Aalto University
School of Science
Master's Programme in Computer, Communication and Information Sciences

Annastiina Ahola

Developing a tool for information retrieval and research purposes utilizing BookSampo data

Master's Thesis
Espoo, February 27, 2023

Supervisor: Professor Eero Hyvönen
Advisor: Heikki Rantala M.Sc., Aalto University

Aalto University
School of Science
Master's Programme in Computer, Communication and Information Sciences

ABSTRACT OF MASTER'S THESIS

Author:	Annastiina Ahola		
Title: Developing a tool for information retrieval and research purposes utilizing BookSampo data			
Date:	February 27, 2023	**Pages:**	103
Major:	Computer Science	**Code:**	SCI3042
Supervisor:	Professor Eero Hyvönen		
Advisor:	Heikki Rantala M.Sc.		

The goal of this thesis was to develop a new tool for information retrieval and research purposes based on the BookSampo data set. The current BookSampo portal uses Linked Data but offers limited search capabilities to the user that do not fully utilize the possibilities of Linked Data. The goal of the new portal was to offer the users both ways to visualize the data as well as a facet-based search, where the user could search and filter the result data set using different facets.

The portal was developed using the Sampo Model based Sampo-UI framework that makes it possible to build new semantic portals without requiring a lot of actual programming work. Developing new semantic portals using the Sampo-UI framework is based on editing configuration files that define the wanted components as well as the data location using a SPARQL endpoint. The data used in the portal was the BookSampo dataset, the quality of which was also assessed during the development process.

The final product is a Sampo-UI based semantic portal that uses unmodified BookSampo data, which was assessed to be of good quality considering the sheer volume of the data set. The portal offers the user five different perspectives that can be used for looking at and searching for data. The perspectives have their own visualizations developed for them based on the annotation coverage of the data that can be used for viewing the data in different formats, e.g., as a chart or on a map.

Sampo-UI framework fit the purpose well and makes it possible to develop the portal easily further in the future. The BookSampo data has and will be receiving the needed fixes, after which it could be published to public use and then improved based on the feedback received from users.

Keywords:	semantic web, linked data, digital libraries, bibliographical linked data, semantic portals
Language:	English

Aalto-yliopisto
Perustieteiden korkeakoulu
korkeakoulu

Aalto-yliopisto
Perustieteiden korkeakoulu
Tieto-, tietoliikenne- ja informaatiotekniikan maisteriohjelma

DIPLOMITYÖN TIIVISTELMÄ

Tekijä:	Annastiina Ahola
Työn nimi:	

Tiedonhakuun ja tutkimukseen tarkoitetun työkalun kehittäminen hyödyntäen Kirjasampo-aineistoja

Päiväys:	27. helmikuuta 2023	**Sivumäärä:**	103
Pääaine:	Tietotekniikka	**Koodi:**	SCI3042
Valvoja:	Professori Eero Hyvönen		
Ohjaaja:	Filosofian maisteri Heikki Rantala		

Tämän diplomityön tavoitteena oli kehittää uusi tiedonhakuun ja tutkimukseen soveltuva työkalu, joka hyödyntää Kirjasampo-aineistoja. Kirjasammon nykyinen portaali käyttää linkitettyä dataa, mutta tarjoaa käyttäjille rajoitetut hakumahdollisuudet, jotka eivät täysin hyödynnä kaikkia linkitetyn datan mahdollisuuksia. Uuden portaalin tarkoituksena oli tarjota käyttäjille niin aineston visualisointeja kuin myös fasettipohjaisen haun, jolla käyttäjä voi fasettien avulla hakea ja rajata tuloksena saatavaa dataa.

Portaali kehitettiin käyttämällä Sampo-malliin perustuvaa Sampo-UI-ohjelmointikehystä, joka mahdollistaa semanttisten portaalien kehityksen ilman suurempaa varsinaista ohjelmointiurakkaa. Semanttisten portaalien kehittäminen Sampo-UI-ohjelmointikehyksellä perustuu konfiguraatiotiedostojen muokkaamiseen, missä määritetään niin portaalin halutut komponentit kuin myös datan sijainti SPARQL-rajapinnan avulla. Kehitettävän portaalin datana toimi Kirjasampo-aineisto, jonka laatua kartoitettiin ja arvioitiin kehityksen ohella.

Lopputuloksena syntyi Sampo-UI-pohjainen portaali, joka käyttää muokkaamattomia Kirjasampo-aineistoja. Käytetyt aineistot osoittautuivat laadultaan tarkoitukseen hyviksi, kun otetaan huomioon datasetin suuruus. Portaali tarjoaa käyttäjille viisi eri perspektiiviä, joista tarkastella ja hakea dataa. Perspektiiveille on datan annotointien kattavuuden perusteella kehitetty omat visualisointinsa, joilla juuri kyseisen perspektiivin dataa voi tarkastella erilaisilla tavoilla kuten esimerkiksi kaaviomuodossa tai kartalla.

Sampo-UI-ohjelmointikehys soveltui tehtävään hyvin ja mahdollistaa portaalin helpon jatkokehityksen tulevaisuudessa. Kirjasammon dataan on tehty ja tullaan tekemään tarvittavia korjauksia, joiden jälkeen portaali olisi mahdollista julkaista yleisölle ja kehittää käyttäjiltä saatavan palautteen perusteella.

Asiasanat:	semanttinen web, linkitetty data, digitaaliset kirjastot, bibliografinen linkitetty data, semanttiset portaalit
Kieli:	Englanti

Acknowledgements

First and foremost, I would like to offer my deepest gratitude to Heikki Rantala and Professor Eero Hyvönen for all the help and guidance I have received during the process of writing this master's thesis. I would also like to thank Aalto University, CSC – IT Center For Science and everyone at Finnish Public Libraries who made this project possible through funding and other resources.

I'm very grateful to Esko Ikkala for introducing me to the Sampo-UI framework as well as work at Semantic Computing Research Group (SeCo). I'd also like to offer my thanks to all my other colleagues at SeCo for all the helpful comments and ideas throughout the project as well as for just being there and brightening my days.

Finally, I would like to thank my family for all the support. I'd like to offer special thanks to my father for providing the best motivation to finish the writing process of this thesis – to graduate faster than he did in his time.

Espoo, February 27, 2023

Annastiina Ahola

Abbreviations and Acronyms

CH	Cultural Heritage
DH	Digital Humanities
GUI	Graphical User Interface
IRI	Internationalized Resource Identifier
LD	Linked Data
LDF	Linked Data Finland
LOD	Linked Open Data
N3	Notation3
OWL	The Web Ontology Language
RDF	Resource Description Framework
SKOS	Simple Knowledge Organization System
SPARQL	SPARQL Protocol and RDF Query Language
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
URN	Uniform Resource Name
WWW	World Wide Web
W3C	World Wide Web Consortium

Contents

Chapter 1

Introduction

1.1 Problem statement

The current BookSampo portal[1] managed by Finnish Public Libraries[2] is one of the most popular portals from the Sampo series of semantic portals[3] with about 1.6 million visits and 1.1 million distinct visitors per year [2, 32]. It utilizes linked bibliographical data to bring users better search functionality and connect relevant data together [41]. The linkedness of the data is however not used to its fullest potential. For example, faceted search [3] used in other Sampo portals is not available. The portal also does not lend itself well to using it for literary research purposes, though the data itself has the potential for that.

Now, after more than 10 years since the first release of the current portal, the BookSampo portal and its data is being revisited. One of the parts of this endeavor was to build a new semantic portal for the data to accompany the original traditional BookSampo portal. The idea of a semantic portal is to enhance the user experience by utilizing semantic web technologies and the linked nature of the data to better the search and exploration capabilities. This thesis outlines the development of this new BookSampo semantic portal that could be used as a tool for both intelligent information retrieval and research related to Finnish literature as well as quality analysis on the BookSampo data used in both this new and the old portal.

The research questions of this thesis are the following:

1. How can the user utilize BookSampo data for intelligent information

[1] `https://www.kirjasampo.fi/`

[2] `https://www.kirjastot.fi/`

[3] Information on Sampo portals and user statistics for different Sampo portals available at: `https://seco.cs.aalto.fi/applications/sampo/`

retrieval and research?

(a) How should the BookSampo knowledge graph be visualized and what kind of visualizations would be the best for these purposes?

(b) What kind of searches/information retrieval should the user be able to do using BookSampo data and the developed user interface?

2. How to configure a new semantic portal using the Sampo-UI framework and a knowledge graph?
3. How to deal with problematic (e.g., missing labels, hierarchy) and/or incomplete data when developing portals like these?
4. What is the quality of the BookSampo data?

1.2 Structure of the thesis

The first few chapters cover the background related to developing a semantic portal like the one presented in the thesis. Chapter 2 gives the brief background of the Semantic Web and Linked Data as well as discusses projects utilizing bibliographical Linked Data. Chapter 3 presents the Sampo Model and Sampo-UI framework that are used in the development of the new BookSampo portal. The BookSampo data used in the portal and its origins are covered in Chapter 4 in which the quality of the data is also assessed.

The later chapters deal with the design, implementation, and evaluation of the created portal. Chapter 5 overviews the design and the choices made regarding that. Chapter 6 covers the implementation itself and presents the created portal. The portal is evaluated in Chapter 7. Finally, Chapter 8 gives the conclusions of the thesis. Appendices are included at the end of the thesis.

Chapter 2

Semantic Web

This chapter starts with an introduction of the Semantic Web in general and its principles. After that it goes over the usage of Linked Data in the Cultural Heritage (CH) field as well as specifically in bibliographical contexts.

2.1 Semantic Web and its origins

This section briefly goes over the origins of the Semantic Web and its evolution throughout the years. The first subsection focuses on the origins and the second subsection afterwards introduces some of the most integral Semantic Web technologies.

2.1.1 Origins

The major turning point for the Semantic Web and its development happened in 2001 when Berners-Lee et al. [11] described the idea of the Semantic Web in their article titled *The Semantic Web*. In comparison to the classic human-readable Web, on the Semantic Web human-readable sites should be augmented with machine-readable data. This data should be semantic by nature that is achieved through adding metadata. The metadata in turn should provide definitions for the terms or reasoning rules for the data. While the article pitches this idea of the Semantic Web as a new form content on the World Wide Web (WWW), the idea itself was not new as Berners-Lee had already mentioned his vision of it in 1994 in the First International Conference on the World Wide Web (WWW1) and in his *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor* from 1999 [10, 55].

Berners-Lee et al. [11] describe the Semantic Web as its own *killer app*, but the initial adoption of the Semantic Web was not without its difficulties and even regarded as failed or doomed to fail by some [26, 55]. Nonetheless, efforts were made by both the users of the WWW as well as some large companies to make use of the possibilities of the Semantic Web technologies that were being developed by the World Wide Web Consortium (W3C) [18].

In Finland Semantic Web development was kicked off with the *Semantic Web Kick-off in Finland* conference held in the autumn of 2001 and the development of semantic content itself started in 2002 [34]. The National Semantic Web Ontology project in Finland (FinnONTO, 2003–2012) project began in 2003 with the aim of building Semantic Web content infrastructure in Finland [29, 36, 59]. This was done by developing metadata standards, ontologies and ontology services, tools for Semantic Web content and its creation as well as pilot applications [36]. An important part of building this Finnish Semantic Web infrastructure was building national ontologies for describing data. The General Finnish Thesaurus YSA[1] was turned into the General Finnish Ontology YSO[2] and then combined with fifteen other domain specific ontologies under a national Linked Open Ontology Cloud (LOO), KOKO, where YSO functioned as the general upper ontology (GUO) [20, 29]. The national Finnish Ontology Library Service ONKI pilot was also developed under the FinnONTO project [61].

While in the initial years of Semantic Web field's development the focus centered around ontologies, during mid-2000s the focus switched instead to Linked Data (LD) [25]. With the rise of LD and Linked Open Data (LOD) the Semantic Web gained some new wind beneath its wings and mainstream services, such as Facebook and Google, started to become interested in utilizing Semantic Web technologies and their benefits [26]. To ensure that published Linked Data had the best possible quality, ways to assess it were developed. There are principles for publishers to follow such as the FAIR-principles[3] [62] (shown in Table 2.1) as well as different star models to assess published Linked Data quality. The 5-star model[4] was proposed by Berners-Lee in 2006. [8]. The 5-star model lists five requirements (listed in Table 2.2) for the data nature with all requirements met meaning the Linked Data gets the maximum 5-star rating. For the Linked Data to also qualify as good linked open data (LOD), the data must be published with an open license to qualify for any stars at all [8]. In 2014 Linked Data Finland[5] proposed the 7-star model, an extension to the 5-star model, which adds the two additional

[1] https://finto.fi/ysa/en/

[2] https://finto.fi/yso/en/

[3] https://www.go-fair.org/fair-principles/

[4] https://5stardata.info/

[5] https://www.ldf.fi/

requirements for 6 and 7 stars as presented in Table 2.3 [35].

In Finland the FinnONTO project's work was continued with the Linked Data Finland (LDF) project from 2012 to 2014 and was then succeeded by Linked Open Data Science Service (LODSCI) from 2015 to 2017 and the Linked Open Data Infrastructure for Digital Humanities in Finland (LODI4DH)[6], a joint initiative of Aalto Semantic Computing Research Group SeCo and University of Helsinki, Helsinki Centre for Digital Humanities HELDIG [31, 33]. During this time, the national ontology service prototype ONKI was deployed into Finnish Thesaurus and Ontology Service Finto maintained by the National Library of Finland and the service was opened to the public at the start of 2014 [59].

In 2019 the general opinion of the Semantic Web community was that Berners-Lee's original vision was yet to be fully realized, but the members of the community felt hopeful regarding a more widespread adoption of the idea of the Semantic Web in the future [26]. While some tools have been developed throughout the years, the field is still very academic in nature and practical tools to help with the adoption of the idea are still few and far between [26]. A lot of the Semantic Web research is done in Europe, where projects have received significant funding from the European Union [25]. On the other hand, after the first half decade of the 2000s, Semantic Web related research in U.S. has been lacking large-scale funding and has been focusing on only specific fields [25].

The next section will briefly introduce some of the fundamental technologies developed during the years.

2.1.2 Technologies

This section briefly introduces some of the essential Semantic Web technologies that are going to be relevant to the BookSampo data used.

Resource Description Framework (RDF)

The most fundamental of the Semantic Web technologies, the Resource Description Framework (RDF), was first defined in 1997 by the W3C [18, 55]. It was added to W3C recommendations and thus became a part of the Semantic Web Standards[7] later in 1999 [55]. The RDF model was developed as a metadata model to describe resources and their relations to other resources in a machine-readable way using statements [19, 40]. These statements consist

[6] `https://seco.cs.aalto.fi/projects/lodi4dh/`

[7] `https://www.w3.org/2001/sw/wiki/`

Findable		
F1	(Meta)data are assigned a globally unique and persistent identifier	
F2	Data are described with rich metadata (defined by R1 below)	
F3	Metadata clearly and explicitly include the identifier of the data they describe	
F4	(Meta)data are registered or indexed in a searchable resource	
Accessible		
A1	(Meta)data are retrievable by their identifier using a standardised communications protocol	
	A1.1	The protocol is open, free, and universally implementable
	A1.2	The protocol allows for an authentication and authorisation procedure, where necessary
A2	Metadata are accessible, even when the data are no longer available	
Interoperable		
I1	(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation	
I2	(Meta)data use vocabularies that follow FAIR principles	
I3	(Meta)data include qualified references to other (meta)data	
Reusable		
R1	(Meta)data are richly described with a plurality of accurate and relevant attributes	
	R1.1	(Meta)data are released with a clear and accessible data usage license
	R1.2	(Meta)data are associated with detailed provenance
	R1.3	(Meta)data meet domain-relevant community standards

Table 2.1: Fair Principles [62]

⋆	Data must be available on the web in some format
⋆ ⋆	Data must be machine-readable structured data
⋆ ⋆ ⋆	Data format must be non-proprietary
⋆ ⋆ ⋆ ⋆	Data must use open standards from W3C
⋆ ⋆ ⋆ ⋆ ⋆	Data must be linked to other people's data for context

Table 2.2: 5-star model requirements [8]

⋆ ⋆ ⋆ ⋆ ⋆ ⋆	Schemas used must be explicitly described and published with the data
⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆	Data quality must be explicated against the schemas used

Table 2.3: 7-star model requirements [35]

of triples—a subject, predicate and object—and refer to resources with Uniform Resource Identifiers (URIs) [40]. The RDF model can be represented using different serializations, such as using the Extensible Markup Language (XML), triples, quads, JSON or as graphs [19, 40]. An example RDF graph is shown in Figure 2.1.

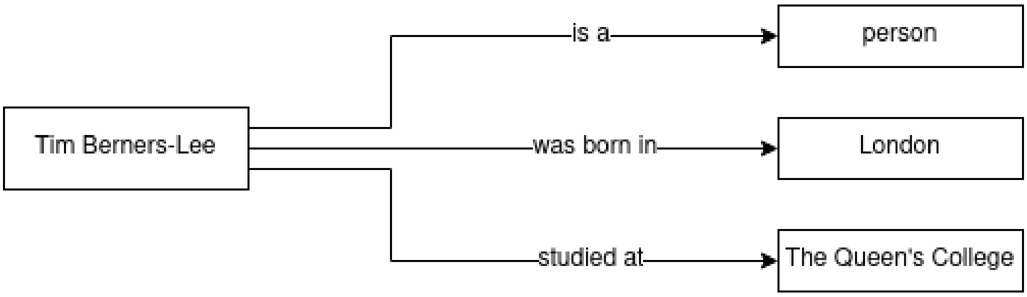


Figure 2.1: An example RDF graph representing a person and some of his basic properties

RDF/XML[8], an XML-based serialization, was envisioned to be a good initial serialization choice and was showcased in the RDF specification [40] in 1999 [19, 40]. The relevance of XML-based serialization options has however lessened over the years due to their weaknesses like scalability issues with bigger RDF data sets [19]. Many instead prefer to use other less verbose but still human-readable options like triple and quad serializations, such as Notation3[9] (N3), its RDF-focused subset Turtle[10] and Turtle's line-based subset N-Triples[11] [6, 7, 9, 19]. Extensions with quad support also exist for N-Triples and Turtle [6, 19]. Aside from triple and quad serializations, some JSON-based serializations like JSON-LD[12], which was specifically designed

[8] https://www.w3.org/TR/REC-rdf-syntax/

[9] https://www.w3.org/TeamSubmission/n3/

[10] https://www.w3.org/TR/turtle/

[11] https://www.w3.org/TR/n-triples/

[12] https://json-ld.org/

JSON compatible and with web-based programming environments in mind, have also gained some popularity over the years [19, 58].

RDF(S) and OWL ontology languages

To describe data with different terms and relations, there needs to exist some ontology with a structured vocabulary of definitions for the terms and relations used [18, 27]. These structured vocabularies can be written using ontology languages [18]. The implemented features of an ontology language differ from language to language [4]. Most commonly ontology languages implement classes and their instances of individual objects as well as properties and relations between objects [44].

Many different ontology languages were developed during the first few years of the Semantic Web's emergence [21]. While RDF Schema[13] (RDF(S)) and Web Ontology Language[14] (OWL)—and later Simple Knowledge Organization System (SKOS)—would top the usage charts in the coming years, many different ontology languages were thought to be promising in the first few years [15, 21, 57].

RDF(S) [13] is a semantic expansion of RDF that provides the necessary vocabulary for modeling RDF data with class and property hierarchies and constraints (restrictions) on subject and object values. While the class and property systems of RDF(S) are similar to those of object-oriented languages, properties are described in terms of their domain and range instead of being attributes of a certain class.

OWL [27] implements class and property hierarchies like RDF(S) but extends the features further. Properties can be assigned with relation types to indicate things like symmetry and transitivity and relations can also be defined in terms of another property like in the case of a property being the inverse of another property. OWL also facilitates the option to define restrictions on property behavior based on the local class instead the restrictions applying to the property universally.

SPARQL

SPARQL Protocol and RDF Query Language (SPARQL) [51] is a query language for RDF data. SPARQL was originally introduced in 2008 with its newer version SPARQL 1.1 [50] having its W3C Recommendation published in 2013. RDF data can be queried using SPARQL's four supported query forms `SELECT`, `CONSTRUCT`, `ASK` and `DESCRIBE`. SPARQL supports querying

[13] `https://www.w3.org/TR/rdf-schema/`

[14] `https://www.w3.org/OWL/`

with required and optional graph patterns as well as value testing and using constraints. Queried data can be combined with conjunctions and disjunctions.

For example, if one wanted to query for all objects of the class `foaf:Person` that have a specified name with the property `foaf:name` and then return the object URIs and the names, the SPARQL query could be formulated in the following format:

```
SELECT ?subject ?name WHERE {
    ?subject a foaf:Person ;
             foaf:name ?name .
}
```

where `?subject` is a variable for the object's URI and `?name` is a variable for the specified name of the object.

2.2 Literature and Linked Data

It did not take long for interest to be shown in the possible usage of Semantic Web technologies in the context of digital libraries. These technologies and what they could mean for digital libraries were discussed by Sure and Struder [60] in 2005 in their article *Semantic Web technologies for digital libraries*.

In 2009, the Library of Congress was taking its first steps into incorporating the Semantic Web technologies into digital libraries by providing digital records in a Linked Data service [52]. They would a few years later start the development of the XML-based Bibliographic Framework[15] (BIBFRAME) Linked Data model [52]. Around the world the leading role in publishing digital bibliographical data was largely taken up by national libraries [52]. Through publishing this digital bibliographical data, the goals were to improve service infrastructure and data as well as to better serve users and their needs with the long-term goal of better cooperation between libraries and other institutions recording relevant data [23].

In 2010 the Library Linked Data Incubator[16] W3C group [5] was formed. The aim of this group was to connect people from the library community together to figure out ways to better collaborate on adopting Linked Data in libraries. Concretely this meant finding ways to develop uniform standards for publishing linked bibliographical data as well as exploring the different

[15] `https://www.loc.gov/bibframe/`

[16] `https://www.w3.org/2005/Incubator/lld/`

schemas and metadata models for it. The working group's final report was published a year later in 2011.

While progress was made the same lack of appropriate tools that was plaguing the adoption of Linked Data also applied to bibliographical data [3]. On top of this there were also challenges specific to the domain [3]. Libraries had long used the MARC format for recording data and moving away from that would require a lot of resources [3]. The terminological differences between the terms of everyday use in libraries and the terms used in standards and models like the FRBR were a challenge [3, 5].

