RAGUENAUD, C., AND HYAM, R. The Prometheus Taxonomic Model: A Practical Approach to Representing Multiple Classifications. *Taxon 49*, 1 (2000), 55–75.

[215] PYLE, R. L., AND MICHEL, E. ZooBank: Developing a nomenclatural tool for unifying 250 years of biological information. *Zootaxa*, 1950 (2008), 39–50.

[216] QING, Z. Building a platform for linked open thesauri. *Library Hi Tech 31*, 4 (2013), 620–637.

[217] QU, Y., AND CHENG, G. Falcons Concept Search: A Practical Search Engine for Web Ontologies. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 41*, 4 (2011), 810–816.

[218] REES, T. Taxamatch, an algorithm for near ('fuzzy') matching of scientific names in taxonomic databases. *PLoS ONE 9*, 9 (2014).

[219] REMSEN, D. The use and limits of scientific names in biological informatics. *ZooKeys 550* (2016), 207–223.

[220] REMSEN, D. P., DÖRING, M., AND ROBERTSON, T. GBIF GNA profile reference guide for Darwin Core Archive, core terms and extensions. Tech. rep., Global Biodiversity Information Facility (GBIF), Copenhagen, Denmark, 2011. version 1.2, released on 1 April 2011.

[221] RICHARDS, K., WHITE, R., NICOLSON, N., AND PYLE, R. A beginner's guide to persistent identifiers. Tech. rep., Global Biodiversity Information Facility (GBIF), Copenhagen, Denmark, 2011. version 1.0, released on 9 February 2011.

[222] RILEY, J. Understanding Metadata. Primer, National Information Standards Organization, 2017.

[223] ROSKOV, Y., ABUCAY, L., ORRELL, T., NICOLSON, D., FLANN, C., BAILLY, N., KIRK, P., BOURGOIN, T., DEWALT, R., DECOCK, W., AND DE WEVER, A. Species 2000 & ITIS Catalogue of Life. Digital resource, Species 2000: Naturalis, Leiden, Netherlands, 2017. 27th February 2017, `http://www.catalogueoflife.org/col`.

[224] RUEDA, C., BERMUDEZ, L., AND FREDERICKS, J. The MMI Ontology Registry and Repository: A Portal for Marine Metadata Interoperability. In *Proceedings of the Oceans Conference 2009, Biloxi, Mississippi, USA, 26–29 October* (2009), IEEE, pp. 1854–1859.

[225] SARKAR, I. N. Biodiversity informatics: Organizing and linking information across the spectrum of life. *Briefings in Bioinformatics 8*, 5 (2007), 347–357.

[226] SCHANDL, T., AND BLUMAUER, A. PoolParty: SKOS thesaurus management utilizing linked data. In *The Semantic Web: Research and Applications: 7th Extended Semantic Web Conference, ESWC 2010, Heraklion, Crete, Greece, May 30 – June 3, 2010, Proceedings, Part II* (2010), L. Aroyo, G. Antoniou, E. Hyvönen, A. ten Teije, H. Stuckenschmidt, L. Cabral, and T. Tudorache, Eds., Springer-Verlag, pp. 421–425.

[227] SCHREIBER, A. T., DUBBELDAM, B., WIELEMAKER, J., AND WIELINGA, B. Ontology-based photo annotation. *IEEE Intelligent Systems and Their Applications 16*, 3 (2001), 66–74.

[228] SCHULZ, S., STENZHORN, H., AND BOEKER, M. The ontology of biological taxa. *Bioinformatics 24*, 13 (2008), 313–321.

[229] SEGERS, H., DE SMET, W. H., FISCHER, C., FONTANETO, D., MICHALOUDI, E., WALLACE, R. L., AND JERSABEK, C. D. Towards a List of Available Names in Zoology, partim Phylum Rotifera. *Zootaxa 3179* (2012), 61–68.

[230] ŠEVČENKO, M. Online Presentation of an Upper Ontology. In *Proceedings of Znalosti 2003, Ostrava, Czech Republic, February 19-21* (2003), pp. 153–162.

[231] SHAW, R., RABINOWITZ, A., GOLDEN, P., AND KANSA, E. A sharing-oriented design strategy for networked knowledge organization systems. *International Journal on Digital Libraries 17*, 1 (2016), 49–61.