The possible benefits like facet-based search and increased discoverability however were apparent [3]. Initiatives like Europeana[17] and its library-domain aggregator the European Library[18], the British Library[19] and the Digital Public Library of America[20] (DPLA) started committing to the use of Linked Data despite the challenges and aimed to develop data models for bibliographical and other cultural heritage data publishing [3, 16, 24, 52]. Existing models like the previously mentioned FRBR model were improved and expanded [63]. FRBR and the other models in the FR family would later be combined into a singular model, the IFLA Library Reference Model (LRM), to meet the modeling needs of library data [54, 63].

In 2016 Hallo et al. [22] surveyed the state of Linked Data adoption in digital libraries. The survey set to find out what kind of vocabularies and ontologies were being used as well as the effects of the adoption of Linked Data into digital libraries and what the future could bring in that field. While benefits were noted, problems relating to things such as data quality and lack of appropriate support tools persisted withing the surveyed libraries and initiatives. More recent surveys indicate that technological challenges still pay a large factor in preventing the adoption of semantic web technology digital libraries even in 2022 [56].

[17] https://www.europeana.eu/

[18] https://www.theeuropeanlibrary.org/

[19] https://www.bl.uk/

[20] https://dp.la/

Chapter 3

Sampo Model & Sampo-UI

This chapter briefly goes over the Sampo Model[1] and the framework developed based on the model, the Sampo-UI framework[2], that were used to create the new BookSampo interface presented in this thesis. The first section focuses on the Sampo Model and its principles. The second section introduces the Sampo-UI framework itself.

3.1 Sampo Model

Sampo Model [32] is a general model to address problems regarding LOD data silos and data publishing that has been developed since 2002. The Sampo model consists of six principles showcased in Table 3.1. The model was based on the FAIR principles (Table 2.1) and was developed following LD principles, standards and W3C's best practices [32, 37].

The Sampo model was originally developed for the Cultural Heritage (CH) domain but can be and has been applied to other domains as well due to its generic nature [32]. The model has been applied in practice in a set of data services and semantic portals called the *Sampo series*.

3.2 Sampo-UI framework

Sampo-UI framework [38] is a framework developed by Semantic Research Group (SeCo) for facilitating the development of user interfaces for semantic portals in accordance with the Sampo model. The source code and docu-

[1] https://seco.cs.aalto.fi/applications/sampo/

[2] https://seco.cs.aalto.fi/tools/sampo-ui/

P1	Support collaborative data creation and publishing
P2	Use a shared open ontology infrastructure
P3	Make clear distinction between the LOD service and the user interface (UI)
P4	Provide multiple perspectives to the same data
P5	Standardize portal usage by a simple filter-analyze two-step cycle
P6	Support data analysis and knowledge discovery in addition to data exploration

Table 3.1: Sampo Model Principles [32]

mentation for Sampo-UI framework is available on GitHub[3] under the MIT license. The framework has been used for Sampo portal development since 2019. Semantic portals build based on the Sampo Model are called Sampo Portals[4].

The client of the framework is built with React and Redux libraries and the backend uses the Node.js and Express frameworks. The client of the framework is built using Material UI for the UI components. The data to be used in the portal is queried from knowledge graphs(s) using SPARQL endpoint(s). The client first requests data by sending read-only API requests to the backend and the SPARQL queries itself are generated and sent to the endpoint(s) by the backend.

Sampo-UI framework includes some ready-made UI components to be used in semantic portals. These ready-to-use components can be added to portals by simply editing the configuration files provided in the framework for the portal and the different perspectives. The facet menu has different options for facet types that can be used:

- *Text facet.* Facet providing free text search capability.
- *Hierarchical checkbox facet.* Facet with checkboxes for all different facet values that the user can select. Values can be grouped hierarchically if the data provides parent/child relations for concepts.
- *Date facet.* Facet for specifying a date range using date pickers.
- *Range facet.* Facet for specifying a value range using input fields.

[3] Sampo-UI on GitHub: `https://github.com/SemanticComputing/sampo-ui`

[4] List available at: `https://seco.cs.aalto.fi/applications/sampo/` (Accessed 1.11.2022)

- *Slider facet.* Facet for specifying a range using sliders.

On top of the facets, the Sampo-UI framework also offers ready-to-use visualization components that can be used to visualize the result datasets as well as for visualizing facet value distributions in some cases. The default option for visualizing data is a paginated table visualization, where each row represents a result set object and columns represent the objects' properties and their values.

The ready-to-use visualization components included in the framework are the following:

- *Leaflet map.* One of the two available map visualization components. The component is built using the Leaflet[5] library. The component supports showing external map layers.

- *DeckGL map.* The other map-based visualization component. The component is built with the DeckGL[6] framework which uses WebGL technology. This component can be used to create, for example, temporal maps.

- *Pie/bar chart.* This component is built using the ApexCharts[7] library. The format can be changed between pie and bar chart formats using a dropdown menu included in the component.

- *Line chart.* Another component built using the ApexCharts library that can be used for, e.g., time series visualizations.

- *Network component.* This component can be used for visualizing networks. The component is built using Cytoscape.js[8].

Before diving into the design of the new implemented BookSampo portal, let us look at the original one. The next chapter introduces the original BookSampo portal, one of the early Sampo portals, and the BookSampo data set.

[5] `https://leafletjs.com/`

[6] `https://deck.gl/`

[7] `https://apexcharts.com/`

[8] `https://js.cytoscape.org/`

Chapter 4

BookSampo & Data

This chapter presents the origins of the current publicly available BookSampo portal[1] as well as over the BookSampo data and its evolution throughout the years.

4.1 Origins

The current BookSampo service maintained by Finnish Public Libraries was originally part of the FinnONTO project introduced in Chapter 2 and developed following the Sampo Model principles as also previously used in CultureSampo [30]. Following the Sampo Model principles meant using shared domain ontologies from the infrastructure created as a part of the FinnONTO project and this guided the development of the metadata model itself. The content from existing databases was mostly automatically transformed into Linked Data format using ontologies based on the thesauri used for fiction literature indexing [41]. This data was then manually checked and corrected by volunteer librarians using Semantic Web editing tools and ONKI ontology services.

The original data is from a data dump of the Helsinki metropolitan area library system from June 2009 and was based on content keyword data on literary works that have been systematically added for literary works in Finland since 1997 [41, 43]. These content keywords are entered using the Finnish fiction content thesaurus Kaunokki that has been developed since 1993 [41]. The BookSampo project itself was started in 2008 as part of the FinnONTO project as a joint venture between the Finnish public libraries and Semantic Computing Research Group (SeCo) located at Aalto University and University of Helsinki [41].

[1] `https://www.kirjasampo.fi/`

The KOKO ontology cloud alone did not contain all the concepts needed for representing the themes and genres used for fiction literature indexing and needed to be supplemented [42]. This supplemental ontology used in the BookSampo data is the bilingual ontology KAUNO that was created by converting the Kaunokki and Bella (the Swedish counterpart of Kaunokki) thesauri first into RDF and then by mapping them to each other [41]. KAUNO was then linked to the KOKO ontology cloud automatically and the links were manually checked [42]. The ontologies used were further supplemented with additional language ontologies Lingvoj[2] and Lexvo[3] as well as geographical ontologies and a nationality ontology [28].

The data model for the literary works in BookSampo is a simplified data model based on the FRBRoo Model [43]. Instead of FRBRoo Model's four-leveled structure of (1) work, (2) expression, (3) manifestation and (4) item, works are represented on only two levels: abstract and physical work levels [43, 53]. The abstract work level of a work describes the details of the work itself that remain unchanged between different editions and is the equivalent to FRBRoo Model's work level [43, 53]. The physical work level of the work on the other hand describes edition-specific information on literary works like page numbers and publishers and is equivalent to FRBRoo Model's manifestation level [43, 53]. An example of this split is shown in Figure 4.1 using Mika Waltari's work *Sinuhe egyptiläinen*.

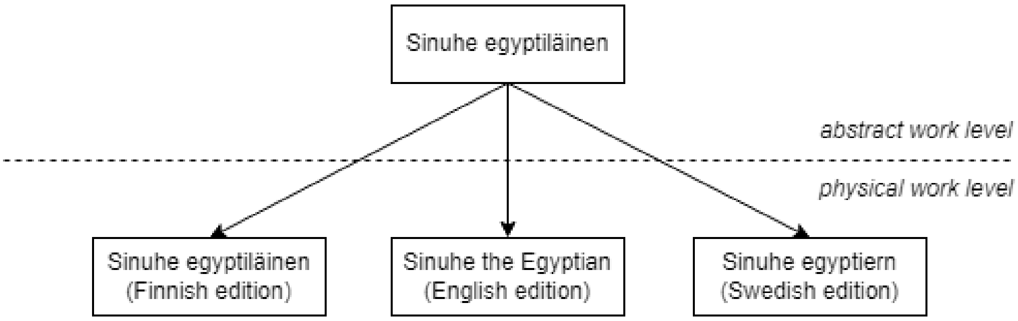

Sinuhe egyptiläinen
abstract work level
physical work level
Sinuhe egyptiläinen (Finnish edition)
Sinuhe the Egyptian (English edition)
Sinuhe egyptiern (Swedish edition)

Figure 4.1: Split between abstract and physical work levels

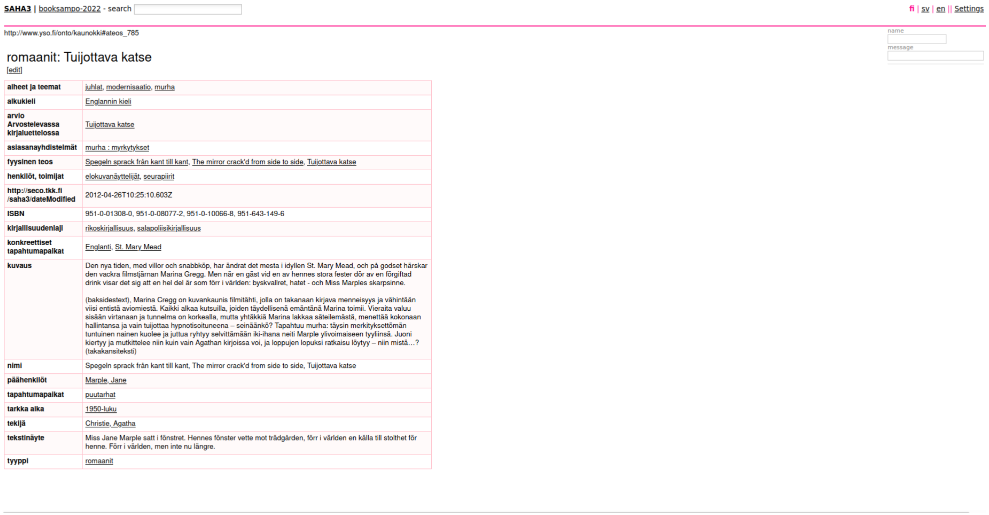
Collections of works (e.g., short story collections) are modeled using these two levels by first defining that a singular story is both an abstract work as well as a physical work that is part of another physical work [43]. This latter physical work is the physical-level manifestation of the collection itself and is linked to the abstract level work representing the collection.

[2] `http://linkedvocabs.org/lingvoj/`

[3] `http://www.lexvo.org/`

The original data used for BookSampo data was recorded in MARC format, which is edition-centric in nature [41]. The transformation from this edition-centric data to abstract works was done automatically and then manually checked by volunteers to get rid of things like duplicates due to different editions being transformed into multiple different abstract works.

In addition to data on literary works themselves, authors and other people related to the literary works had to be modeled as well. Based on user research, the biographical information of authors was recorded in attribute form instead of using events for things like the birth and death of an author [41]. Listing these things as just attribute values was found to be simpler for the people inputting all the data and thus a better solution than the classic CIDOC-CRM and BIO way of using events.



http://www.yso.fi/onto/kaunokki#ateos_785

romaanit: Tuijottava katse

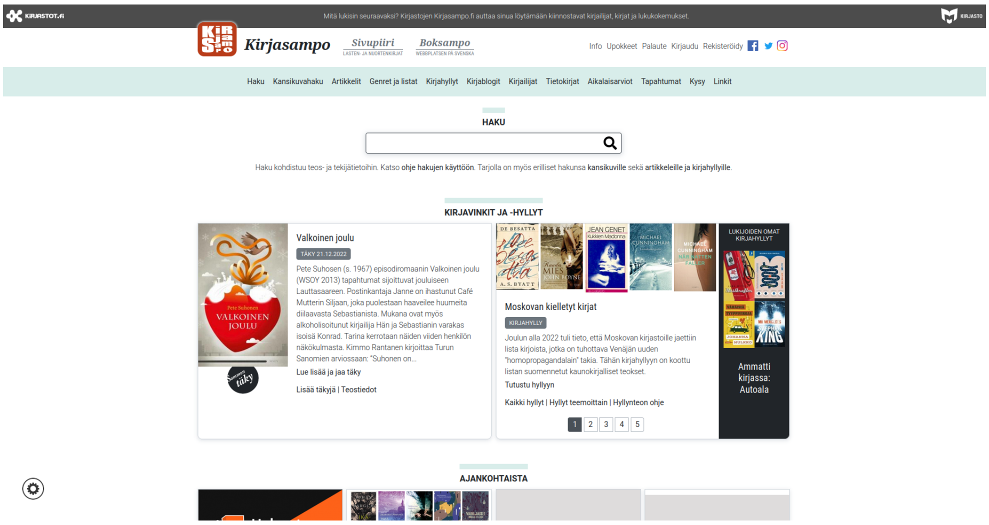
aiheet ja teemat	juhlat, modernisaatio, murha
alkukieli	Englannin kieli
arvio Arvostelevassa kirjaluettelossa	Tuijottava katse
asiasanayhdistelmät	murha : myrkytykset
fyysinen teos	Spegeln sprack från kant till kant, The mirror crack'd from side to side, Tuijottava katse
henkilöt, toimijat	elokuvanäyttelijät, seurapiirit
http://seco.tkk.fi /saha3/dateModified	2012-04-26T10:25:10.603Z
ISBN	951-0-01308-0, 951-0-08077-2, 951-0-10066-8, 951-643-149-6
kirjallisuudenlaji	rikoskirjallisuus, salapoliisikirjallisuus
konkreettiset tapahtumapaikat	Englanti, St. Mary Mead
kuvaus	Den nya tiden, med villor och snabbköp, har ändrat det mesta i idyllen St. Mary Mead, och på godset härskar den vackra filmstjärnan Marina Gregg. Men när en gäst vid en av hennes stora fester dör av en förgiftad drink visar det sig att en hel del är som förr i världen: byskvallret, hatet - och Miss Marples skarpsinne. (baksidestext), Marina Cregg on kuvankaunis filmitähti, jolla on takanaan kirjava menneisyys ja vähintään viisi entistä aviomiestä. Kaikki alkaa kutsuilla, joiden täydellisenä emäntänä Marina toimii. Vieraita valuu sisään virtanaan ja tunnelma on korkealla, mutta yhtäkkiä Marina lakkaa säteilemästä, menettää kokonaan hallintansa ja vain tuijottaa hypnotisoituneena – seinäänkö? Tapahtuu murha: täysin merkityksettömän tuntuinen nainen kuolee ja juttua ryhtyy selvittämään iki-ihana neiti Marple ylivoimaiseen tyyliinsä. Juoni kiertyy ja mutkittelee niin kuin vain Agathan kirjoissa voi, ja loppujen lopuksi ratkaisu löytyy – niin mistä…? (takakansiteksti)
nimi	Spegeln sprack från kant till kant, The mirror crack'd from side to side, Tuijottava katse
päähenkilöt	Marple, Jane
tapahtumapaikat	puutarhat
tarkka aika	1950-luku
tekijä	Christie, Agatha
tekstinäyte	Miss Jane Marple satt i fönstret. Hennes fönster vette mot trädgården, förr i världen en källa till stolthet för henne. Förr i världen, men inte nu längre.
tyyppi	romaanit

Figure 4.2: SAHA metadata editor user interface

RDF-based metadata editor SAHA[4] (shown in Figure 4.2) that was developed during FinnONTO was adopted to be used for the editing environment for the data [42]. The BookSampo portal itself was developed using the Drupal portal system [42]. The user interface for the portal is shown in Figure 4.3.

The end-user portal offers search functionality to the user through a text-based search that utilizes the linked nature of the data to map the final search result by combining sub-result sets matching the provided keywords [41]. The user can for example search with a combination of an author's name and a

[4] https://seco.cs.aalto.fi/services/saha/



Mitä lukisin seuraavaksi? Kirjastojen Kirjasampo.fi auttaa sinua löytämään kiinnostavat kirjailijat, kirjat ja lukukokemukset.

Haku kohdistuu teos- ja tekijätietoihin. Katso **ohje hakujen käyttöön**. Tarjolla on myös erilliset hakunsa **kansikuville** sekä **artikkeleille ja kirjahyllyille**.

Valkoinen joulu

TÄKY 21.12.2022

Pete Suhosen (s. 1967) episodiromaanin Valkoinen joulu (WSOY 2013) tapahtumat sijoittuvat jouluiseen Lauttasaareen. Postinkantaja Janne on ihastunut Café Mutterin Siljaan, joka puolestaan haaveilee huumeita diilaavasta Sebastianista. Mukana ovat myös alkoholisoitunut kirjailija Hän ja Sebastianin varakas isoisä Konrad. Tarina kerrotaan näiden viiden henkilön näkökulmasta. Kimmo Rantanen kirjoittaa Turun Sanomien arviossaan: "Suhonen on...

Lue lisää ja jaa täky

Lisää täkyjä | Teostiedot

Moskovan kielletyt kirjat

KIRJAHYLLY

Joulun alla 2022 tuli tieto, että Moskovan kirjastoille jaettiin lista kirjoista, jotka on tuhottava Venäjän uuden "homopropagandalain" takia. Tähän kirjahyllyyn on koottu listan suomennetut kaunokirjalliset teokset.

Tutustu hyllyyn

Kaikki hyllyt | Hyllyt teemoittain | Hyllynteon ohje

Figure 4.3: The original BookSampo portal user interface

keyword that would match a cover of a book written by the specified author. The sub-results of the search would lead from the cover all the way to the book written by the specified author itself and this book would be the final returned result.

4.2 Data

This section first goes over the past and current state of the BookSampo data itself and afterwards the most important data models are introduced. Lastly this section covers some of the difficulties arising from the nature of the data.

The BookSampo data has grown over the years. In 2013 over three million triples with 400,000 subject URIs were stored in the data [43]. In 2022, the data has more than 8.7 million triples and over one million distinct subject URIs. The instance count values in comparison to the values of 2013 are showcased in Table 4.1. Back in spring 2013 the data on literary works was primarily focused on fiction for an adult demographic, but the scope was expanded afterwards to include metadata also on works for children and teens [1].

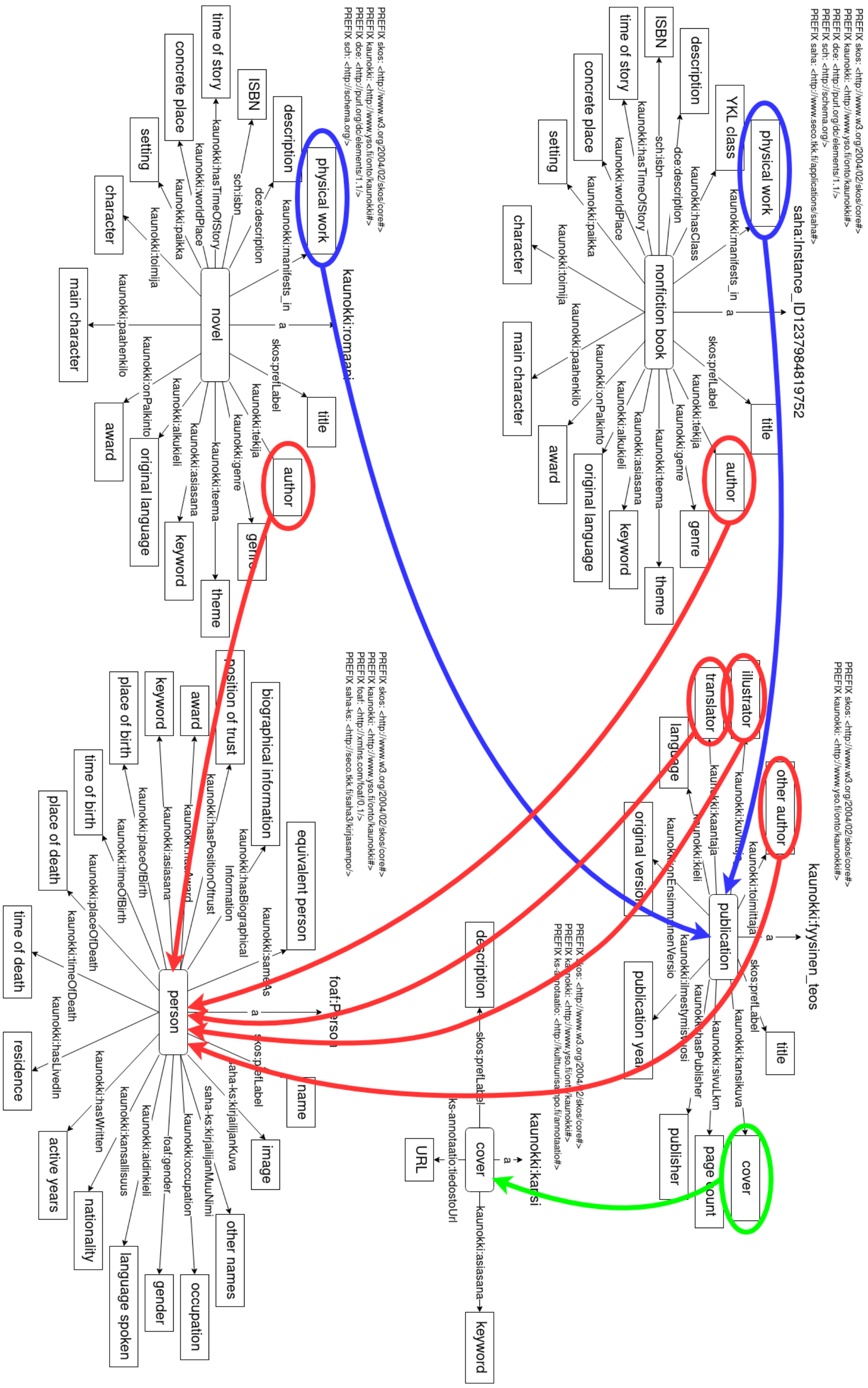


Figure 4.4: Combined data model of the most important classes

Literary works	93,000	Literary works	209,000
Publications	127,000	Publications	214,000
Covers	27,000	Covers	113,000
People	29,000	People	62,000
Reviews	15,000	Reviews	15,000
Series	2,900	Series	8,300

Table 4.1: Instance counts in 2013 [43] (left) vs. 2022 (right)

4.2.1 Data models

This subsection introduces some of the most relevant classes and their most relevant properties in relation to the development of the new BookSampo portal. For an overview of all the classes mentioned in this section and how they are related to each other through their properties, see Figure 4.4. The colored lines indicate the relations between the classes. The individual data model charts for the covered classes are included later in this section.