[232] SHIRI, A. Thesauri: Introduction and Recent Developments. In *Powering search: The role of thesauri in new information environments*. Information Today, 2012, pp. 1–41.

[233] SHVAIKO, P., AND EUZENAT, J. Ontology Matching: State of the Art and Future Challenges. *IEEE Transactions on Knowledge and Data Engineering 25*, 1 (2013), 158–176.

[234] SLIMANI, T. Semantic Annotation: The Mainstay of Semantic Web. *International Journal of Computer Applications Technology and Research 2*, 6 (2013), 763–770.

[235] SMITH, B., ASHBURNER, M., ROSSE, C., BARD, J., BUG, W., CEUSTERS, W., GOLDBERG, L. J., EILBECK, K., IRELAND, A., MUNGALL, C. J., OBI CONSORTIUM, LEONTIS, N., ROCCA-SERRA, P., RUTTENBERG, A., SANSONE, S.-A., SCHEUERMANN, R. H., SHAH, N., WHETZEL, P. L., AND LEWIS, S. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology 25*, 11 (2007), 1251–1255.

[236] SOERGEL, D. The Arts and Architecture Thesaurus (AAT): A Critical Appraisal. *Visual Resources 10*, 4 (1995), 369–400.

[237] SOLOMOU, G., AND PAPATHEODOROU, T. The use of SKOS vocabularies in digital repositories: The DSpace case. In *ICSC 2010: 2010 IEEE Fourth International Conference on Semantic Computing, Proceedings, Pittsburgh, Pennsylvania, USA, September 22-24* (2010), IEEE, pp. 542–547.

[238] SOUZA, R. R., TUDHOPE, D., AND ALMEIDA, M. B. The KOS spectra A tentative typology of knowledge organization systems. In *Paradigms and conceptual systems in knowledge organization: Proceedings of the Eleventh International ISKO Conference, Rome, Italy, 23-26 February* (2010), C. Gnoli and F. Mazzocchi, Eds., Ergon-Verlag, pp. 122–128.

[239] STAAB, S., MAEDCHE, A., AND HANDSCHUH, S. An annotation framework for the semantic Web. In *Proceedings of the First Workshop on Multimedia Annotation, Tokyo, Japan* (2001).

[240] STAAB, S., AND STUDER, R., Eds. *Handbook on Ontologies*. International Handbooks on Information Systems. Springer-Verlag, Berlin Heidelberg, 2004.

[241] STOJANOVIC, L. *Methods and Tools for Ontology Evolution*. PhD thesis, University of Karlsruhe, Germany, Aug. 2004.

[242] STOJANOVIC, L., MAEDCHE, A., MOTIK, B., AND STOJANOVIC, N. User-driven Ontology Evolution Management. In *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web: 13th International Conference, EKAW 2002, Sigüenza, Spain, October 1–4, 2002, Proceedings* (2002), A. Gómez-Pérez and V. R. Benjamins, Eds., Springer-Verlag, pp. 285–300.

[243] STUCKENSCHMIDT, H., AND KLEIN, M. Integrity and change in modular ontologies. In *IJCAI'03: Proceedings of the 18th international joint conference on Artificial intelligence* (2003), Morgan Kaufmann Publishers, pp. 900–905.

[244] STUDER, R., BENJAMINS, V. R., AND FENSEL, D. Knowledge engineering: Principles and methods. *Data & Knowledge Engineering 25*, 1-2 (1998), 161–197.

[245] SUMMERS, E., ISAAC, A., REDDING, C., AND KRECH, D. LCSH, SKOS and Linked Data. In *Proceedings of the International Conference on Dublin Core and Metadata Applications 2008, Berlin, Germany, September 22-26* (2008), Dublin Core Metadata Initiative, pp. 25–33.

[246] SUNAGAWA, E., KOZAKI, K., KITAMURA, Y., AND MIZOGUCHI, R. An environment for distributed ontology development based on dependency management. In *The Semantic Web - ISWC 2003: Second International Semantic Web Conference, Sanibel Island, FL, USA, October 20-23, 2003, Proceedings* (2003), D. Fensel, K. Sycara, and J. Mylopoulos, Eds., Springer-Verlag, pp. 453–468.

[247] SUOMINEN, O., HYVÖNEN, E., VILJANEN, K., AND HUKKA, E. HealthFinland-A national semantic publishing network and portal for health information. *Journal of Web Semantics 7*, 4 (2009), 287–297.

[248] SUOMINEN, O., JOHANSSON, A., YLIKOTILA, H., TUOMINEN, J., AND HYVÖNEN, E. Vocabulary Services Based on SPARQL Endpoints: ONKI Light on SPARQL. In *Poster proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2012), Galway, Ireland, October 8-12* (2012).

[249] SUOMINEN, O., PESSALA, S., TUOMINEN, J., LAPPALAINEN, M., NYKYRI, S., YLIKOTILA, H., FROSTERUS, M., AND HYVÖNEN, E. Deploying national ontology services: From ONKI to Finto. In *Proceedings of the Industry Track at the International Semantic Web Conference 2014, Riva del Garda, Italy, October 19-23* (2014), A. Polleres, A. Garcia, and R. Benjamins, Eds., CEUR Workshop Proceedings.

[250] TAM, A. M., AND LEUNG, C. H. C. Structured Natural-Language Description for Semantic Content Retrieval. *Journal of the American Society for Information Science and Technology 52*, 11 (2001), 930–937.

[251] TAXONOMIC NAMES AND CONCEPTS INTEREST GROUP. Taxonomic concept transfer schema (TCS). TDWG current (2005) standard, Biodiversity Information Standards (TDWG), 2006. [`http://www.tdwg.org/standards/117`.

[252] TENNIS, J. T., AND SUTTON, S. A. Extending the Simple Knowledge Organization System for Concept Management in Vocabulary Development Applications. *Journal of the American Society for Information Science and Technology 59*, 1 (2008), 25–37.

[253] THESSEN, A. E., CUI, H., AND MOZZHERIN, D. Applications of natural language processing in biodiversity science. *Advances in Bioinformatics 2012* (2012).

[254] THOMAS, E., PAN, J. Z., AND SLEEMAN, D. ONTOSEARCH2: Searching ontologies semantically. In *Proceedings of the OWLED 2007 Workshop on OWL: Experiences and Directions, Innsbruck, Austria, June 6-7* (2007), C. Golbreich, A. Kalyanpur, and B. Parsia, Eds., CEUR Workshop Proceedings.

[255] TOPQUADRANT. Controlled vocabularies, taxonomies, and thesauruses (and ontologies). Whitepaper, TopQuadrant, 2013.

[256] TUDHOPE, D., ALANI, H., AND JONES, C. Augmenting thesaurus relationships: Possibilities for retrieval. *Journal of Digital Information 1*, 8 (2001).

[257] TUDHOPE, D., AND BINDING, C. Toward Terminology Services: Experiences with a Pilot Web Service Thesaurus Browser. *Bulletin of the American Society for Information Science and Technology 32*, 5 (2006), 6–9.

[258] TUOMINEN, J., SALONOJA, M., VILJANEN, K., AND HYVÖNEN, E. A user interface for ontology repositories. In *Proceedings of the 1st Workshop on Ontology Repositories and Editors for the Semantic Web, Hersonissos, Crete, Greece, May 31* (2010), M. d'Aquin, A. García Castro, C. Lange, and K. Viljanen, Eds., CEUR Workshop Proceedings, pp. 17–21.

[259] TUOMINEN, J., AND VILJANEN, K. Ontology services: results of the ONKI user inquiry, presentation at the FinnONTO 2.0 project board meeting 13.12.2011, 2011.

[260] USCHOLD, M., AND GRUNINGER, M. Ontologies: Principles, methods and applications. *Knowledge Engineering Review 11*, 2 (1996), 93–136.

[261] VAN ASSEM, M., MALAISÉ, V., MILES, A., AND SCHREIBER, G. A method to convert thesauri to SKOS. In *The Semantic Web: Research and Applications: 3rd European Semantic Web Conference, ESWC 2006, Budva, Montenegro, June 11-14, 2006, Proceedings* (2006), Y. Sure and J. Domingue, Eds., Springer-Verlag, pp. 95–109.