The most relevant classes in the data in regard to the development of the new BookSampo portal were the following:

1. the classes for books itself
   - `kaunokki`[5]`:romaani` (novel) and `saha`[6]`:Instance_ID1237984819752` (nonfiction book) classes for the abstract work levels
   - `kaunokki:fyysinen_teos` (publication, physical work) class for the physical work level
2. `kaunokki:kansi` class for book covers
3. `foaf`[7]`:Person` class for all authors and other people related to literature

The data model for the novel class and its most important properties are displayed in Figure 4.5. This class encompasses all novels and their information on an abstract level. The physical level of the work is connected to the novel class through the `kaunokki:manifests_in` property that takes the physical work object URI as its value. The novels are connected to person objects representing authors using the `kaunokki:tekija` property.

The data model for the nonfiction book class and its most important properties is remarkably similar to the novel class one and is showcased in

[5] Prefix `kaunokki` is used for `http://www.yso.fi/onto/kaunokki#` namespace

[6] Prefix `saha` is used for `http://www.seco.tkk.fi/applications/saha#` namespace

[7] Prefix `foaf` is used for `http://xmlns.com/foaf/0.1/` namespace

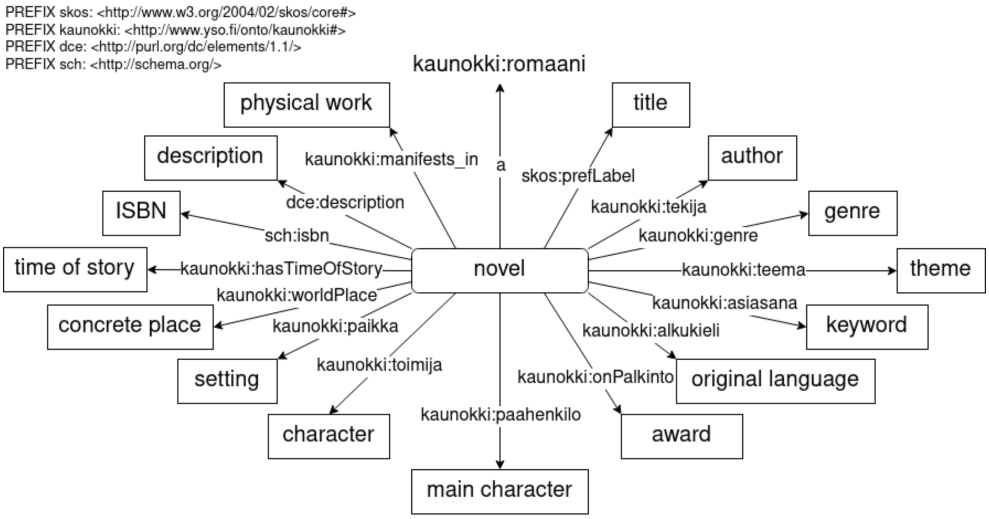


Figure 4.5: Novel class and its most important properties

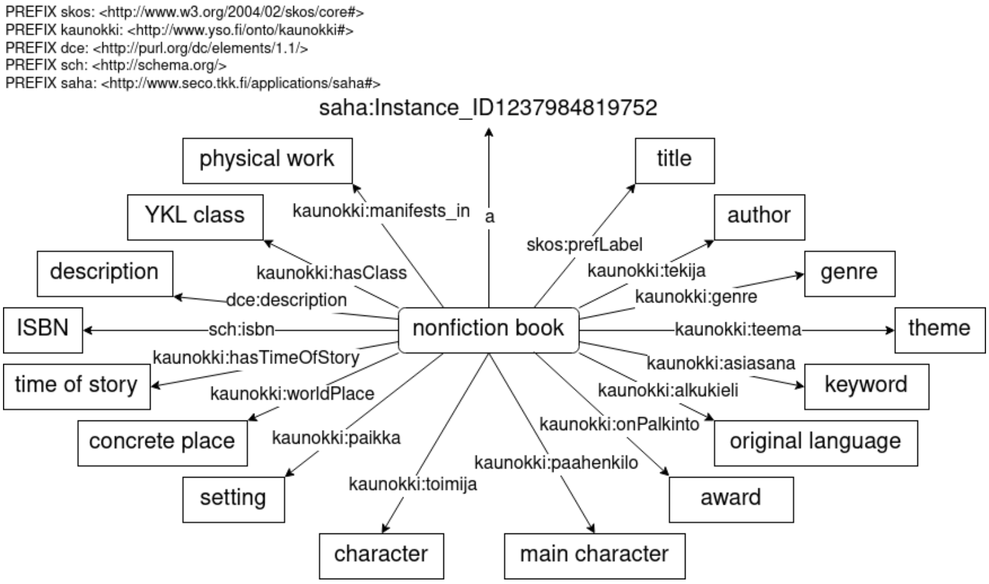


Figure 4.6: Nonfiction book class and its most important properties

Figure 4.6. In addition to the properties shared with the novel class, non-fiction books also have a property for the YKL (Yleisten kirjastojen luokitusjärjestelmä) class[8] that represents the category in which the book would be stored in a library catalogue.

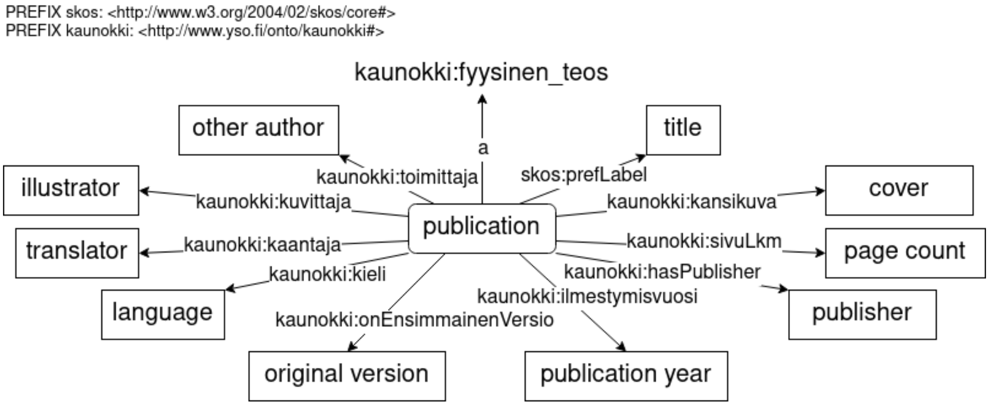


Figure 4.7: Publication class and its most important properties

The data model for the publication class and its most important properties are showcased in Figure 4.7. The publication class deals with the publication specific information like a publication's publication year and page count. The original publication is denoted with `kaunokki:onEnsimmainenVersio` which has the value `kaunokki:true` if the publication is the first publication in the original language. The publication class is connected to the book cover class through its property `kaunokki:kansikuva` which takes cover object URIs as its value. People related to a particular publication specifically, e.g., illustrators and translators, are linked to the publication using `kaunokki:kuvittaja`, `kaunokki:kaantaja` and `kaunokki:toimittaja` properties.

The data model for the book cover class and its most important properties are showcased in Figure 4.8. Covers are tagged with keywords such as color keywords and keywords for the beings depicted in the covers with the `kaunokki:asiasana` keyword property. The cover image itself is stored as a URL value of the property `ks-annotaatio:tiedostoUrl`.

The data model for the person class and its most important properties are showcased in Figure 4.9. The person class includes not only authors but other people related to literature, such as translators, reviewers and illustrators.

[8] Information about YKL classification available at: `https://www.kiwi.fi/x/bIcdCw` (Accessed 24.1.2023)

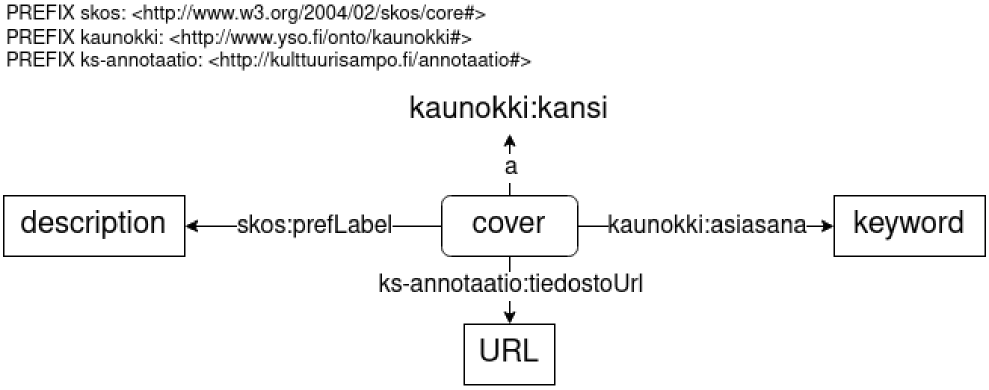

PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX kaunokki: <http://www.yso.fi/onto/kaunokki#>
PREFIX ks-annotaatio: <http://kulttuurisampo.fi/annotaatio#>
kaunokki:kansi
a
description
skos:prefLabel
cover
kaunokki:asiasana
keyword
ks-annotaatio:tiedostoUrl
URL

Figure 4.8: Cover class and its most important properties

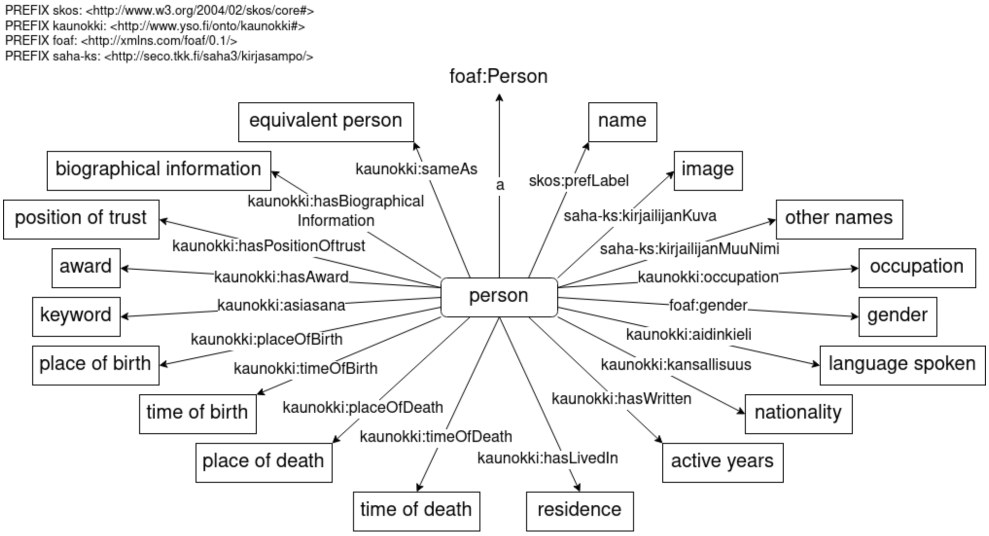


Figure 4.9: Person class and its most important properties

4.2.2 Data problems

This subsection goes over some of the problems of the data that arose during the development of the new BookSampo portal.

Label coverage

Novels	**80,718**
with skos:prefLabel	80,698
with a language tag	79,982
with FI language tag	48,405
with SV language tag	45,878
with EN language tag	23,314
with other language tag	11,012
without any language tag	716
without skos:prefLabel	20
Nonfiction books	**1,956**
with skos:prefLabel	1,928
with a language tag	1,819
with FI language tag	1,554
with SV language tag	452
with EN language tag	177
with other language tag	171
without any language tag	109
without skos:prefLabel	28
Publications	**213,877**
with skos:prefLabel	213,829
with a language tag	203,766
with FI language tag	90,677
with SV language tag	69,682
with EN language tag	27,913
with other language tag	25,945
without any language tag	10,063
without skos:prefLabel	48

Table 4.2: Label coverage for novels, nonfiction books and publications (spring 2022)

The biggest problem in terms of the development of the new portal was the preferred label coverage. Table 4.2 lists the preferred label counts by language tag for novel, nonfiction book and publication names, which gives

Original languages	**118**
with skos:prefLabel	7
with FI language tag	7
without skos:prefLabel	111
Main characters	**45,598**
with skos:prefLabel	45,595
with FI language tag	1,098
multiple skos:prefLabels with @fi	14
without skos:prefLabel	3
Awards	**6,310**
with skos:prefLabel	6,276
with FI language tag	3,952
multiple skos:prefLabels with @fi	1
without skos:prefLabel	34
Genres	**630**
with skos:prefLabel	628
with FI language tag	627
multiple skos:prefLabels with @fi	4
without skos:prefLabel	2
Publishers	**3,084**
with skos:prefLabel	3,084
with FI language tag	51
multiple skos:prefLabels with @fi	2
without skos:prefLabel	0

Table 4.3: Finnish label coverage for some of the novel properties' values in use in the data (spring and autumn 2022)

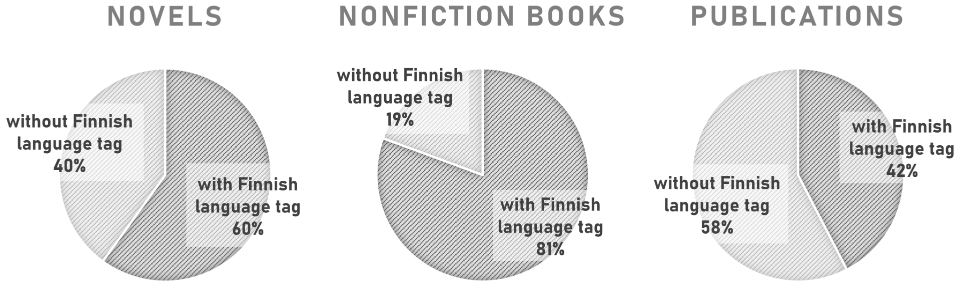


Figure 4.10: Finnish language tagged preferred label coverage for novel, nonfiction book and publication instances that have some kind of preferred label

some insight into label coverage percentages. Out of these three, novels and publications have nearly 100% coverage for preferred labels and nonfiction books nearly 99% coverage. The coverage however drops when looking at labels with Finnish language tags. Figure 4.10 illustrates the proportion of instances with a Finnish language tagged preferred label to those without for the three classes.

To preserve uniformity withing the new portal, querying for labels would ideally be done with a language tag filter so all returned labels would be in for example Finnish. The coverage in Finnish is however not 100% for all preferred labels as showcased in the aforementioned table as well as Table 4.3 showcasing the Finnish preferred label coverage for some example novel properties. In cases of high coverage, the missing labels could be replaced with a label from another language or one without a language tag, but with cases of low coverage the uniformity of the portal would suffer.

In addition to missing language tags, the usage of `skos:prefLabel` was not always guaranteed. For language resources the language names were provided mostly with `rdfs:label` while some still provided a single language-tagged label with `skos:prefLabel`. These language resources using `rdfs:label` had multiple labels with the same language tag which would have to be reduced to one preferred one for the portal if they were to be used. This problem of duplicate labels also applies to some of the `skos:prefLabel` values in the data as well, where there are incorrectly multiple preferred labels listed for things like characters and publishers. For example, the character *Scrooge McDuck* from the cartoon *Donald Duck* has two different preferred labels with the Finnish language tag in the data: *Roope-setä* ('Uncle Scrooge') and *Roope Ankka* ('Scrooge McDuck'). Having multiple preferred labels for the same language tag in this way goes against the SKOS integrity constraints for preferred labels, so the duplicate labels should be listed under a different label type [45].

Missing data

In some individual cases related objects had missing links to each other. This would result in orphan physical works with no information regarding their abstract level work and lone cover images with no information on the physical work they appeared in.

Some of the resources used in annotating things, such as years or other time periods, were also problematic as they were possibly created just for that annotation. These resources would often lack information regarding the start and end time of the time period and hierarchy information (some statistics for time resources are shown in Table 4.4), which makes time-based

Time resources	**15,954**
with skos:prefLabel	15,935
with FI language tag	689
without skos:prefLabel	19
missing earliest start time	14,396
missing latest end time	14,397
without parent time resource (hierarchy)	15,505

Table 4.4: Time resource data property coverage (spring 2022)

visualizations tricky to implement with high coverage for the sampled objects.

Annotation choices also differ greatly as annotation work is done by humans. One annotator might feel something is worth annotating and have their own annotation style, while another annotator might not find the same things worth annotating or use different kind of annotations. This can skew the visualization results towards certain kind of works that typically get annotated more thoroughly in certain aspects, e.g., action books might have more precise world place annotations than a romance novel, where the setting might not play a large role.

There were also some systematic problems with certain kinds of resources lacking class information which would be useful for filtering for all of these kinds of resources. Places in the data, for example, do not have any kind of class signifying that they are places. Place objects can thus only be found through filtering for objects that appear as property values to properties that specifically refer to places.

Format issues

All the time resources in the data have their end and start times stored as strings without any proper enforcement on the format itself (e.g., both `1971-01-01T00:00:00.0Z` and `1.1.1971` are acceptable formats in the data). Since there is no proper enforcement on the string format, automatic converting of the data type to types like `dateTime` would be problematic.

Properties like page counts are also stored in string format instead of as integers. This leads to the values containing extra information in string format (e.g., edition number before the page count) and non-integer values that cannot properly be parsed for filtering and visualization purposes.

Missing hierarchy

As mentioned previously and briefly showcased in Table 4.4, a lot of the time resources in the data are lacking hierarchy and thus cannot be used for creating hierarchical facets. Implementing decade-based visualizations also suffers from the lack of hierarchy and requires playing around with the string values, which in turn negatively affects querying speed.

The same missing hierarchy problem also arises when looking at place resources. The place resources are from GeoNames[9], but are missing the hierarchy information present in there. This becomes problematic when trying to visualize data based on the places listed. For places in Finland the annotators might likely list the city itself as the setting for a book or as a place of birth, but for foreign countries the country itself might be listed instead. This leads to situations where it seems like some country is the most written about country based on the visualizations, while in truth the instance counts would be completely different if the annotations for cities inside all countries were added to the total.

Historical context

While not an issue with the data per se, historical context is important especially for annotated places. Borders change and what might have constituted a part of a country in some year but this might not apply in the next decade, and some countries might cease to exist completely. This raises the question of how these places should be shown on a map in the map-based visualizations: should the setting locations for now non-existent countries like the USSR be shown on the map and if so, where and based on what borders?

Miscellaneous issues

Possibly due to the nature of the SAHA metadata editor's interface and how the text search for possible property values works, there are some false links in the data. This is apparent in the case of place links, where multiple places share the same or similar name. For example, some books annotated with Malmi, Finland, are incorrectly marked to take place in Malmi, South Korea.

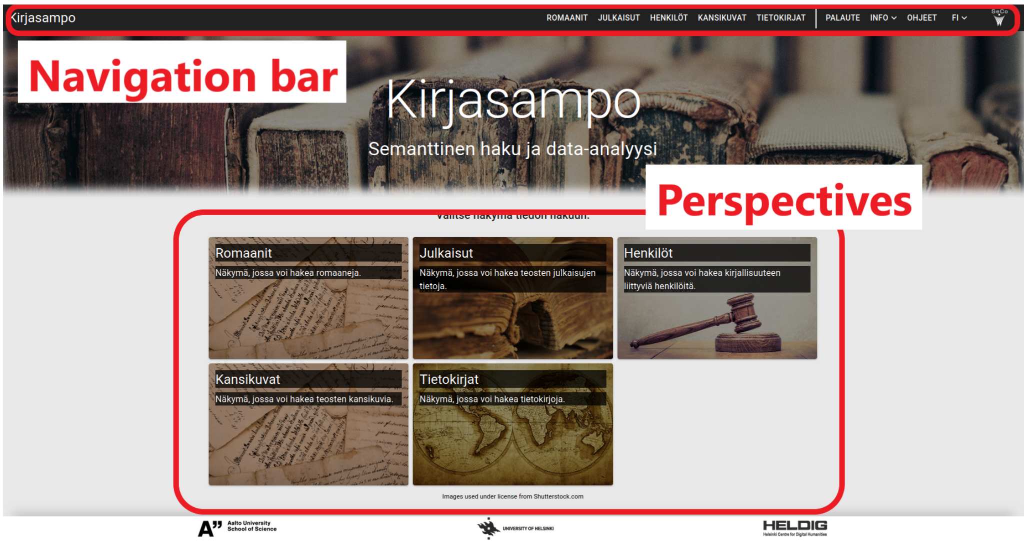
[9] `https://www.geonames.org/`

Chapter 5

Design

Now that the data used has been introduced, this and the following chapter will cover the design and the implementation of the portal itself. This chapter overviews the design of the implemented BookSampo portal. The first subsection presents how the perspectives are split and what kind of interfaces and functionality they have. The second subsection covers what kind of visualizations are included in the portal.

5.1 Perspectives



Valitse näkymä tiedon hakuun.

Näkymä, jossa voi hakea romaaneja.

Näkymä, jossa voi hakea teosten julkaisujen tietoja.

Näkymä, jossa voi hakea kirjallisuuteen liittyviä henkilöitä.

Näkymä, jossa voi hakea teosten kansikuvia.

Näkymä, jossa voi hakea tietokirjoja.

Images used under license from Shutterstock.com

Figure 5.1: Portal home page

The design of the new BookSampo UI follows the Sampo Model principle of using same data from multiple perspectives without modifications in the data itself. Due to the sheer size of the data, including all the different classes as perspectives would have been a monumental task. This lead to narrowing down and prioritizing the included perspectives to cover different aspects of literature as broadly as possible with just select few perspectives.

Novels	80,718
Nonfiction books	1,956
Publications	213,877
Covers	112,971
People	62,207
Reviews	14,644
Series	8,374

Table 5.1: Exact instance counts for perspective classes in 2022

Covering at least literary works on both abstract and physical levels and people related to these works was essential. Another aspect chosen was the covers of literary works due to the popularity of the cover search function of the original BookSampo portal [2]. In the end the new BookSampo portal's perspectives were decided to be the following:

1. *Novels.* This perspective covers all the novels on an abstract level. Edition-specific information is covered in *Publications* perspective. This perspective was chosen to cover the abstract level fiction works aspect of literature due having the most instances out of all the fiction work classes. Exact instance counts for this and the other perspective classes are shown in Table 5.1.

2. *Nonfiction books.* This perspective covers all the nonfiction books on an abstract level. Similarly to *Novels*, edition-specific information is covered in *Publications* perspective. This perspective was chosen to complement the abstract level work aspect with nonfiction works instead of just covering fiction.