[262] VAN ASSEM, M., MENKEN, M. R., SCHREIBER, G., WIELEMAKER, J., AND WIELINGA, B. A Method for Converting Thesauri to RDF/OWL. In *The Semantic Web – ISWC 2004: Third International Semantic Web Conference, Hiroshima, Japan, November 7-11, 2004, Proceedings* (2004), S. A. McIlraith, D. Plexousakis, and F. van Harmelen, Eds., Springer-Verlag, pp. 17–31.

[263] VAN DEN BOSCH, A., BOGERS, T., AND DE KUNDER, M. Estimating search engine index size variability: a 9-year longitudinal study. *Scientometrics 107*, 2 (2016), 839–856.

[264] VAN DER MEIJ, L., ISAAC, A., AND ZINN, C. A web-based repository service for vocabularies and alignments in the cultural heritage domain. In *The Semantic Web: Research and Applications: 7th Extended Semantic Web Conference, ESWC 2010, Heraklion, Crete, Greece, May 30 –June 3, 2010, Proceedings, Part I* (2010), L. Aroyo, G. Antoniou, E. Hyvönen, A. ten Teije, H. Stuckenschmidt, L. Cabral, and T. Tudorache, Eds., Springer-Verlag, pp. 394–409.

[265] VAN OSSENBRUGGEN, J., AMIN, A., AND HILDEBRAND, M. Why evaluating semantic web applications is difficult. In *Proceedings of the Fifth International Workshop on Semantic Web User Interaction (SWUI 2008), collocated with CHI 2008, Florence, Italy, April 5* (2008), D. Degler, M. Schraefel, J. Golbeck, A. Bernstein, and L. Rutledge, Eds., CEUR Workshop Proceedings.

[266] VANDEN BERGHE, E., CORO, G., BAILLY, N., FIORELLATO, F., ALDEMITA, C., ELLENBROEK, A., AND PAGANO, P. Retrieving taxa names from large biodiversity data collections using a flexible matching workflow. *Ecological Informatics 28* (2015), 29–41.

[267] VANDENBUSSCHE, P.-Y., ATEMEZING, G. A., POVEDA-VILLALÓN, M., AND VATANT, B. Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web. *Semantic Web*, In press (2016).

[268] VILJANEN, K., HYVÖNEN, E., MÄKELÄ, E., SUOMINEN, O., AND TUOMINEN, J. Mash-up ontology services for the semantic web. In *Demo track at the European Semantic Web Conference ESWC 2007, Innsbruck, Austria, June 4* (2007).

[269] VIZINE-GOETZ, D., HOUGHTON, A., AND CHILDRESS, E. Web Services for Controlled Vocabularies. *Bulletin of the American Society for Information Science and Technology 32*, 5 (2006), 9–12.

[270] VOORHEES, E. M. Expansion using lexical-semantic relations. In *SIGIR'94 Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval* (1994), pp. 61–69.

[271] WALLS, R. L., DECK, J., GURALNICK, R., BASKAUF, S., BEAMAN, R., BLUM, S., BOWERS, S., BUTTIGIEG, P. L., DAVIES, N., ENDRESEN, D., GANDOLFO, M. A., HANNER, R., JANNING, A., KRISHTALKA, L., MATSUNAGA, A., MIDFORD, P., MORRISON, N., TUAMA, É. Ó., SCHILDHAUER, M., SMITH, B., STUCKY, B. J., THOMER, A., WIECZOREK, J., WHITACRE, J., AND WOOLEY, J. Semantics in Support of Biodiversity Knowledge Discovery: An Introduction to the Biological Collections Ontology and Related Ontologies. *PLoS ONE 9*, 3 (2014).

[272] WANG, S., SCHLOBACH, S., AND KLEIN, M. Concept drift and how to identify it. *Journal of Web Semantics 9*, 3 (2011), 247–265.

[273] WANG, X., ALMEIDA, J. S., AND OLIVEIRA, A. L. Ontology design principles and normalization techniques in the web. In *Data Integration in the Life Sciences: 5th International Workshop, DILS 2008, Evry, France, June 25-27, 2008, Proceedings* (2008), A. Bairoch, S. Cohen-Boulakia, and C. Froidevaux, Eds., Springer-Verlag, pp. 28–43.