3. *Publications.* This perspective covers different publications published of all Finnish literature works. In practice all first editions published in a relevant language, i.e. Finnish, Swedish and the original language of a translated work, are added as publications.

4. *Covers.* This perspective covers all the book covers created for different publications.

5. *People.* This perspective covers all the people directly related to the literature presented, e.g., authors and translators, but also people like illustrators and reviewers.

The split between novels and publications follows the split made in the BookSampo data introduced in Chapter 4, that is to say that the novels perspective deals with the abstract works themselves while publications contains the physical editions. While some edition-specific data are also shown in the novels perspective, all visualizations dealing with that data are presented in the publications perspective. This choice was made to maintain the idea that the objects being visualized should be the same as the objects found in the table search result view.

The above-mentioned perspectives are referred here as 'full' due to their configuration regarding the possible views they can have. Full perspectives are configured to be able to be viewed in two different kind of views:

1. *Faceted search.* Faceted search includes a table view of paginated results as well as different visualizations defined for the said perspective included in additional tabs. The results in both the table view and the visualizations can be affected by filtering the results using the facets provided in the perspective. This is the view that the user is directed to when they choose a perspective from the landing page.

2. *Instance pages.* Instance pages cover all the data about an entity. These pages contain both all the data shown about the particular entity in the table view in addition to possible additional data that exists. This additional data is usually something that couldn't meaningfully be used for filtering with facets, e.g., a property value containing free-form string instead of pointing to another entity. Including it in a column would end up bloating the table too much.

In addition to the full perspectives specified above, there are also perspectives that have no faceted search view but rather consist only of instance pages. This way these entities' individual information can still be viewed even if the data examined in the perspective does not warrant having a full perspective view. These perspectives with only instance pages specified are the following:

1. *Places.* This perspective covers all the places relating to works and people. The exact instance counts for place instances are not known and thus missing from the exact instance counts in Table 5.1 due to the lack of class for place instances in the BookSampo data.

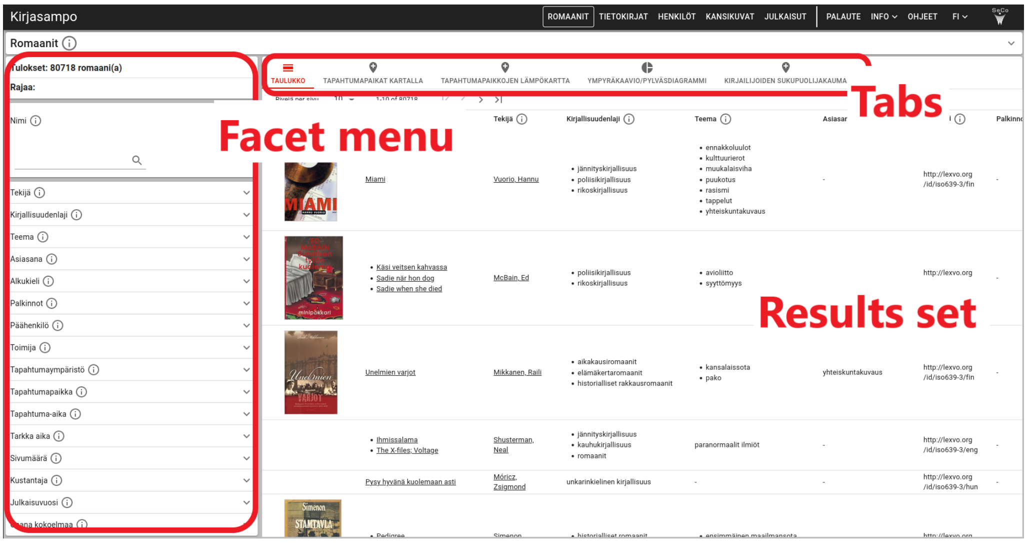
2. *Reviews.* This perspective covers all the contemporary reviews of works.

3. *Series.* This perspective covers all the series of works.

These perspectives can only be reached by clicking on hyperlinks of entities belonging to these perspectives from any of the main perspectives, and are not shown on the landing page with the initial view of all perspectives.

5.1.1 Faceted search view

The first view the user is presented with when (s)he selects a perspective is a faceted search view where the results are shown in paginated table format. By default no facets are applied and the results show all the data objects that match the facet class specified for the perspective, i.e., novels perspective faceted search view shows all novels denoted by the `rdf:type` of `kaunokki:romaani` in the data by default.



	Tekijä	Kirjallisuudenlaji	Teema	Asiasar	i	Palkinnc
Miami	Vuorio, Hannu	• jännityskirjallisuus • poliisikirjallisuus • rikoskirjallisuus	• ennakkoluulot • kulttuurierot • muukalaisviha • puukotus • rasismi • tappelut • yhteiskuntakuvaus	-	http://lexvo.org /id/iso639-3/fin	-
• Käsi veitsen kahvassa • Sadie när hon dog • Sadie when she died	McBain, Ed	• poliisikirjallisuus • rikoskirjallisuus	• avioliitto • syyttömyys		http://lexvo.org	-
Unelmien varjot	Mikkanen, Raili	• aikakausiromaanit • elämäkertaromaanit • historialliset rakkausromaanit	• kansalaissota • pako	yhteiskuntakuvaus	http://lexvo.org /id/iso639-3/fin	-
• Ihmissalama • The X-files; Voltage	Shusterman, Neal	• jännityskirjallisuus • kauhukirjallisuus • romaanit	paranormaalit ilmiöt	-	http://lexvo.org /id/iso639-3/eng	-
Pysy hyvänä kuolemaan asti	Móricz, Zsigmond	unkarinkielinen kirjallisuus	-	-	http://lexvo.org /id/iso639-3/hun	-
• Pedigree	Simenon	• historialliset romaanit	• ensimmäinen maailmansota		http://lexvo.org	

Figure 5.2: Table view of results set

The faceted search view composition is shown in Figure 5.2. The view can be divided into three major sections: (1) tabs for visualizing the results (2) facet menu and (3) results set.

The top section of the view has an expandable container used for displaying information relating to the current perspective. By default it is in unexpanded state, and has the perspective name visible. If the container is expanded further information about the perspective is shown in text format.

The facet menu is in the left side of the screen. The menu contains all the facets that can be used for filtering the data. Individual facets consist of an expandable container whose content depends on the type of the facet. The facet types used in the BookSampo portal are the following:

1. *String input.* This facet is used for string-based search facets e.g. searching based on novel names. The entered value is applied after the user presses enter.

2. *Checkbox.* This facet includes checkboxes for all the facet values that can be found from the data. A search field is included at the top for searching for specific values from withing the facet values. The selected checkbox values are applied immediately.

3. *Integer range.* This facet is used for specifying a range of integers in which values of an attribute should fall in, e.g., for specifying a page number range. This facet must be applied by using the apply button after at least one value is entered for the range.

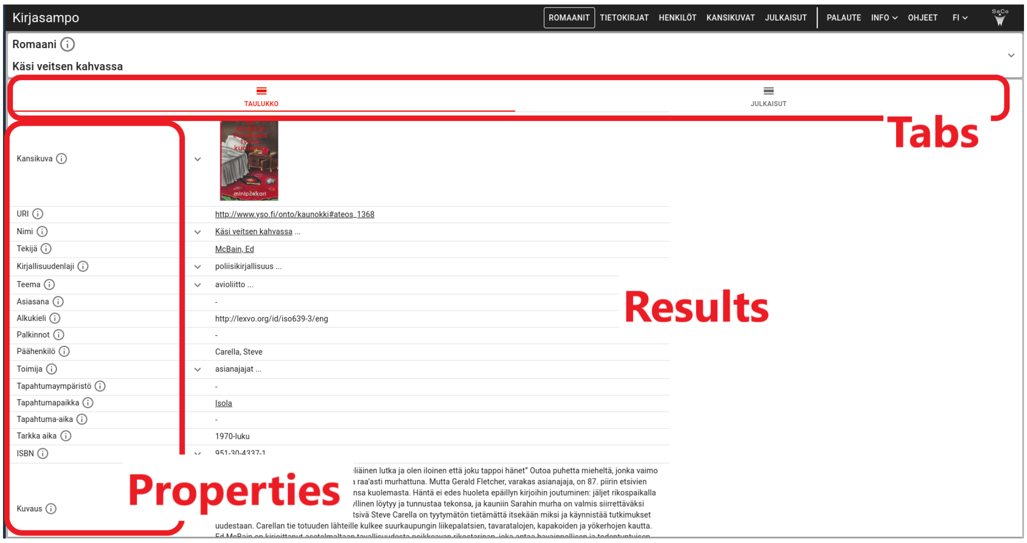
The top of the facet menu updates with the number of results after facets are applied. As facets are automatically applied after a value is chosen or applied inside a particular facet, no universal apply button is included in the facet menu.

The right side of the view contains the search results themselves. This container is split into different tabs. The first tab is the paginated table of the results. By default, this is the tab that is included in all perspectives. This table view of search results shows the results split into rows and columns. Each row represents an object of the searched facet class. Columns hold values of different properties relating to the object. Columns should exist for at least all of the properties listed in the facet menu but do not necessarily include all the possible properties to prevent bloating the view too much. Extra properties not included in the column view may be included on the instance pages of the objects. Other tabs in the view are optional and used for different visualizations of the perspective data. Visualizations are discussed in more detail in Section 5.2.

In order to view more detailed information regarding an entity, the user has to see its instance page. This can be done by clicking on an object's preferred label that are indicated by the label being underlined. Instance pages and their layout will be covered in the next section.

5.1.2 Instance pages

Instance pages are pages for displaying the full information regarding an entity. By default they have at least one tab containing information in a table format with the attribute being displayed on the left and its value on the right. The other tabs can be used for visualizations or other custom tabs like tabs for displaying publications in case of novels and nonfiction books.



Romaani

Käsi veitsen kahvassa

Kansikuva	
URI	http://www.yso.fi/onto/kaunokki#ateos_1368
Nimi	Käsi veitsen kahvassa ...
Tekijä	McBain, Ed
Kirjallisuudenlaji	poliisikirjallisuus ...
Teema	avioliitto ...
Asiasana	-
Alkukieli	http://lexvo.org/id/iso639-3/eng
Palkinnot	-
Päähenkilö	Carella, Steve
Toimija	asianajajat ...
Tapahtumaympäristö	-
Tapahtumapaikka	Isola
Tapahtuma-aika	-
Tarkka aika	1970-luku
ISBN	951-30-4337-1
Kuvaus	iiäinen lutka ja olen iloinen että joku tappoi hänet" Outoa puhetta mieheltä, jonka vaimo a raa'asti murhattuna. Mutta Gerald Fletcher, varakas asianajaja, on 87. piirin etsivien nsa kuolemasta. Häntä ei edes huoleta epäillyn kirjoihin joutuminen: jäljet rikospaikalla yllinen löytyy ja tunnustaa tekonsa, ja kauniin Sarahin murha on valmis siirrettäväksi tsivä Steve Carella on tyytymätön tietämättä itsekään miksi ja käynnistää tutkimukset uudestaan. Carellan tie totuuden lähteille kulkee suurkaupungin liikepalatsien, tavaratalojen, kapakoiden ja yökerhojen kautta. Ed McBain on kirjoittanut asetelmaltaan tavallisuudesta poikkeavan rikostarinan, joka antaa havainnollisen ja todentuntuisen

Figure 5.3: Instance page of an object

The rows containing multiline information can be expanded by clicking them or be enabled to be expanded by default in cases where it makes sense. Property values can also include links to other pages, e.g. the instance page for a novel would have a link to the author's page that can be reached by clicking the author's name.

For an overview on how all these views in the portal are navigated through, see Figure 5.4. The figure illustrates a quick example usage cycle for the system in which a user wants to find a book fulfilling certain criteria. The steps illustrated are:

1. Selecting the appropriate perspective. In this example case the user is looking for a novel, so the novels perspective is selected.

2. Select the wanted facets. In the example case the user sets the genre facet to *romance novels*, setting to *castles* and characters to *nobility*.

3. The selected facets are automatically applied, and the results set to the right is updated. The books on the right now all fulfill the user's basic criteria.

4. The user can look at a book in more detail by clicking one the name of a book from the faceted search results table and opening the book's instance page.

The next chapter covers all the different perspectives and the choices made regarding data in them in more detail.

5.2 Visualizations

Klink et al. [39] present some different ways of visualizing bibliographical data of scientific publications in both textual and non-textual formats in their article *Browsing and Visualizing Digital Bibliographic Data*. The article presents the usage of timelines and histograms for visualizing publications and networks for visualizing relationships between co-authors.

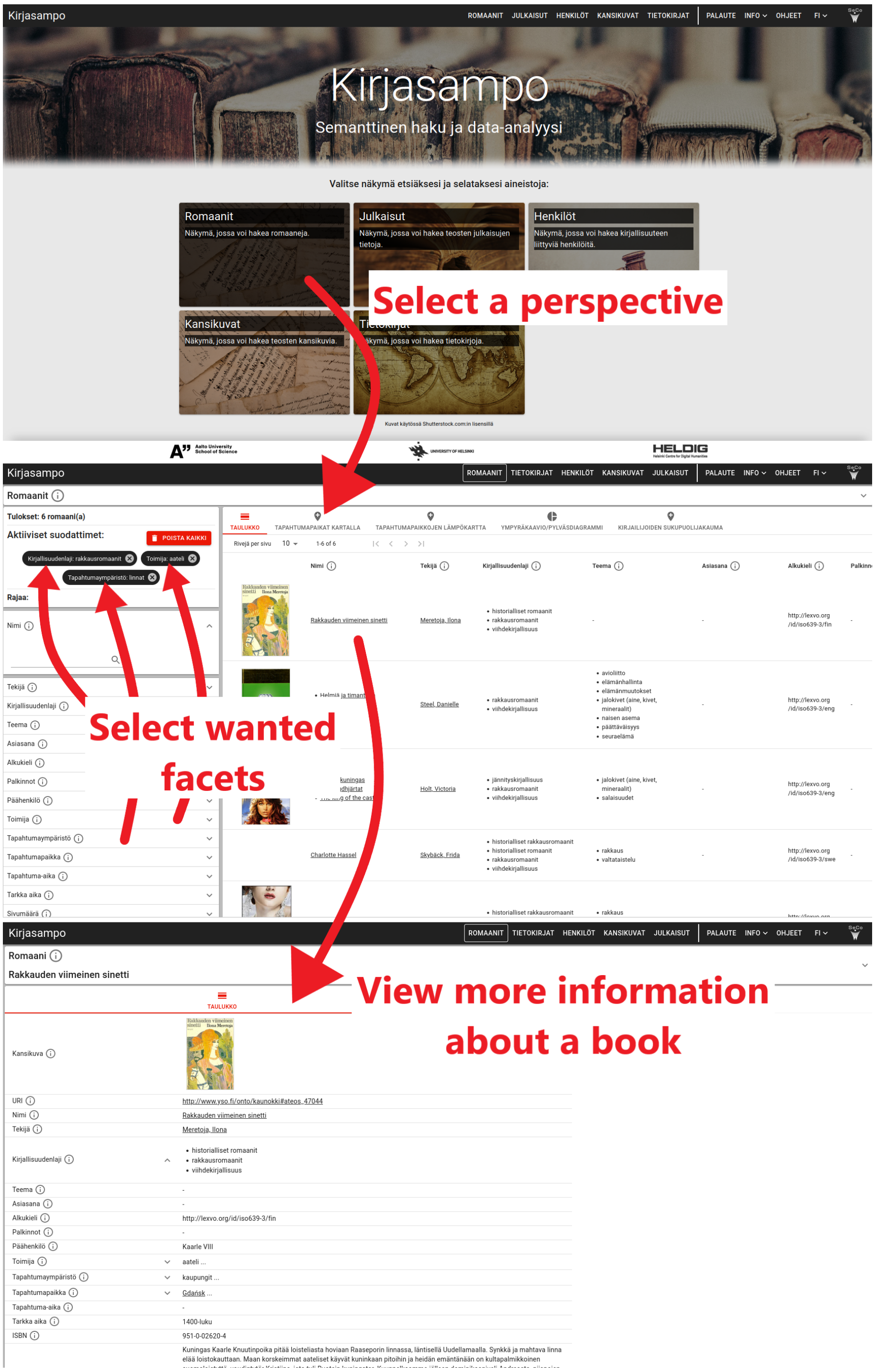
Börner et al. [12] similarly discuss visualizing a dataset with data from Science Citation Index (SCI) and Social Science Citation Index (SSCI). The article presents similar visualizations of showing publication counts through time and different networks based on co-author relationships as well as publication topics. The article also presents a visualization based on publication topic popularity.

Time series visualizations can easily be applied to all the publications in the BookSampo data, but visualizing co-author networks would not necessarily work well for the authors present in the data because books tend to be written and credited to one author outside of listing translators and illustrators. As all data related to publication years are related to publications themselves, these visualizations would best be created for the publications perspective.

The visualization showcased in the article of Börner et al. [12] could be adapted to use the ready-to-use pie chart component provided in Sampo-UI framework and be used for showcasing a variety of topics related to both publications as well as abstract works: genres, themes, characters and languages.

In contrast to the scientific publications covered in the aforementioned articles, BookSampo data set also offers location information for book settings as well as for events like birth and death in an author's life. This presents an opportunity for a variety of map-based visualizations showcasing both static information like a setting in a book as well as dynamic information like migrations of authors during their life.

Another source for possible visualization ideas comes from the HS Open event hosted by Helsingin Sanomat. BookSampo data set was used in the


Kirjasampo
ROMAANIT
JULKAISUT
HENKILÖT
KANSIKUVAT
TIETOKIRJAT
PALAUTE
INFO
OHJEET
FI
Kirjasampo
Semanttinen haku ja data-analyysi
Valitse näkymä etsiäksesi ja selataksesi aineistoja:
Romaanit
Näkymä, jossa voi hakea romaaneja.
Julkaisut
Näkymä, jossa voi hakea teosten julkaisujen tietoja.
Henkilöt
Näkymä, jossa voi hakea kirjallisuuteen liittyviä henkilöitä.
Kansikuvat
Näkymä, jossa voi hakea teosten kansikuvia.
Tietokirjat
Näkymä, jossa voi hakea tietokirjoja.
Kuvat käytössä Shutterstock.com:in lisenssillä
Select a perspective
Romaanit
Tulokset: 6 romaani(a)
Aktiiviset suodattimet:
POISTA KAIKKI
Kirjallisuudenlaji: rakkausromaanit
Toimija: aateli
Tapahtumaympäristö: linnat
Rajaa:
Nimi
Tekijä
Kirjallisuudenlaji
Teema
Asiasana
Alkukieli
Palkinnot
Päähenkilö
Toimija
Tapahtumaympäristö
Tapahtumapaikka
Tapahtuma-aika
Tarkka aika
Sivumäärä
TAULUKKO
TAPAHTUMAPAIKAT KARTALLA
TAPAHTUMAPAIKKOJEN LÄMPÖKARTTA
YMPYRÄKAAVIO/PYLVÄSDIAGRAMMI
KIRJAILIJOIDEN SUKUPUOLIJAKAUMA
Rivejä per sivu 10
1-6 of 6
Select wanted facets
View more information about a book

Nimi	Tekijä	Kirjallisuudenlaji	Teema	Asiasana	Alkukieli	Palkinn
Rakkauden viimeinen sinetti	Meretoja, Ilona	historialliset romaanit rakkausromaanit viihdekirjallisuus	-	-	http://lexvo.org/id/iso639-3/fin	-
Helmiä ja timant	Steel, Danielle	rakkausromaanit viihdekirjallisuus	avioliitto elämänhallinta elämänmuutokset jalokivet (aine, kivet, mineraalit) naisen asema päättäväisyys seuraelämä	-	http://lexvo.org/id/iso639-3/eng	-
kuningas dhjärtat ... of the cast	Holt, Victoria	jännityskirjallisuus rakkausromaanit viihdekirjallisuus	jalokivet (aine, kivet, mineraalit) salaisuudet	-	http://lexvo.org/id/iso639-3/eng	-
Charlotte Hassel	Skybäck, Frida	historialliset rakkausromaanit historialliset romaanit rakkausromaanit viihdekirjallisuus	rakkaus valtataistelu	-	http://lexvo.org/id/iso639-3/swe	-
		historialliset rakkausromaanit	rakkaus			

Romaani

Rakkauden viimeinen sinetti

TAULUKKO

Kansikuva	
URI	http://www.yso.fi/onto/kaunokki#ateos_47044
Nimi	Rakkauden viimeinen sinetti
Tekijä	Meretoja, Ilona
Kirjallisuudenlaji	historialliset romaanit rakkausromaanit viihdekirjallisuus
Teema	-
Asiasana	-
Alkukieli	http://lexvo.org/id/iso639-3/fin
Palkinnot	-
Päähenkilö	Kaarle VIII
Toimija	aateli ...
Tapahtumaympäristö	kaupungit ...
Tapahtumapaikka	Gdańsk ...
Tapahtuma-aika	-
Tarkka aika	1400-luku
ISBN	951-0-02620-4
	Kuningas Kaarle Knuutinpoika pitää loisteliasta hoviaan Raaseporin linnassa, läntisellä Uudellamaalla. Synkkä ja mahtava linna elää loistokauttaan. Maan korskeimmat aateliset käyvät kuninkaan pitoihin ja heidän emäntänään on kultapalmikkoinen

Figure 5.4: Chart showing a basic usage cycle of the interface

2011 event to explore Finnish literature and its evolution. Three articles by Mäkinen regarding the findings were published in Helsingin Sanomat: *Näin keräät kirjaasi kaikki kliseet* [46], *Yhden kirjan kirjoittajien määrä kasvaa* [48] and *Romaanit pidentyneet* [47]. Out of these three *Näin keräät kirjaasi kaikki kliseet* [46] and *Romaanit pidentyneet* [47] are about questions that could be easily visualized.

Näin keräät kirjaasi kaikki kliseet [46] describes the most common traits or "cliches" of the literature included in the BookSampo data set back in 2011. This includes things like most common themes, settings and times of story and what your generic author is like. As mentioned previously, these most common traits can be easily visualized with Sampo-UI's pie chart components.

The article *Romaanit pidentyneet* [47] on the other hand is about the average page length of books and how the length has evolved throughout the years. The average page length could be visualized using a similar time series component as the ones presented in the articles of Klink et al. [39] and Börner et al. for scientific publication counts.

Based on these articles and the BookSampo data set, at least three different types of visualizations are clearly of interest:

1. *Pie charts* showcasing the most common values of properties

2. *Time series* showcasing the evolution of publication counts or of some of their properties

3. *Map visualizations* showcasing settings or places of birth and death

All of these components from the aforementioned categories can be easily implemented with Sampo-UI's ready-to-use components or by developing custom ones with the libraries already included in the framework depending on the individual needs for specific visualizations.

The next chapter covers the actual implementation of the new BookSampo portal and its visualizations.

Chapter 6

Implementation

The new BookSampo portal was built using the Sampo-UI framework as a base. The data is queried from a Jena Apache Fuseki[1] SPARQL server and queried results are cached using Varnish Cache[2]. The data stored in the Jena Apache Fuseki server is a BookSampo data dump from spring 2022 given by Finnish Public Libraries.

6.1 Portal configuration

Sampo-UI's basic configuration is mostly done in three directories. These directories have folders with the name `sampo` that should be changed to match the ID of the portal. This can be done by either renaming the original `sampo` folders or duplicating these folders and then renaming them. The relevant directories and their purposes are the following:

1. `src/client/translations/`. This is the directory where different translations for things like menu items and labels are stored for different locales.

2. `src/configs/`. This is the directory where the JSON configuration files for the portal and all the perspectives are stored.

3. `src/server/sparql/`. This is the directory where all the SPARQL query files are stored. These file names are referred to in the previously mentioned JSON configuration files to connect the queries to the perspective.

[1] `https://jena.apache.org/documentation/fuseki2/`

[2] `https://varnish-cache.org/`

In addition to these directories, setting up a new semantic portal also requires the `sampo` folder in `src/client/components/perspectives/` to be either renamed or duplicated and then renamed to match the new portal ID. Afterwards this directory can be ignored if changes to the main layout and footer are not needed. Some basic instructions for setting up a new semantic portal using Sampo-UI can also be found from the Sampo-UI GitHub Wiki[3].

Portal specific settings are configured through Sampo-UI's portal configuration JSON file. This section will cover the most important configuration options from that file in detail. The relevant sections from the file discussed in this chapter are highlighted in Figure 6.1. For a full example portal configuration file, see Appendix A.

The most important attribute in this file is the `perspectives` attribute. Listing 6.1 shows this attribute's value in the BookSampo portal configuration file.

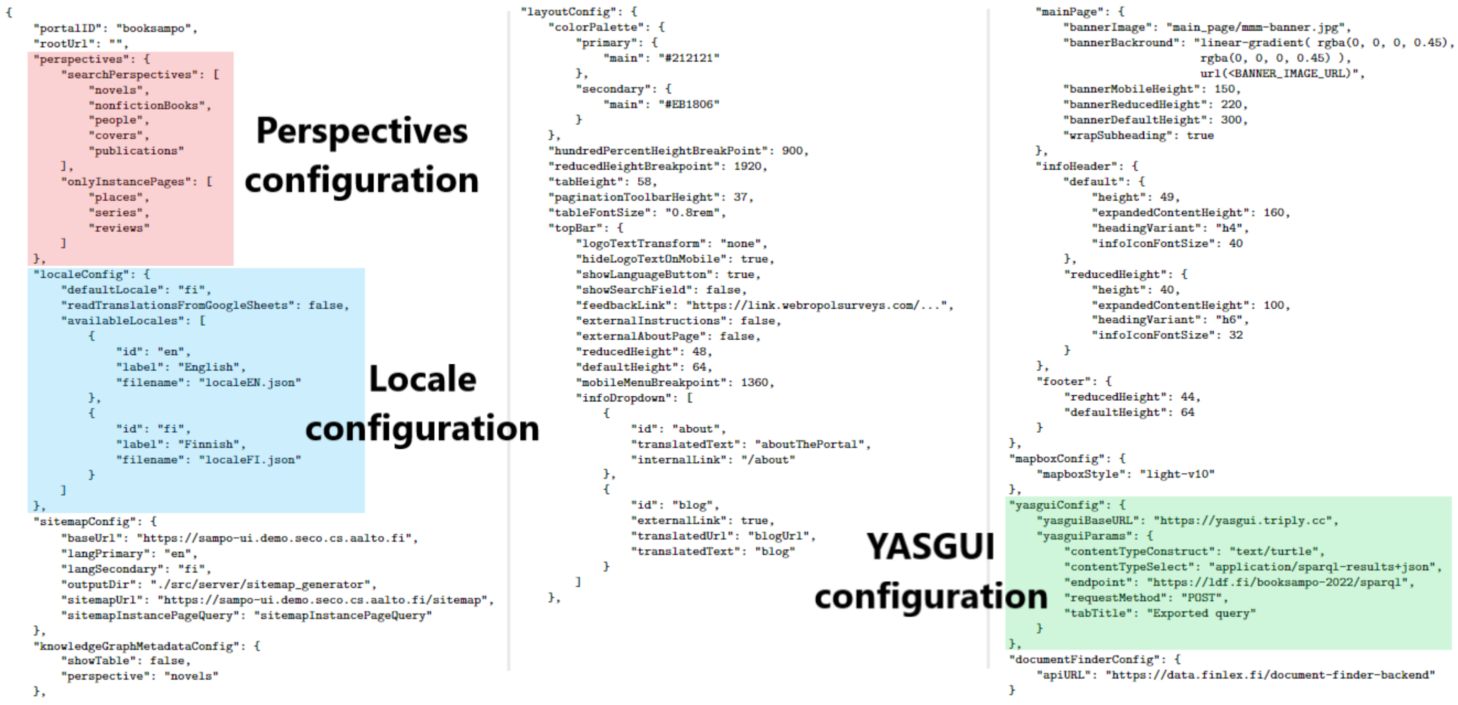
```
"perspectives": {
  "searchPerspectives": [
    "novels",
    "nonfictionBooks",
    "people",
    "covers",
    "publications"
  ],
  "onlyInstancePages": [
    "places",
    "series",
    "reviews"
  ]
}
```

Listing 6.1: `perspectives` attribute in the BookSampo portal configuration

This attribute has two sub-attributes within it for listing perspectives based on their nature, `searchPerspectives` and `onlyInstancePages`. This `perspectives` attribute is used for listing the perspectives to be included in the created portal by their ID that will be matched to the perspective-specific file of the perspective. The perspective ID is added to either of the sub-attribute lists based on whether the included perspective is meant to be a full search perspective (`searchPerspectives`) or a perspective with just instance pages and no search view (`onlyInstancePages`). As in the design, *novels*, *nonfiction books*, *people*, *covers*, and *publications* are configured as full search perspectives with *places*, *series*, and *reviews* perspectives being only available as instance pages that can be reached through links from other perspectives.

```
"localeConfig": {
  "defaultLocale": "fi",
```

[3] `https://github.com/SemanticComputing/sampo-ui/wiki`



```
{
    "portalID": "booksampo",
    "rootUrl": "",
    "perspectives": {
        "searchPerspectives": [
            "novels",
            "nonfictionBooks",
            "people",
            "covers",
            "publications"
        ],
        "onlyInstancePages": [
            "places",
            "series",
            "reviews"
        ]
    },
    "localeConfig": {
        "defaultLocale": "fi",
        "readTranslationsFromGoogleSheets": false,
        "availableLocales": [
            {
                "id": "en",
                "label": "English",
                "filename": "localeEN.json"
            },
            {
                "id": "fi",
                "label": "Finnish",
                "filename": "localeFI.json"
            }
        ]
    },
    "sitemapConfig": {
        "baseUrl": "https://sampo-ui.demo.seco.cs.aalto.fi",
        "langPrimary": "en",
        "langSecondary": "fi",
        "outputDir": "./src/server/sitemap_generator",
        "sitemapUrl": "https://sampo-ui.demo.seco.cs.aalto.fi/sitemap",
        "sitemapInstancePageQuery": "sitemapInstancePageQuery"
    },
    "knowledgeGraphMetadataConfig": {
        "showTable": false,
        "perspective": "novels"
    },
    "layoutConfig": {
        "colorPalette": {
            "primary": {
                "main": "#212121"
            },
            "secondary": {
                "main": "#EB1806"
            }
        },
        "hundredPercentHeightBreakPoint": 900,
        "reducedHeightBreakpoint": 1920,
        "tabHeight": 58,
        "paginationToolbarHeight": 37,
        "tableFontSize": "0.8rem",
        "topBar": {
            "logoTextTransform": "none",
            "hideLogoTextOnMobile": true,
            "showLanguageButton": true,
            "showSearchField": false,
            "feedbackLink": "https://link.webropolsurveys.com/...",
            "externalInstructions": false,
            "externalAboutPage": false,
            "reducedHeight": 48,
            "defaultHeight": 64,
            "mobileMenuBreakpoint": 1360,
            "infoDropdown": [
                {
                    "id": "about",
                    "translatedText": "aboutThePortal",
                    "internalLink": "/about"
                },
                {
                    "id": "blog",
                    "externalLink": true,
                    "translatedUrl": "blogUrl",
                    "translatedText": "blog"
                }
            ]
        },
        "mainPage": {
            "bannerImage": "main_page/mmm-banner.jpg",
            "bannerBackround": "linear-gradient( rgba(0, 0, 0, 0.45),
                                rgba(0, 0, 0, 0.45) ),
                                url(<BANNER_IMAGE_URL)",
            "bannerMobileHeight": 150,
            "bannerReducedHeight": 220,
            "bannerDefaultHeight": 300,
            "wrapSubheading": true
        },
        "infoHeader": {
            "default": {
                "height": 49,
                "expandedContentHeight": 160,
                "headingVariant": "h4",
                "infoIconFontSize": 40
            },
            "reducedHeight": {
                "height": 40,
                "expandedContentHeight": 100,
                "headingVariant": "h6",
                "infoIconFontSize": 32
            }
        },
        "footer": {
            "reducedHeight": 44,
            "defaultHeight": 64
        }
    },
    "mapboxConfig": {
        "mapboxStyle": "light-v10"
    },
    "yasguiConfig": {
        "yasguiBaseURL": "https://yasgui.triply.cc",
        "yasguiParams": {
            "contentTypeConstruct": "text/turtle",
            "contentTypeSelect": "application/sparql-results+json",
            "endpoint": "https://ldf.fi/booksampo-2022/sparql",
            "requestMethod": "POST",
            "tabTitle": "Exported query"
        }
    },
    "documentFinderConfig": {
        "apiURL": "https://data.finlex.fi/document-finder-backend"
    }
```

Figure 6.1: Portal configuration file with highlighted explanations for sections

```
  "readTranslationsFromGoogleSheets": false,
  "availableLocales": [
    {
      "id": "en",
      "label": "English",
      "filename": "localeEN.json"
    },
    {
      "id": "fi",
      "label": "Finnish",
      "filename": "localeFI.json"
    }
  ]
}
```

Listing 6.2: `localeConfig` attribute in the BookSampo portal configuration

The `localeConfig` attribute (see Listing 6.2) is used for configuring the locale settings for the whole portal. The new BookSampo user interface supports two languages (`availableLocales`), Finnish and English, with data labels being queried primarily in Finnish due to largest label coverage in that language as found in Chapter 4. Finnish is set as the default language (`defaultLocale`) for the portal.

```
"yasguiConfig": {
  "yasguiBaseURL": "https://yasgui.triply.cc",
  "yasguiParams": {
    "contentTypeConstruct": "text/turtle",
    "contentTypeSelect": "application/sparql-results+json",
    "endpoint": "https://ldf.fi/booksampo-2022/sparql",
    "requestMethod": "POST",
    "tabTitle": "Exported query"
  }
}
```

Listing 6.3: `yasguiConfig` attribute in the BookSampo portal configuration

Sampo-UI framework supports adding tabs for exporting queries into the YASGUI[4] SPARQL Graphical User Interface (GUI). The endpoint used for these queries in YASGUI as well as other relevant YASGUI settings are configured in the portal configuration file through the `yasguiConfig` attribute (see Listing 6.3). In the case of the new BookSampo portal, the endpoint is configured to the aforementioned Jena Apache Fuseki server's endpoint although the export tabs itself are not currently used in the portal.

The other attributes in the portal configuration file deal with general site-wide settings and graphics. This includes things such as the links in the site header and general layout configurations.

All different perspectives are solely configured through their own perspective JSON configuration files. These configuration files are used to specify the endpoint that is used for querying the data for that specific perspective

[4] `https://yasgui.triply.cc/`

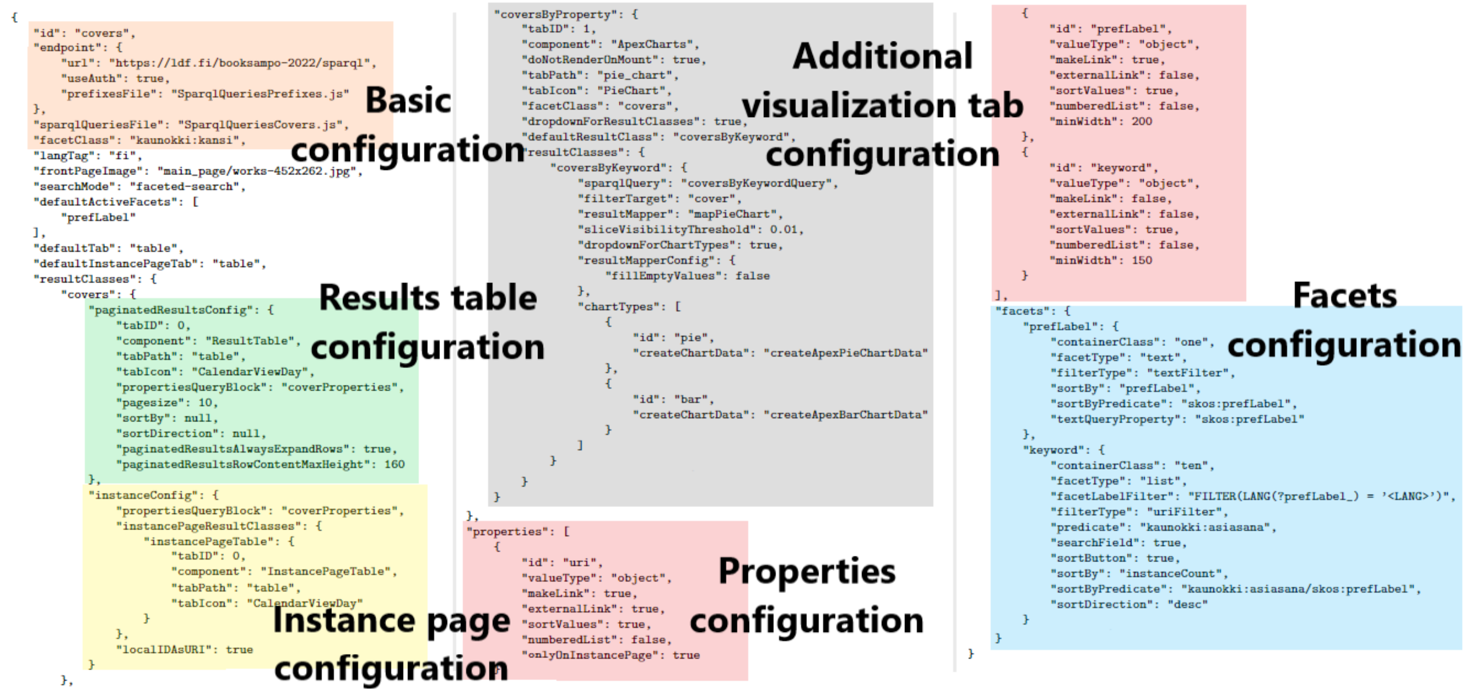
as well as the JavaScript file containing the SPARQL queries themselves. Figure 6.2 shows an abridged version of an example perspective file with relevant sections from the file highlighted and labeled. A slightly longer example perspective configuration file with more properties and facets can be found in Appendix B.

There are three attributes for specifying the content of the perspective that is to be shown to the user: `resultClasses`, `properties` and `facets`.

```
"resultClasses": {
  "covers": {
    "paginatedResultsConfig": {
      "tabID": 0,
      "component": "ResultTable",
      "tabPath": "table",
      "tabIcon": "CalendarViewDay",
      "propertiesQueryBlock": "coverProperties",
      "pagesize": 10,
      "sortBy": null,
      "sortDirection": null,
      "paginatedResultsAlwaysExpandRows": true,
      "paginatedResultsRowContentMaxHeight": 160
    },
    "instanceConfig": {
      "propertiesQueryBlock": "coverProperties",
      "instancePageResultClasses": {
        "instancePageTable": {
          "tabID": 0,
          "component": "InstancePageTable",
          "tabPath": "table",
          "tabIcon": "CalendarViewDay"
        }
      },
      "localIDAsURI": true
    }
  },
  ...
}
```

Listing 6.4: Abridged `resultClasses` attribute in the BookSampo covers perspective configuration

The `resultClasses` attribute (see Listing 6.4) contains all the components shown on the different tabs of the faceted search view. The most basic `resultClasses` attribute contains at least a single attribute with the perspective ID as its name (e.g., `covers` in Listing 6.4) and an object as its value. This object includes the configuration for the initial results view (e.g., the default table view) as well as the configuration for the instance pages of this perspective. Things like the included component, tab path as well as icon are specified in these configurations in addition to possible component-specific configuration options. The specific query to be used for the results is listed here using the name of the variable containing the query in the queries file. Other tabs are included by adding additional attributes to the `resultClasses` attribute. These should contain the same kind of tab and



```
{
    "id": "covers",
    "endpoint": {
        "url": "https://ldf.fi/booksampo-2022/sparql",
        "useAuth": true,
        "prefixesFile": "SparqlQueriesPrefixes.js"
    },
    "sparqlQueriesFile": "SparqlQueriesCovers.js",
    "facetClass": "kaunokki:kansi",
    "langTag": "fi",
    "frontPageImage": "main_page/works-452x262.jpg",
    "searchMode": "faceted-search",
    "defaultActiveFacets": [
        "prefLabel"
    ],
    "defaultTab": "table",
    "defaultInstancePageTab": "table",
    "resultClasses": {
        "covers": {
            "paginatedResultsConfig": {
                "tabID": 0,
                "component": "ResultTable",
                "tabPath": "table",
                "tabIcon": "CalendarViewDay",
                "propertiesQueryBlock": "coverProperties",
                "pagesize": 10,
                "sortBy": null,
                "sortDirection": null,
                "paginatedResultsAlwaysExpandRows": true,
                "paginatedResultsRowContentMaxHeight": 160
            },
            "instanceConfig": {
                "propertiesQueryBlock": "coverProperties",
                "instancePageResultClasses": {
                    "instancePageTable": {
                        "tabID": 0,
                        "component": "InstancePageTable",
                        "tabPath": "table",
                        "tabIcon": "CalendarViewDay"
                    }
                },
                "localIDAsURI": true
            }
        },
```

```
        "coversByProperty": {
            "tabID": 1,
            "component": "ApexCharts",
            "doNotRenderOnMount": true,
            "tabPath": "pie_chart",
            "tabIcon": "PieChart",
            "facetClass": "covers",
            "dropdownForResultClasses": true,
            "defaultResultClass": "coversByKeyword",
            "resultClasses": {
                "coversByKeyword": {
                    "sparqlQuery": "coversByKeywordQuery",
                    "filterTarget": "cover",
                    "resultMapper": "mapPieChart",
                    "sliceVisibilityThreshold": 0.01,
                    "dropdownForChartTypes": true,
                    "resultMapperConfig": {
                        "fillEmptyValues": false
                    },
                    "chartTypes": [
                        {
                            "id": "pie",
                            "createChartData": "createApexPieChartData"
                        },
                        {
                            "id": "bar",
                            "createChartData": "createApexBarChartData"
                        }
                    ]
                }

            }
        }
    },
    "properties": [
        {
            "id": "uri",
            "valueType": "object",
            "makeLink": true,
            "externalLink": true,
            "sortValues": true,
            "numberedList": false,
            "onlyOnInstancePage": true
        },
```

```
        {
            "id": "prefLabel",
            "valueType": "object",
            "makeLink": true,
            "externalLink": false,
            "sortValues": true,
            "numberedList": false,
            "minWidth": 200
        },
        {
            "id": "keyword",
            "valueType": "object",
            "makeLink": false,
            "externalLink": false,
            "sortValues": true,
            "numberedList": false,
            "minWidth": 150
        }
    ],
    "facets": {
        "prefLabel": {
            "containerClass": "one",
            "facetType": "text",
            "filterType": "textFilter",
            "sortBy": "prefLabel",
            "sortByPredicate": "skos:prefLabel",
            "textQueryProperty": "skos:prefLabel"
        },
        "keyword": {
            "containerClass": "ten",
            "facetType": "list",
            "facetLabelFilter": "FILTER(LANG(?prefLabel_) = '<LANG>')",
            "filterType": "uriFilter",
            "predicate": "kaunokki:asiasana",
            "searchField": true,
            "sortButton": true,
            "sortBy": "instanceCount",
            "sortByPredicate": "kaunokki:asiasana/skos:prefLabel",
            "sortDirection": "desc"
        }
    }
}
```

Figure 6.2: Abridged perspective configuration file with highlighted explanations for sections

component configuration as the initial results view object.

```
"properties": [
  {
    "id": "image",
    "valueType": "image",
    "previewImageHeight": 150,
    "makeLink": true,
    "externalLink": true,
    "sortValues": true,
    "numberedList": false,
    "hideHeader": true
  },
  {
    "id": "uri",
    "valueType": "object",
    "makeLink": true,
    "externalLink": true,
    "sortValues": true,
    "numberedList": false,
    "onlyOnInstancePage": true
  },
  ...
]
```

Listing 6.5: Abridged `properties` attribute in the BookSampo covers perspective configuration

The properties that are included in the table results view are specified through adding objects to the `properties` attribute (see Listing 6.5) list. Each property is configured through these objects to have properties like an ID, value type (e.g., object or string) and whether its values should function as a link. The values gotten from the queries are linked to the specified properties through the ID that should match between the configuration and queries file. Sampo-UI also offers optional visual configuration options to set things like column minimum width and showing properties only on instance pages.

```
 "facets": {
  "prefLabel": {
    "containerClass": "one",
    "facetType": "text",
    "filterType": "textFilter",
    "sortBy": "prefLabel",
    "sortByPredicate": "skos:prefLabel",
    "textQueryProperty": "skos:prefLabel"
  },
  "keyword": {
    "containerClass": "ten",
    "facetType": "list",
    "facetLabelFilter": "FILTER(LANG(?prefLabel_) = ’fi’)",
    "filterType": "uriFilter",
    "predicate": "kaunokki:asiasana",
    "searchField": true,
    "sortButton": true,
    "sortBy": "instanceCount",
    "sortByPredicate": "kaunokki:asiasana/skos:prefLabel",
    "sortDirection": "desc"
  },
  ...
}
```

Listing 6.6: Abridged `facets` attribute in the BookSampo covers perspective configuration

The facets included in the facet menu are configured through the `facets` attribute (see Listing 6.6). Each facet has its own attribute named after the property that is used as the facet. The values for these attributes should be objects that contain the configuration options for the facet. These configuration objects are used to specify the facet type as well as what predicate is used for the facet. Similar to the `properties` attribute, visual aspects of the facets can be configured in the objects.