[274] WANG, Y.-C., VANDENDORPE, J., AND EVENS, M. Relational thesauri in information retrieval. *Journal of the American Society for Information Science 36*, 1 (1985), 15–27.

[275] WHETZEL, P. L., NOY, N. F., SHAH, N. H., ALEXANDER, P. R., NYULAS, C., TUDORACHE, T., AND MUSEN, M. A. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Research 39*, Web Server issue (2011), W541–W545.

[276] WHITE, H., WILLIS, C., AND GREENBERG, J. HIVEing: The Effect of a Semantic Web technology on Inter-Indexer Consistancy. *Journal of Documentation 70*, 3 (2014), 307–329.

[277] WIECZOREK, J., BLOOM, D., GURALNICK, R., BLUM, S., DÖRING, M., GIOVANNI, R., ROBERTSON, T., AND VIEGLAIS, D. Darwin Core: An

evolving community-developed biodiversity data standard. *PLoS ONE 7*, 1 (2012).

[278] WIELINGA, B. J., SCHREIBER, A. T., WIELEMAKER, J., AND SANDBERG, J. A. C. From thesaurus to ontology. In *Proceedings of the 1st International Conference on Knowledge Capture: K-CAP 2001, Victoria, Canada, October 21-23* (2001), ACM, pp. 194–201.

[279] XIANG, Z., MUNGALL, C., RUTTENBERG, A., AND HE, Y. Ontobee: A linked data server and browser for ontology terms. In *Proceedings of the 2nd International Conference on Biomedical Ontology, Buffalo, NY, USA, July 26-30* (2011), O. Bodenreider, M. E. Martone, and A. Ruttenberg, Eds., CEUR Workshop Proceedings.

[280] YTOW, N., MORSE, D. R., AND ROBERTS, D. M. Nomencurator: a nomenclatural history model to handle multiple taxonomic views. *Biological Journal of the Linnean Society 73* (2001), 81–98.

[281] ZABLITH, F., ANTONIOU, G., D'AQUIN, M., FLOURIS, G., KONDYLAKIS, H., MOTTA, E., PLEXOUSAKIS, D., AND SABOU, M. Ontology evolution: a process-centric survey. *The Knowledge Engineering Review 30*, 1 (2015), 45–75.

[282] ZAPILKO, B., SCHAIBLE, J., MAYR, P., AND MATHIAK, B. TheSoz: A SKOS representation of the thesaurus for the social sciences. *Semantic Web 4*, 3 (2013), 257–263.

[283] ZENG, M. L., AND CHAN, L. M. Trends and issues in establishing interoperability among knowledge organization systems. *Journal of the American Society for Information Science and Technology 55*, 5 (2004), 377–395.

[284] ZENG, M. L., AND HODGE, G. Developing a Dublin Core Application Profile for the Knowledge Organization Systems (KOS) Resources. *Bulletin of the American Society for Information Science & Technology 37*, 4 (2011), 30–34.

[285] ZERMOGLIO, P. F., GURALNICK, R. P., AND WIECZOREK, J. R. A standardized reference data set for vertebrate taxon name resolution. *PLoS ONE 11*, 1 (2016).

[286] ZHAO, L., AND ICHISE, R. Ontology Integration for Linked Data. *Journal on Data Semantics 3*, 4 (2014), 237–254.

Aalto-DD 101/2017

Knowledge organization systems such as thesauri and ontologies are used for improving the findability of information. They harmonize the content descriptions and provide interoperability in and between information systems, such as museum collections, digital libraries, and online stores.

Ontology services facilitate the use of knowledge organization systems in applications. This thesis presents methods for cost-effective ontology access for content indexers, information searchers, and application developers.

The developed tools are provided as a public living lab service, and they facilitate the simultaneous use of ontologies from different domains, such as music, economics, and agriculture. A method is presented for managing the changes of biological taxonomies as an ontology.

Cover image:
The concept hierarchy of the top three levels of the KOKO ontology

9 789526 074566

ISBN 978-952-60-7456-6 (printed)
ISBN 978-952-60-7455-9 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Science
Department of Computer Science
www.aalto.fi

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

DOCTORAL
DISSERTATIONS