Visualization components included in Sampo-UI framework were configured for the data by just using the configuration files and providing the queries for the data itself. For visualization components described as 'custom', the code for the components were written and added into the source code and made usable by using the same configurations as the Sampo-UI framework components were using. This encompassed both modifying the source code for components to expand them to handle new configuration options as well as writing new mapping functions for transforming the SPARQL query results data into a format that could be used with the components.

```
"visualizationTabNameHere": {
  "tabID": 1,
  "tabPath": "some_path",
  "tabIcon": "CalendarViewDay",
  "component": "componentNameHere",
  "sparqlQuery": "exampleQuery",
  "resultMapper": "exampleMapperFunction",
  ...
}
```

Listing 6.7: General visualization configuration

The general format of adding visualization tabs, or other additional tabs for that matter, can be seen in Listing 6.7. `tabID` determines the order of included tabs with `0` usually being reserved for the default table view and `tabIcon` determines the icon used for the tab from the list of options included in Sampo-UI. `tabPath` determines the path added to the URL of the page when the tab is opened. The type of the component that the tab contains and what data is used for that component is determined with `component` and `sparqlQuery` attributes respectively.

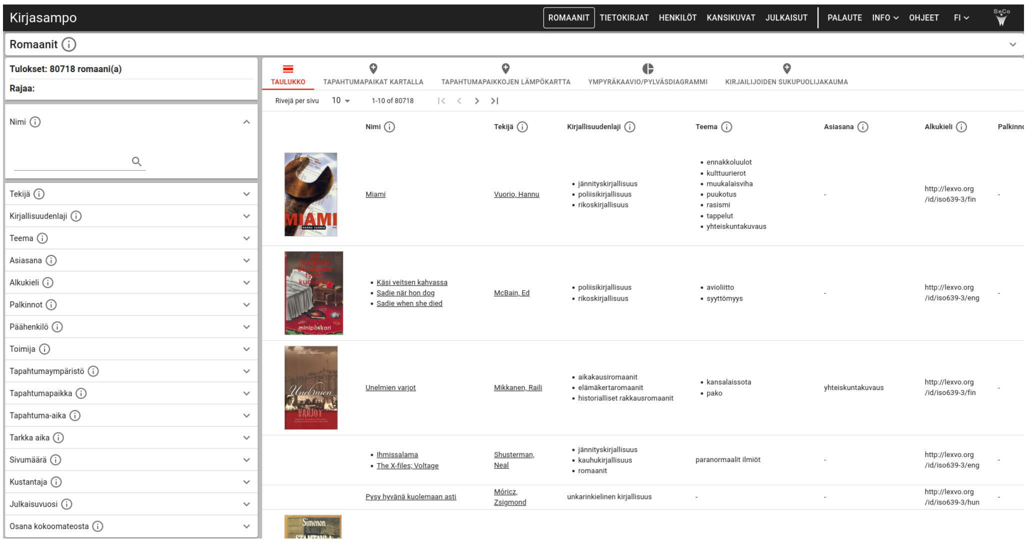
To transform the data into the correct format for the component, mapping functions are used. The `resultMapper` attribute is used for specifying the name of the mapping function in the `Mappers.js` file (found in the `src/server/sparql/` directory) containing all the different mapping functions. This `Mappers.js` is the file that should be used for adding custom mapping functions for new custom components. If any additional processing is required for the data after mapping, the `createChartData` attribute is added to the configuration for specifying the function name in the respective component configuration file.

Any additional configurations are passed through this same object based on the configurations the used component supports. The actual configurations used in the new portal will be presented in more detail for every different type of visualization when they are first introduced in the next sections.

6.2 Novels

This section covers the novels perspective. This perspective has all data and visualizations relating to novels on an abstract work level with its facet class set to `kaunokki:romaani`. The configuration of the portal is done using a Sampo-UI perspective-specific JSON configuration file. The physical manifestations of novels are covered in the publications perspective discussed later in this chapter.

6.2.1 Faceted search view



Tulokset: 80718 romaani(a)

Rajaa:

1-10 of 80718

	Nimi	Tekijä	Kirjallisuudenlaji	Teema	Asiasana	Alkukieli	Palkinn
	Miami	Vuorio, Hannu	• jännityskirjallisuus • poliisikirjallisuus • rikoskirjallisuus	• ennakkoluulot • kulttuurierot • muukalaisviha • puukotus • rasismi • tappelut • yhteiskuntakuvaus	-	http://lexvo.org/id/iso639-3/fin	-
	• Käsi veitsen kahvassa • Sadie när hon dog • Sadie when she died	McBain, Ed	• poliisikirjallisuus • rikoskirjallisuus	• avioliitto • syyttömyys	-	http://lexvo.org/id/iso639-3/eng	-
	Unelmien varjot	Mikkanen, Raili	• aikakausiromaanit • elämäkertaromaanit • historialliset rakkausromaanit	• kansalaissota • pako	yhteiskuntakuvaus	http://lexvo.org/id/iso639-3/fin	-
	• Ihmissalama • The X-files; Voltage	Shusterman, Neal	• jännityskirjallisuus • kauhukirjallisuus • romaanit	paranormaalit ilmiöt	-	http://lexvo.org/id/iso639-3/eng	-
	Pysy hyvänä kuolemaan asti	Móricz, Zsigmond	unkarinkielinen kirjallisuus	-	-	http://lexvo.org/id/iso639-3/hun	-

Figure 6.3: Novels perspective faceted search view

The default tab of the perspective is the table view showcasing the results dataset with other tabs including various visualizations as shown in Figure 6.3. The next subsections go into more detail on the available facets and visualizations.

Facets

All of the columns apart from that for cover images can be used to filter the data using the facets in the facet menu. The facet menu has one available free text search facet for searching for novels by name.

Facets for *author*, *genre*, *theme*, *keyword*, *original language*, *awards*, *main character*, *(other) characters*, *concrete setting*, *setting time*, *exact time of story*, *publisher*, *publication year* and *part of collective work* are all checkbox based. The values are ordered by default based on descending instance counts and labels with Finnish language tags are prioritized.

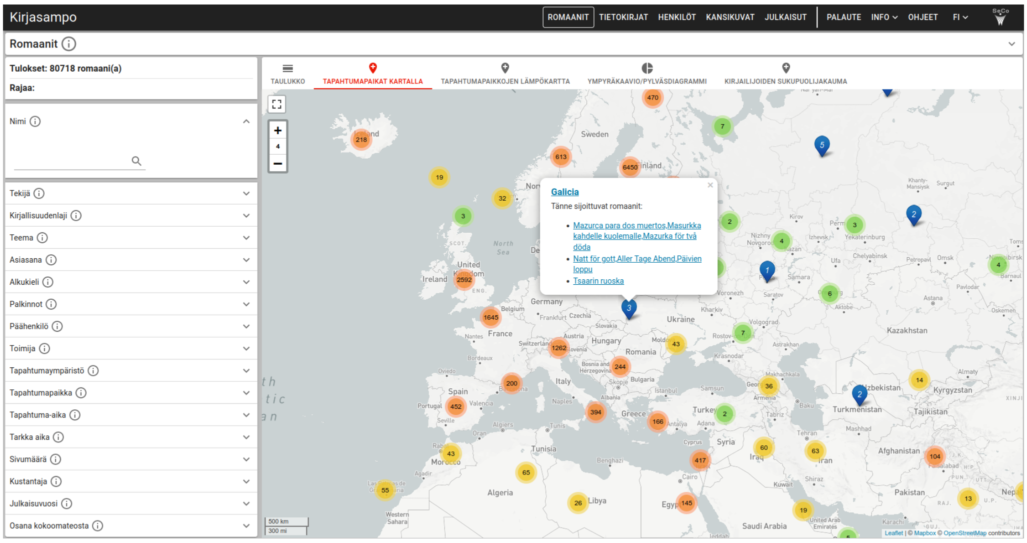
The exception with labels is the *original language* facet. As previously discussed in the last section of Chapter 4, most of the used language resources do not have proper preferred labels listed in the data. To keep the facet value labels as uniform as possible, URIs are shown for all languages instead of it being a mix of URIs and proper labels. Using `rdfs:label` values instead of the preferred labels listed with `skos:prefLabel` would have led to duplicates being listed in the facet list, so that solution was out of question as well.

The *awards* facet has a hierarchical structure defined through the parent property `kaunokki:palkintosarja`, where awards series are the parent property and the awards for specific years are the children. The facet for *publication year* could have been an integer range facet, but due to the non-uniform formats parsing would not have been guaranteed. Due to lacking hierarchical structure for most of the time periods the facet does not have any decade hierarchy set either, but rather all possible values are shown individually.

The *page count* facet is a range facet. It attempts to match the page count range to any provided page count for an object and ignores object values that cannot be converted into integers. This leads to some objects not being returned even if they technically fit the range, but the alternative of using checkbox facet would not really feel appropriate for page numbers as the facet would be cumbersome to use due to having to manually select all applicable page counts for a range and full of unique values only matching one object.

Visualizations

The novels perspective has four different visualizations tabs in addition to the default table view.



Tulokset: 80718 romaani(a)

Rajaa:

Galicia

Tänne sijoittuvat romaanit:

- Mazurca para dos muertos,Masurkka kahdelle kuolemalle,Mazurka för två döda
- Natt för gott,Aller Tage Abend,Päivien loppu
- Tsaarin ruoska

Figure 6.4: Leaflet map showcasing concrete novel settings

```
"novelsPlaces": {
  "tabID": 1,
  "tabPath": "map",
  "tabIcon": "AddLocation",
  "component": "LeafletMap",
  "showExternalLayers": false,
  "customMapControl": true,
  "sparqlQuery": "novelsPlacesQuery",
  "facetClass": "novels",
  "filterTarget": "novels",
  "resultMapper": "mapPlaces",
  "instanceConfig": {
    "propertiesQueryBlock": "placePropertiesInfoWindow",
    "relatedInstances": "novelsTakingPlaceAt",
    "createPopUpContent": "createPopUpContentBookSampo"
  }
}
```

Listing 6.8: Example Leaflet map configuration

The first visualization tab has a ready-to-use Sampo-UI Leaflet map component showcasing the different concrete settings for novels as shown in Figure 6.4. The configuration for this component can be seen in Listing 6.8. Each node on the map shows the number of novels taking place at that location. The query behind this visualizations queries all novels with a defined concrete setting that has coordinates and groups them based on the concrete setting's URI as well as latitude and longitude values. Duplicate coordinates are filtered out to avoid same place having multiple nodes with the same novels listed. Clicking on a node opens a custom-made tooltip (configured in `instanceConfig` in the listing) which shows the name of the place as well as a list of novels with that location as their concrete setting.

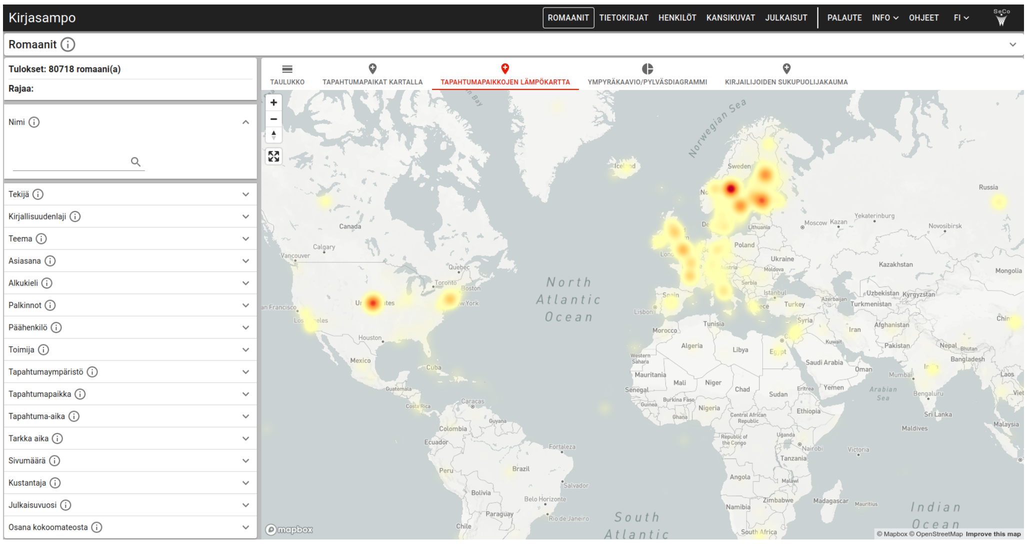
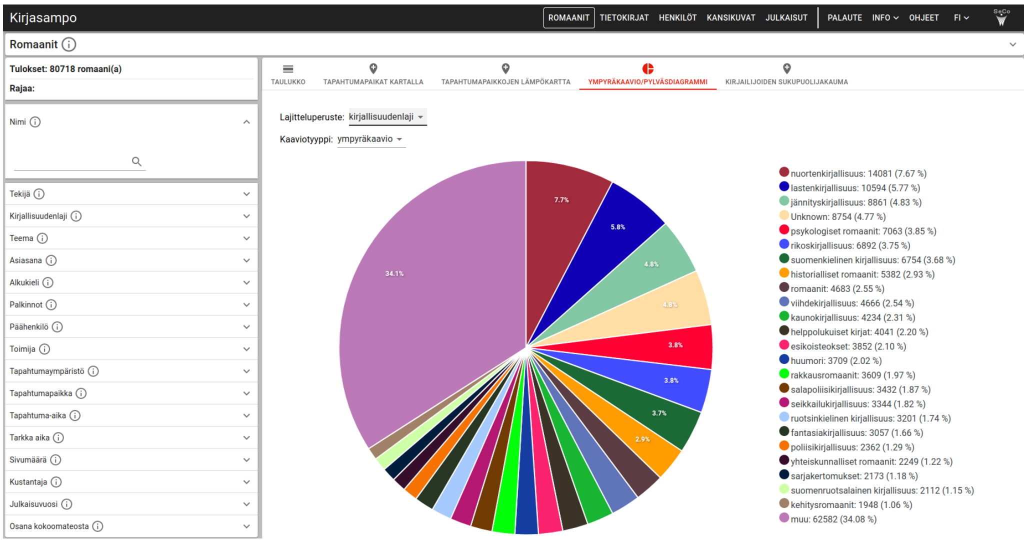


Figure 6.5: Heatmap showcasing concrete novel settings

```
"novelsPlacesHeatmap": {
  "tabID": 2,
  "tabPath": "heatmap",
  "tabIcon": "AddLocation",
  "component": "Deck",
  "layerType": "heatmapLayer",
  "sparqlQuery": "novelsPlacesQuery",
  "facetClass": "novels",
  "filterTarget": "novels",
  "resultMapper": "mapPlaces",
  "heatmapRadiusPixels": 50,
  "heatmapThreshold": 0.05,
  "heatmapIntensity": 1
}
```

Listing 6.9: Example heatmap configuration

The second visualization tab has a ready-to-use Sampo-UI heatmap component showcasing the same information of concrete novel settings as the previous one, but this time in a heatmap format as shown in Figure 6.5. The configuration for this component can be seen in Listing 6.9. The configuration for the rendering of the colored heatmap areas can passed through this configuration using the optional attributes heatmapRadiusPixels, heatmapThreshold, and heatmapIntensity. The query behind the data is the same as the previous visualization.



Tulokset: 80718 romaani(a)

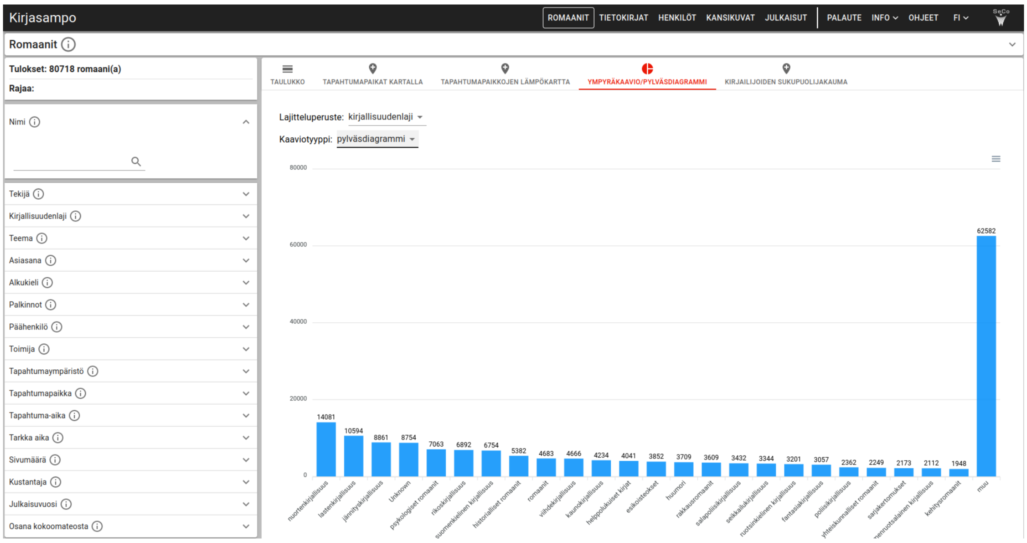
Rajaa:

Lajitteluperuste: kirjallisuudenlaji

Kaaviotyyppi: ympyräkaavio

- ● nuortenkirjallisuus: 14081 (7.67 %)
- ● lastenkirjallisuus: 10594 (5.77 %)
- ● jännityskirjallisuus: 8861 (4.83 %)
- ● Unknown: 8754 (4.77 %)
- ● psykologiset romaanit: 7063 (3.85 %)
- ● rikoskirjallisuus: 6892 (3.75 %)
- ● suomenkielinen kirjallisuus: 6754 (3.68 %)
- ● historialliset romaanit: 5382 (2.93 %)
- ● romaanit: 4683 (2.55 %)
- ● viihdekirjallisuus: 4666 (2.54 %)
- ● kaunokirjallisuus: 4234 (2.31 %)
- ● helppolukuiset kirjat: 4041 (2.20 %)
- ● esikoisteokset: 3852 (2.10 %)
- ● huumori: 3709 (2.02 %)
- ● rakkausromaanit: 3609 (1.97 %)
- ● salapoliisikirjallisuus: 3432 (1.87 %)
- ● seikkailukirjallisuus: 3344 (1.82 %)
- ● ruotsinkielinen kirjallisuus: 3201 (1.74 %)
- ● fantasiakirjallisuus: 3057 (1.66 %)
- ● poliisikirjallisuus: 2362 (1.29 %)
- ● yhteiskunnalliset romaanit: 2249 (1.22 %)
- ● sarjakertomukset: 2173 (1.18 %)
- ● suomenruotsalainen kirjallisuus: 2112 (1.15 %)
- ● kehitysromaanit: 1948 (1.06 %)
- ● muu: 62582 (34.08 %)

Figure 6.6: Pie chart showcasing top properties for novels



Tulokset: 80718 romaani(a)

Rajaa:

Lajitteluperuste: kirjallisuudenlaji

Kaaviotyyppi: pylväsdiagrammi

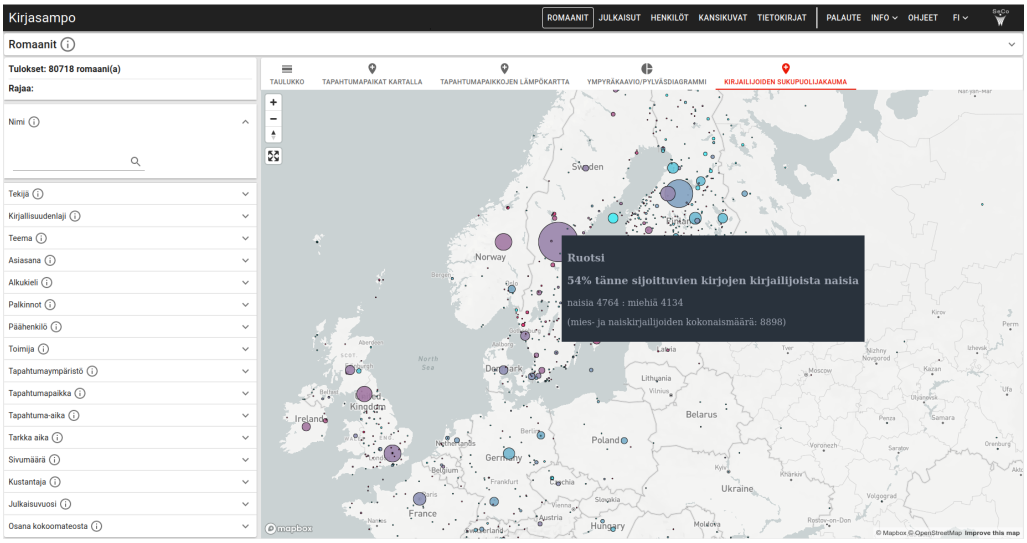
Figure 6.7: Bar chart showcasing top properties for novels

```
"novelsByProperty": {
  "tabID": 3,
  "component": "ApexCharts",
  "doNotRenderOnMount": true,
  "tabPath": "pie_chart",
  "tabIcon": "PieChart",
  "facetClass": "novels",
  "dropdownForResultClasses": true,
  "defaultResultClass": "novelsByOriginalLanguage",
  "resultClasses": {
    "novelsByGenre": {
      "sparqlQuery": "novelsByGenreQuery",
      "filterTarget": "novel",
      "resultMapper": "mapPieChart",
      "sliceVisibilityThreshold": 0.01,
      "dropdownForChartTypes": true,
      "resultMapperConfig": {
        "fillEmptyValues": false
      },
      "chartTypes": [
        {
          "id": "pie",
          "createChartData": "createApexPieChartData"
        },
        {
          "id": "bar",
          "createChartData": "createApexBarChartData"
        }
      ]
    },
    ...
  }
}
```

Listing 6.10: Abridged example pie/bar chart configuration

The third visualization tab has a pie chart for showcasing the top properties for novels as shown in Figure 6.6. The shown property can be chosen from a dropdown menu with the options of *original language*, *genre*, *theme*, *publisher*, *character* and *author gender*. Property values with less than specified percentage shares are combined under the label 'Other' in the pie chart, with the threshold value being configured specific to each different property. The pie chart can be changed into a bar chart format as shown in Figure 6.7 by using the second dropdown menu. The abridged configuration for both the pie and bar chart formats can be seen in Listing 6.10. The properties that can be visualized are defined as objects inside sub-attributes in the `resultClasses` attribute such as the `novelsByGenre` in the listing. Supported formats (i.e. pie and bar formats in `chartTypes`) and property-specific configurations like visibility thresholds are defined individually for each result class in `resultClasses`.

In the state of the data in the used BookSampo data dump, this pie/bar chart visualization behaves somewhat problematically. Publisher resources in the data are one of the resource types with preferred label problems: They largely do not have language tags and thus there are multiple preferred labels returned. This leads to the pie/bar chart (and the facet) to showing each publisher as many times as there are labels, which are counted in the chart as their own slices or bars. This leads to the percentages being skewed, but the visualization was chosen to be included as it is both a good way to showcase what kind of problems the data has and the visualization can also easily be modified to only get the correct preferred labels as soon as the underlying data is fixed.



Tulokset: 80718 romaani(a)
Rajaa:

Ruotsi
54% tänne sijoittuvien kirjojen kirjailijoista naisia
naisia 4764 : miehiä 4134
(mies- ja naiskirjailijoiden kokonaismäärä: 8898)

Figure 6.8: Map showcasing gender ratio of authors for novels taking place at location

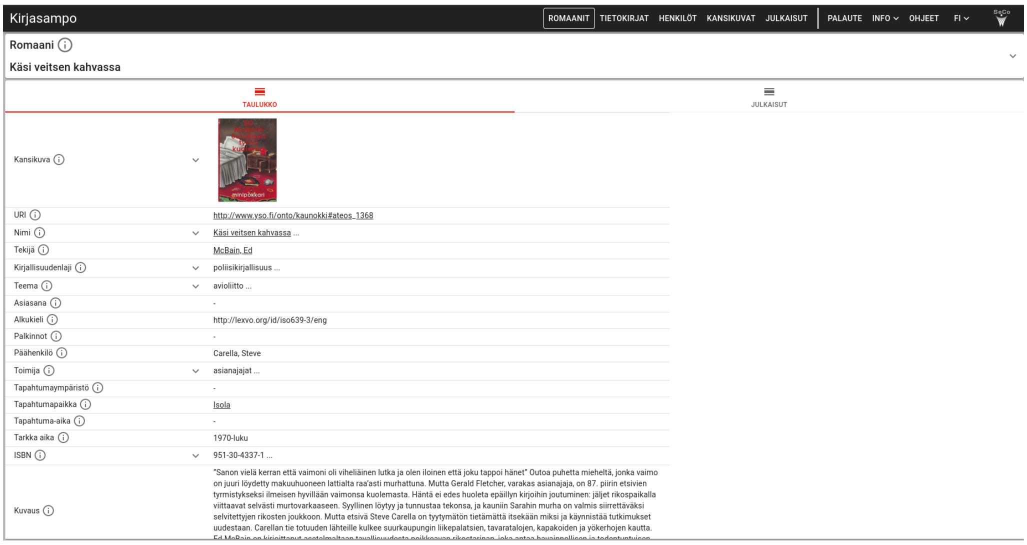
```
"authorsGenderScatterplot": {
  "tabID": 4,
  "tabPath": "scatterplot",
  "tabIcon": "AddLocation",
  "component": "Deck",
  "layerType": "scatterplotLayer",
  "sparqlQuery": "authorsGenderQuery",
  "facetClass": "novels",
  "filterTarget": "novels",
  "resultMapper": "mapPlacesRatio",
  "scatterplotStartColor": [61, 250, 255],
  "scatterplotEndColor": [233, 30, 99]
}
```

Listing 6.11: Example scatterplot configuration

The last visualization tab as shown in Figure 6.8 has a custom scatterplot component built with the DeckGL framework. The configuration can be seen in Listing 6.11. The component used is the `Deck` component included in Sampo-UI framework but with a layer type of `scatterplotLayer`.

The circles and their size on the map indicate the number of novels with that location as the setting and the color of the circle indicates the gender ratio of the authors of those novels. If all the authors are male, the circle is light blue (defined through the additional configuration option `scatterplotStartColor`) and pink (defined through the additional configuration option `scatterplotEndColor`) in the opposite case. Hovering over a circle opens a tooltip which indicates the name of the place, the exact ratio percentage as well as the exact author counts.

6.2.2 Instance page



Romaani

Käsi veitsen kahvassa

Kansikuva	
URI	http://www.yso.fi/onto/kaunokki#ateos_1368
Nimi	Käsi veitsen kahvassa ...
Tekijä	McBain, Ed
Kirjallisuudenlaji	poliisikirjallisuus ...
Teema	avioliitto ...
Asiasana	-
Alkukieli	http://lexvo.org/id/iso639-3/eng
Palkinnot	-
Päähenkilö	Carella, Steve
Toimija	asianajajat ...
Tapahtumaympäristö	-
Tapahtumapaikka	Isola
Tapahtuma-aika	-
Tarkka aika	1970-luku
ISBN	951-30-4337-1 ...
Kuvaus	"Sanon vielä kerran että vaimoni oli viheliäinen lutka ja olen iloinen että joku tappoi hänet" Outoa puhetta mieheltä, jonka vaimo on juuri löydetty makuuhuoneen lattialta raa'asti murhattuna. Mutta Gerald Fletcher, varakas asianajaja, on 87. piirin etsivien tyrmistykseksi ilmeisen hyvillään vaimonsa kuolemasta. Häntä ei edes huoleta epäillyn kirjoihin joutuminen: jäljet rikospaikalla viittaavat selvästi murtovarkaaseen. Syyllinen löytyy ja tunnustaa tekonsa, ja kauniin Sarahin murha on valmis siirrettäväksi selvitettyjen rikosten joukkoon. Mutta etsivä Steve Carella on tyytymätön tietämättä itsekään miksi ja käynnistää tutkimukset uudestaan. Carellan tie totuuden lähteille kulkee suurkaupungin liikepalatsien, tavaratalojen, kapakoiden ja yökerhojen kautta. Ed McBain on kirjoittanut asetelmaltaan tavallisuudesta poikkeavan rikostarinan, joka antaa havainnollisen ja todentuntuisen

Figure 6.9: Novels perspective instance page

The default view in a novel instance page is the table view as shown in Figure 6.9. The information included in the table is the same as in the facet search table view with the addition of the following fields:

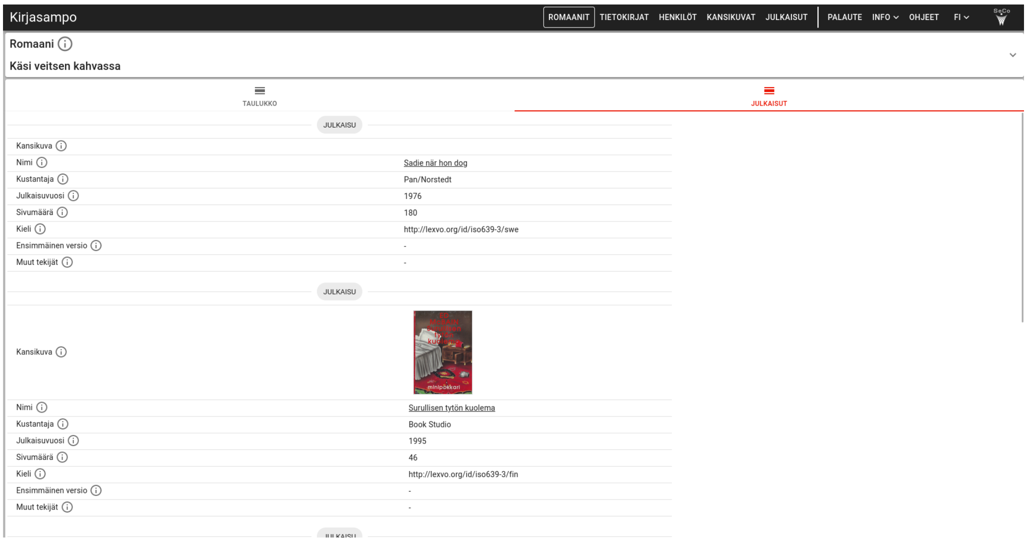
1. *URI.* The URI of the object is listed and the link leads to the SAHA editor page for the object.

2. *ISBN.* The ISBN number of the book.

3. *Description.* Field for describing the contents of the book from the back cover of the book.

4. *Review.* Link to a possible review page of the book, if one exists.

5. *Link to the novel on the original BookSampo portal.* Link that leads to the novel's page on the original BookSampo portal.



Romaani

Käsi veitsen kahvassa

JULKAISU

Kansikuva	
Nimi	Sadie när hon dog
Kustantaja	Pan/Norstedt
Julkaisuvuosi	1976
Sivumäärä	180
Kieli	http://lexvo.org/id/iso639-3/swe
Ensimmäinen versio	-
Muut tekijät	-

JULKAISU

Kansikuva	
Nimi	Surullisen tytön kuolema
Kustantaja	Book Studio
Julkaisuvuosi	1995
Sivumäärä	46
Kieli	http://lexvo.org/id/iso639-3/fin
Ensimmäinen versio	-
Muut tekijät	-

JULKAISU

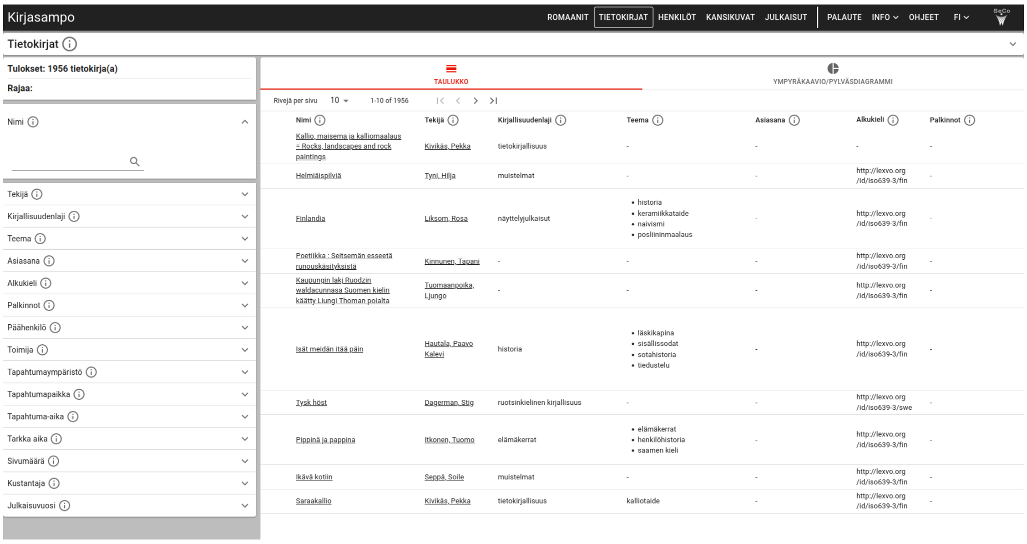
Figure 6.10: Publications tab of a novels perspective instance page

```
"novelInstancePagePublications": {
  "tabID": 1,
  "component": "InstancePageTableList",
  "fetchResultsWhenMounted": true,
  "tabPath": "publications",
  "tabIcon": "CalendarViewDay",
  "sparqlQuery": "novelPublicationsQuery",
  "filterTarget": "novel",
  "properties": [
    {
      "id": "image",
      "valueType": "image",
      "previewImageHeight": 150,
      "makeLink": true,
      "externalLink": true,
      "sortValues": true,
      "numberedList": false
    },
    {
      "id": "prefLabel",
      "valueType": "object",
      "makeLink": true,
      "externalLink": false,
      "sortValues": true,
      "numberedList": false,
      "minWidth": 200
    },
    ...
  ]
}
```

Listing 6.12: Abridged example configuration for a list of tables

In addition to the table view, novel instance pages also have a second tab for showing all the publications of the current novel. The component in that tab is a custom component of multiple tables added after each other in a list, one table for each publication. The abridged configuration for this component is listed in Listing 6.12. These tables list the publication specific information like *cover image*, *publication name*, *language*, *page count*, *publication year*, *publisher*, *other possible authors* and *whether or not the publication is the original publication of the novel* for all the publications if they are specified. The included properties in the configuration are listed in the same format as the general perspective-wide properties.

6.3 Nonfiction books



Tulokset: 1956 tietokirja(a)

1-10 of 1956

Nimi	Tekijä	Kirjallisuudenlaji	Teema	Asiasana	Alkukieli	Palkinnot
Kallio, maisema ja kalliomaalaus = Rocks, landscapes and rock paintings	Kivikäs, Pekka	tietokirjallisuus	-	-	-	-
Helmiäispilviä	Tyni, Hilja	muistelmat	-	-	http://lexvo.org /id/iso639-3/fin	-
Finlandia	Liksom, Rosa	näyttelyjulkaisut	• historia • keramiikkataide • naivismi • posliininmaalaus	-	http://lexvo.org /id/iso639-3/fin	-
Poetiikka : Seitsemän esseetä runouskäsityksistä	Kinnunen, Tapani	-	-	-	http://lexvo.org /id/iso639-3/fin	-
Kaupungin lakj Ruodzin waldacunnasa Suomen kielin käätty Liungi Thoman poialta	Tuomaanpoika, Ljungo	-	-	-	http://lexvo.org /id/iso639-3/fin	-
Isät meidän itää päin	Hautala, Paavo Kalevi	historia	• läskikapina • sisällissodat • sotahistoria • tiedustelu	-	http://lexvo.org /id/iso639-3/fin	-
Tysk höst	Dagerman, Stig	ruotsinkielinen kirjallisuus	-	-	http://lexvo.org /id/iso639-3/swe	-
Pippinä ja pappina	Itkonen, Tuomo	elämäkerrat	• elämäkerrat • henkilöhistoria • saamen kieli	-	http://lexvo.org /id/iso639-3/fin	-
Ikävä kotiin	Seppä, Soile	muistelmat	-	-	http://lexvo.org /id/iso639-3/fin	-
Saraakallio	Kivikäs, Pekka	tietokirjallisuus	kalliotaide	-	http://lexvo.org /id/iso639-3/fin	-

Figure 6.11: Nonfiction books perspective faceted search view

This section covers the nonfiction books perspective. The facet class of this perspective is set to `saha:Instance_ID1237984819752` and showcases all data and visualizations relating to nonfiction books on the abstract level. As in the previous case of novels perspective, this perspective is configured through its own specific JSON configuration file.

6.3.1 Faceted search view

Like in the previous novels perspective, this perspective's default tab in the faceted search view is the results table view as shown in Figure 6.11. As nonfiction books share most of their properties with novels, the chosen facets and columns to be shown to the user are the same for nonfiction books as they were for novels. Nonfiction books have a lot smaller number of objects in the data (1,956 nonfiction books vs. 80,718 novels). They are also being fairly scarcely annotated for fields like concrete settings, so the perspective only shares the pie/bar chart visualization with the novels perspective with *genre*, *original language*, *theme*, *publisher* and *author gender* being the properties showcased.

6.3.2 Instance pages

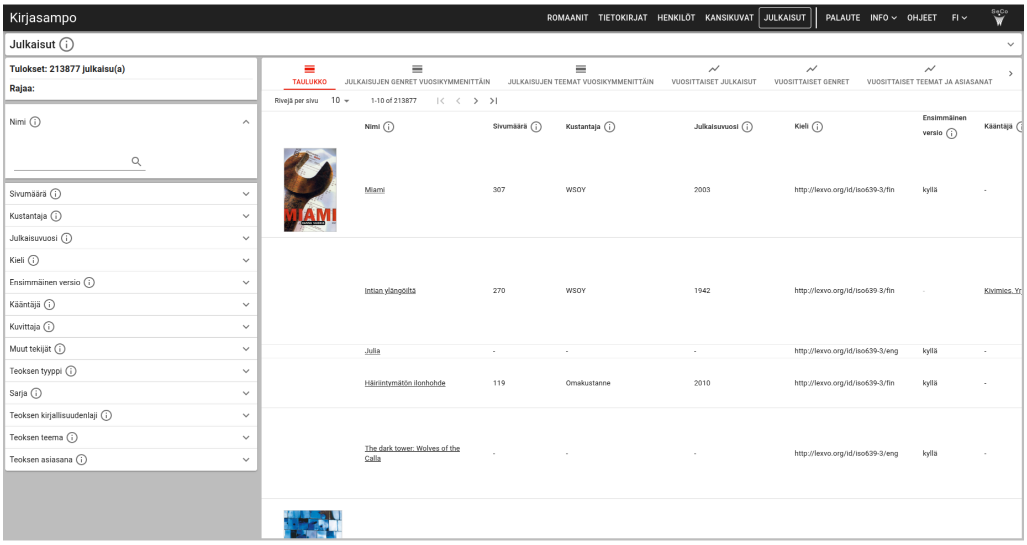
Just as in the faceted search view, the instance pages for nonfiction books are largely the same as they are for novels. The only new addition is the property for *YKL class*, the class for the book in the classification system for public libraries. *Review* and *Link to the book on the original BookSampo portal* properties from the novels perspective have been omitted from the nonfiction book instance pages. Similarly to novels, the nonfiction book instance pages have the same publications tab using the same custom component for different publications.

6.4 Publications

This section covers the publications perspective. The facet class of this perspective is set to `kaunokki:fyysinen_teos` and covers all data and visualizations relating to all publications on the physical level. As in the previous cases, this perspective is configured through its own specific JSON configuration file.

6.4.1 Faceted search view

As previously, the default tab of the perspective is the table view showcasing the results dataset with other tabs including various visualizations as shown in Figure 6.12. Available facets and visualizations are introduced in the next subsections.



Nimi	Sivumäärä	Kustantaja	Julkaisuvuosi	Kieli	Ensimmäinen versio	Kääntäjä
Miami	307	WSOY	2003	http://lexvo.org/id/iso639-3/fin	kyllä	-
Intian ylängöiltä	270	WSOY	1942	http://lexvo.org/id/iso639-3/fin	-	Kivimies, Yr
Julia	-	-	-	http://lexvo.org/id/iso639-3/eng	kyllä	-
Häiriintymätön ilonhohde	119	Omakustanne	2010	http://lexvo.org/id/iso639-3/fin	kyllä	-
The dark tower: Wolves of the Calla	-	-	-	http://lexvo.org/id/iso639-3/eng	kyllä	-

Figure 6.12: Publications perspective faceted search view

Facets

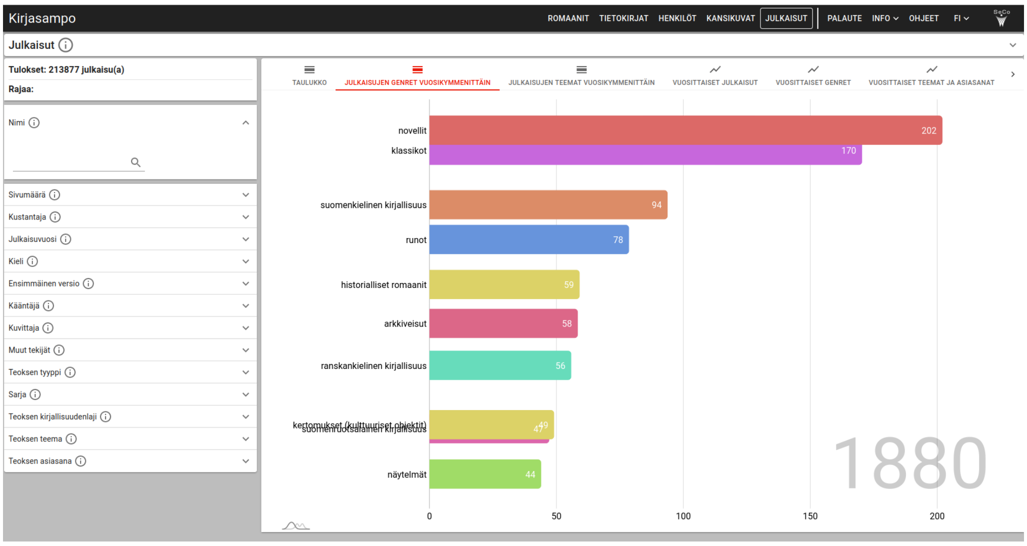
All of the columns apart from the cover image, the abstract level work and the parts of the publication can be used to filter the data using the facets in the facet menu. A free text search facet is available for searching by publication name.

Facets related directly to the publications for *publisher*, *publication year*, *language*, *first version*, *translator*, *illustrator*, *other authors*, and *series* are all checkbox facets with the facet values ordered based on instance counts in descending order by default. The facet for *page count* is a range facet as it was for novels, with the same caveats.

In addition to facets directly related to the publications itself, the facet menu and results table include information regarding the abstract level work. The user has checkbox facets for *work type*, *work genre*, *work theme* and *work keyword* for filtering the results based on the abstract level works. If the abstract level work is either a novel or a nonfiction book, the column with the abstract level work name has a link to the instance page of that work. Otherwise just the name is shown without a hyperlink as instance pages do not exist for other kinds of works.

Visualizations

As all release-related time information is included in the data about the publications, time series visualizations are included in this perspective. The first two visualizations are animated and use the default results set without any filters. The visualizations afterward change depending on the chosen facets.



Tulokset: 213877 julkaisu(a)

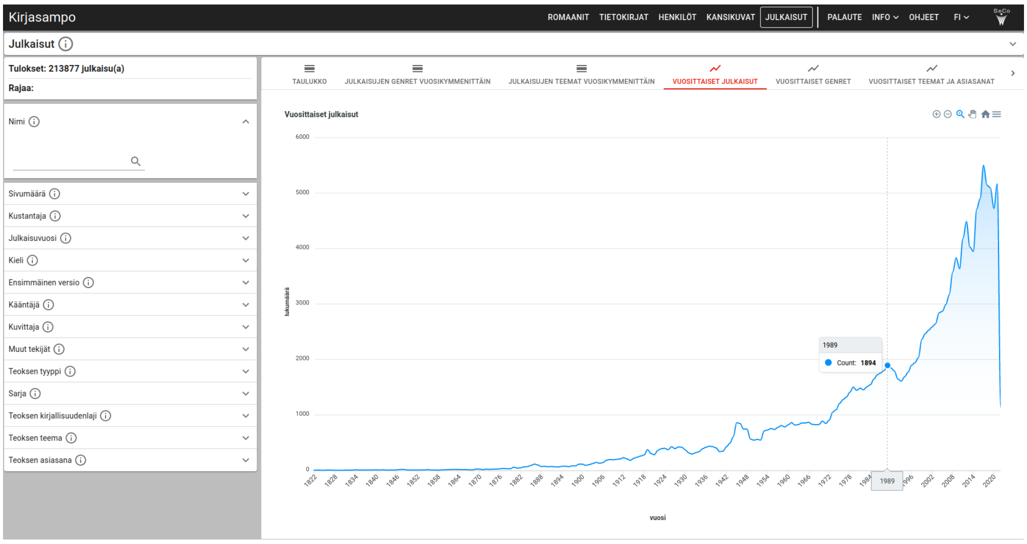
Rajaa:

Figure 6.13: Bar chart race of publication genres throughout decades

```
"publicationsByDecadeAndGenre": {
  "tabID": 1,
  "tabIcon": "CalendarViewDay",
  "tabPath": "publications_by_decade_and_genre",
  "component": "BarChartRace",
  "stepBegin": 1700,
  "stepEnd": 2020,
  "stepIncrement": 10,
  "stepDuration": 3000,
  "sparqlQuery": "publicationsByDecadeAndGenreQuery",
  "facetClass": "publications",
  "filterTarget": "publication",
  "resultMapper": "makeObjectList",
  "postprocess": {
    "func": "toBarChartRaceFormat",
    "config": {
      "step": 10
    }
  }
}
```

Listing 6.13: Example bar chart race configuration

The first two visualizations are bar chart races showing the top genres (shown in Figure 6.13, configuration in Listing 6.13) and top themes throughout the decades. These are implemented using the bar chart race component included in Sampo-UI that is built with the amCharts[5] library. The visualizations are set to begin from 1700 (specified in `stepBegin`) and go all the way to 2020 (`stepEnd`). The start date was chosen to be as early as possible while still having enough annotated books in that period to make it reasonable to visualize. Due to the problems with the time resources mentioned in Chapter 4 as well as the sheer amount of data, queries for these two visualizations can take a while to finish.



Tulokset: 213877 julkaisu(a)

Rajaa:

1989

Count: 1894

Figure 6.14: Publication counts throughout years

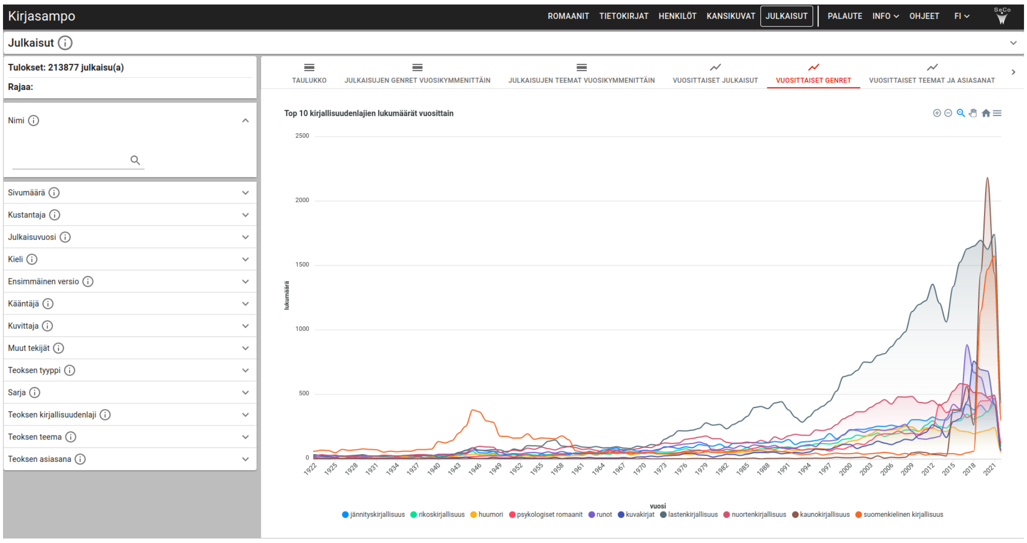
```
"publicationsByYearLineChart": {
  "tabID": 3,
  "component": "ApexCharts",
  "tabPath": "publication_years",
  "tabIcon": "ShowChart",
  "sparqlQuery": "publicationsByYearLineChartQuery",
  "facetClass": "publications",
  "filterTarget": "publication",
  "resultMapper": "mapLineChart",
  "resultMapperConfig": {
    "fillEmptyValues": true
  },
  "createChartData": "createZoomableTimeSeriesData",
  "xaxisTitle": "year",
  "xaxisType": "category",
  "xaxisTickAmount": 30,
```

[5] https://www.amcharts.com/

```
  "yaxisTitle": "count",
  "seriesTitle": "Count",
  "stroke": {
    "width": 2
  }
}
```

Listing 6.14: Example singleline time series configuration

The next visualization tab has a custom time series component for showcasing the absolute number of publications by years (shown in Figure 6.14, configuration in Listing 6.14). The component is configured to fill in years with no values with zeroes (specified through `fillEmptyValues`) to create a continuous time series. The query attempts to convert all publications' preferred publication years to integers and groups them based on that integer number while filtering out the ones where no valid integer was gotten. The counts shown in the visualization by default include all publications instead of just including first publications. To see the first publication specific information, the user can just check the checkbox for *first version* facet.

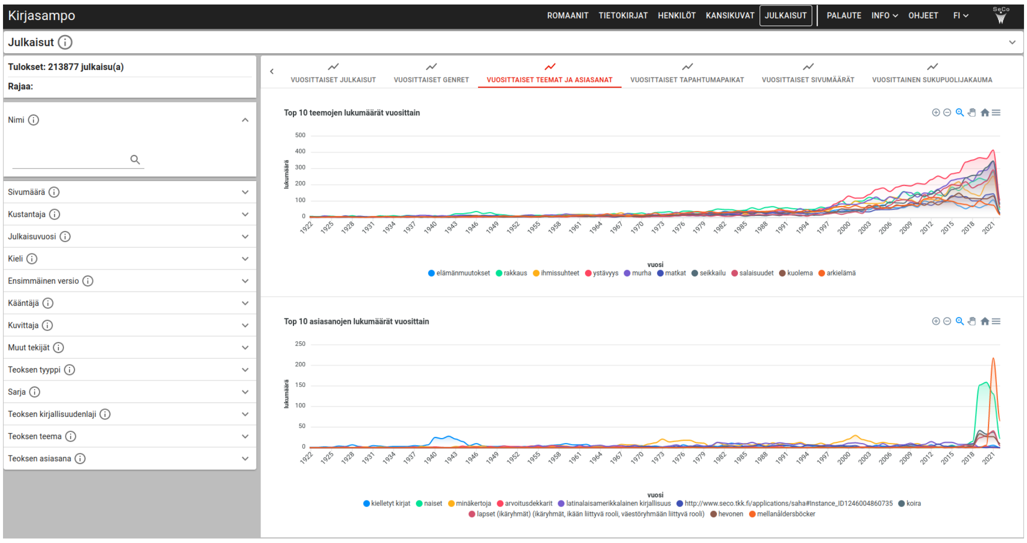


Tulokset: 213877 julkaisu(a)

Rajaa:

Figure 6.15: Top 10 publication genres throughout years

```
"publicationGenresByYearLineChart": {
  "tabID": 4,
  "component": "ApexCharts",
  "tabPath": "publication_genre_years",
  "tabIcon": "ShowChart",
  "sparqlQuery": "genresByYearTimeSeriesQuery",
  "facetClass": "publications",
  "filterTarget": "publication",
  "resultMapper": "mapZoomableMultipleLineTimeSeries",
```



Tulokset: 213877 julkaisu(a)

Rajaa:

Figure 6.16: Top 10 publication themes and keywords throughout years

```
  "resultMapperConfig": {
    "fillEmptyValues": true
  },
  "createChartData": "createZoomableMultipleLineTimeSeriesData",
  "xaxisTitle": "year",
  "xaxisType": "category",
  "xaxisTickAmount": 30,
  "yaxisTitle": "count",
  "seriesTitle": "Count",
  "stacked": false,
  "stroke": {
    "width": 2
  }
}
```

Listing 6.15: Example multiline time series configuration

The next three visualization tabs have custom multiline time series components that group publications by some property and show the publication counts for top ten of these properties throughout the years. The first of these visualizations, shown in Figure 6.15 and configured like in Listing 6.15, shows the publication count for the top ten genres present in the results set. By applying additional facets the top ten genres are calculated again.

```
"themeAndKeywordTimespanLineChart": {
  "tabID": 5,
  "component": "ApexChartsDouble",
  "tabPath": "themes_and_keywords",
  "tabIcon": "ShowChart",
  "upperResultClass": "themeTimespanLineChart",
  "lowerResultClass": "keywordTimespanLineChart",
  "resultClasses": {
    "themeTimespanLineChart": {
      "height": "50%",
      "sparqlQuery": "themesByYearTimeSeriesQuery",
      "facetClass": "publications",
      "filterTarget": "publication",
      "resultMapper": "mapZoomableMultipleLineTimeSeries",
      "resultMapperConfig": {
        "fillEmptyValues": true
      },
      "createChartData": "createZoomableMultipleLineTimeSeriesData",
      "xaxisTitle": "year",
      "xaxisType": "category",
      "xaxisTickAmount": 30,
      "yaxisTitle": "count",
      "seriesTitle": "Count",
      "stacked": false,
      "stroke": {
        "width": 2
      }
    },
    "keywordTimespanLineChart": {
      "height": "50%",
      "sparqlQuery": "keywordsByYearTimeSeriesQuery",
      "facetClass": "publications",
      "filterTarget": "publication",
      "resultMapper": "mapZoomableMultipleLineTimeSeries",
      "resultMapperConfig": {
        "fillEmptyValues": true
      },
      "createChartData": "createZoomableMultipleLineTimeSeriesData",
      "xaxisTitle": "year",
      "xaxisType": "category",
      "xaxisTickAmount": 30,
      "yaxisTitle": "count",
      "seriesTitle": "Count",
      "stacked": false,
      "stroke": {
        "width": 2
      }
    }
  }
}
```

Listing 6.16: Example configuration for multiple stacked multiline time series components

The next visualization has a component with two custom multiline time series components stacked on top of each other as shown in Figure 6.16. The configuration for this component can be seen in Listing 6.16. The component handles the data using the same functions as the previous time series component but combines them together into one component where the `resultClasses` attribute holds the configurations for individual time series charts. The top component showcases the top ten themes throughout the years while the bottom one showcases the top ten keywords. These two components are combined into one tab due to their relation to each other: theme is the preferred field to be used for including themes of the book, but if there is no appropriate theme in the data, keywords are used for supplementing them. Showing both of these at the same time with same date range being shown should then give a slightly better image of theme evolution by year. Like in the previous visualization, choosing additional facets will cause the top ten themes and keywords to be recalculated based on the resulting result set.

The last of the multiline time series components deals with novel concrete settings and uses the previously introduced non-stacked multiline time series component. The query groups publications by their setting and then returns the publication counts for the top ten settings. The top ten places are recalculated if additional facets are chosen as in the previous cases.

The second to last visualization page for publications has the average page count for publications showcased by year as a time series visualization similar to the previously mentioned publication count visualization. This component was inspired by Mäkinen's article *Romaanit pidentyneet* [47] mentioned at the end of Chapter 5. The query attempts to convert the provided publication page numbers to integers and filters out the ones that fail. The query then returns the average page number of the publications for each year.

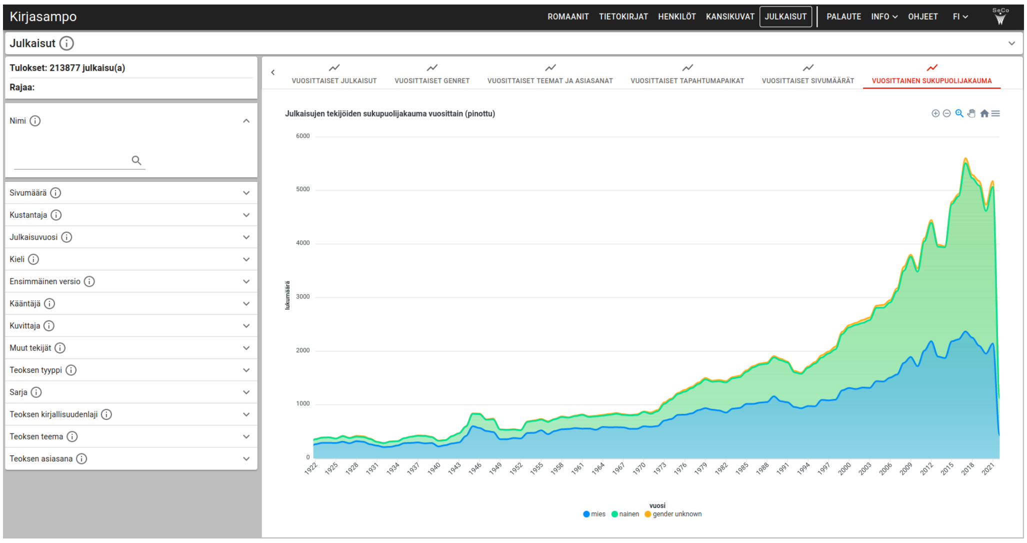


Figure 6.17: Publication author gender ratio throughout years

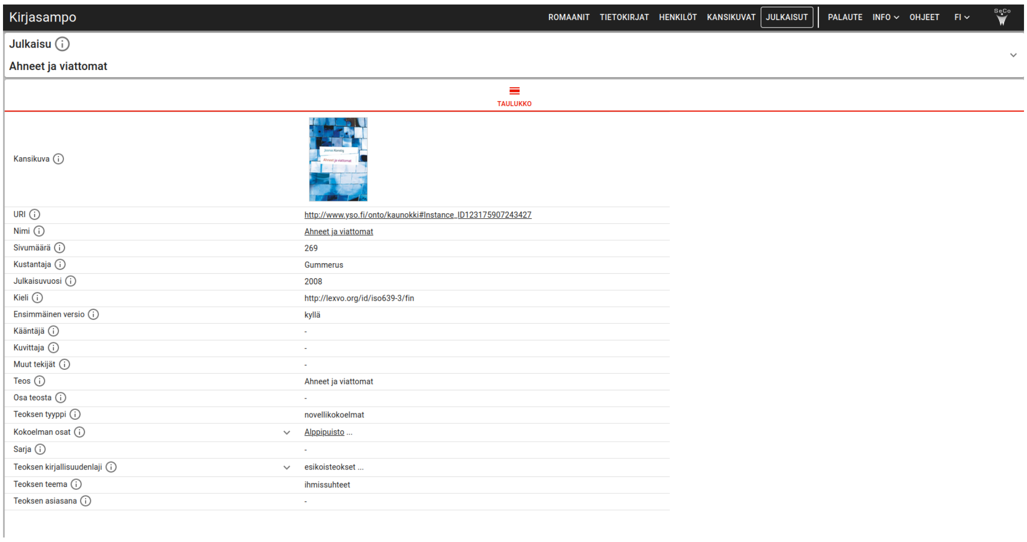
```
"publicationGenderRatioByYearLineChart": {
  "tabID": 8,
  "component": "ApexCharts",
  "tabPath": "publication_gender_ratio_years",
  "tabIcon": "ShowChart",
  "sparqlQuery": "genderRatiosByYearTimeSeriesQuery",
  "facetClass": "publications",
  "filterTarget": "publication",
  "resultMapper": "mapZoomableMultipleLineTimeSeries",
  "resultMapperConfig": {
    "fillEmptyValues": true
  },
  "createChartData": "createZoomableMultipleLineTimeSeriesData",
  "xaxisTitle": "year",
  "xaxisType": "category",
  "xaxisTickAmount": 30,
  "yaxisTitle": "count",
  "seriesTitle": "Count",
  "stacked": true,
  "stroke": {
    "width": 3
  },
  "fill": {
    "type": "gradient",
    "gradient": {
      "opacityFrom": 0.6,
      "opacityTo": 0.8
    }
  }
}
```

Listing 6.17: Example stacked time series configuration

The last visualization tab has a custom stacked time series chart component as shown in Figure 6.17. The configuration for this component can be seen in Listing 6.17. This component uses the same mapping functions for handling the data as the previously introduced multiline time series components but has additional configuration through the `fill` attribute and the `stacked` attribute set to true. The query behind this visualization returns the number of publications by year for all different specified gender values. Publications with a listed author with no information about their gender are included under a 'gender unknown' category. Publications with no information regarding their author are ignored by the query. These publications counts gotten from the query for each year are then shown as a stacked time series graph.

6.4.2 Instance pages

An example of a publication instance page is shown in Figure 6.18. Publication instance pages show the same information as the faceted search table view with the only addition of showing the URI of the object with a hyperlink to the instance's SAHA metadata editor page.



Julkaisu ⓘ

Ahneet ja viattomat

Kansikuva ⓘ	
URI ⓘ	http://www.yso.fi/onto/kaunokki#Instance_ID123175907243427
Nimi ⓘ	Ahneet ja viattomat
Sivumäärä ⓘ	269
Kustantaja ⓘ	Gummerus
Julkaisuvuosi ⓘ	2008
Kieli ⓘ	http://lexvo.org/id/iso639-3/fin
Ensimmäinen versio ⓘ	kyllä
Kääntäjä ⓘ	-
Kuvittaja ⓘ	-
Muut tekijät ⓘ	-
Teos ⓘ	Ahneet ja viattomat
Osa teosta ⓘ	-
Teoksen tyyppi ⓘ	novellikokoelmat
Kokoelman osat ⓘ	Alppipuisto ...
Sarja ⓘ	-
Teoksen kirjallisuudenlaji ⓘ	esikoisteokset ...
Teoksen teema ⓘ	ihmissuhteet
Teoksen asiasana ⓘ	-

Figure 6.18: Publications perspective instance page

6.5 Covers

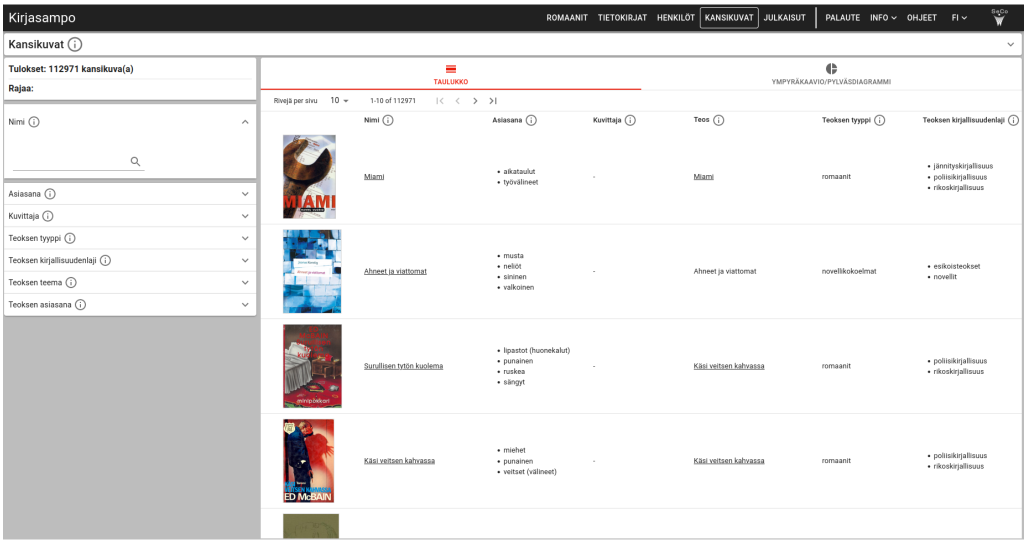
This section covers the covers perspective with the facet class of `kaunokki:kansi`. This perspective covers all data and visualizations relating to all book cover images. Like before this perspective is configured through its own specific JSON configuration file.

6.5.1 Faceted search view

The default tab of the perspective is the table view showcasing the results dataset with the other tab including a pie/bar chart visualization as shown in Figure 6.19. Available facets and visualizations are introduced in the next subsections.

Facets

The facet menu includes a free text search for searching for covers by their name. The other facets related to the covers, *keyword* and *illustrator*, are checkbox facets with their facet values ordered in a descending order by their instance counts. In addition to these facets, there are also a few checkbox facets relating to the work that the cover is for: *work type*, *work genre*, *work theme* and *work keyword*.



Tulokset: 112971 kansikuva(a)

Rajaa:

Rivejä per sivu 10 1-10 of 112971

Nimi	Asiasana	Kuvittaja	Teos	Teoksen tyyppi	Teoksen kirjallisuudenlaji
Miami	aikataulut työvälineet	-	Miami	romaanit	jännityskirjallisuus poliisikirjallisuus rikoskirjallisuus
Ahneet ja viattomat	musta neliöt sininen valkoinen	-	Ahneet ja viattomat	novellikokoelmat	esikoisteokset novellit
Surullisen tytön kuolema	lipastot (huonekalut) punainen ruskea sängyt	-	Käsi veitsen kahvassa	romaanit	poliisikirjallisuus rikoskirjallisuus
Käsi veitsen kahvassa	miehet punainen veitset (välineet)	-	Käsi veitsen kahvassa	romaanit	poliisikirjallisuus rikoskirjallisuus

Figure 6.19: Covers perspective faceted search view

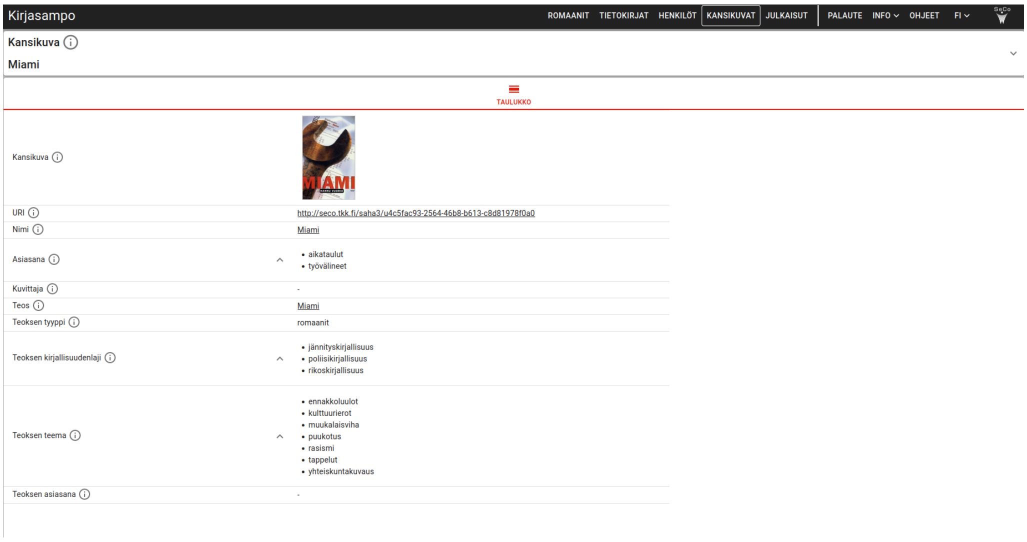
All these facets have a corresponding column in the table view. In addition, the table has a column for the name of the work for which the cover is made for. If the cover is made for a novel or a nonfiction book, the name of the work is also a hyperlink to an instance page of the work.

Visualizations

The covers perspective has one visualization tab with a Sampo-UI's ready-to-use pie and bar chart component. The component is configured to support showing bar and pie charts of *cover keywords*, *work types*, *work genres* and *work themes*.

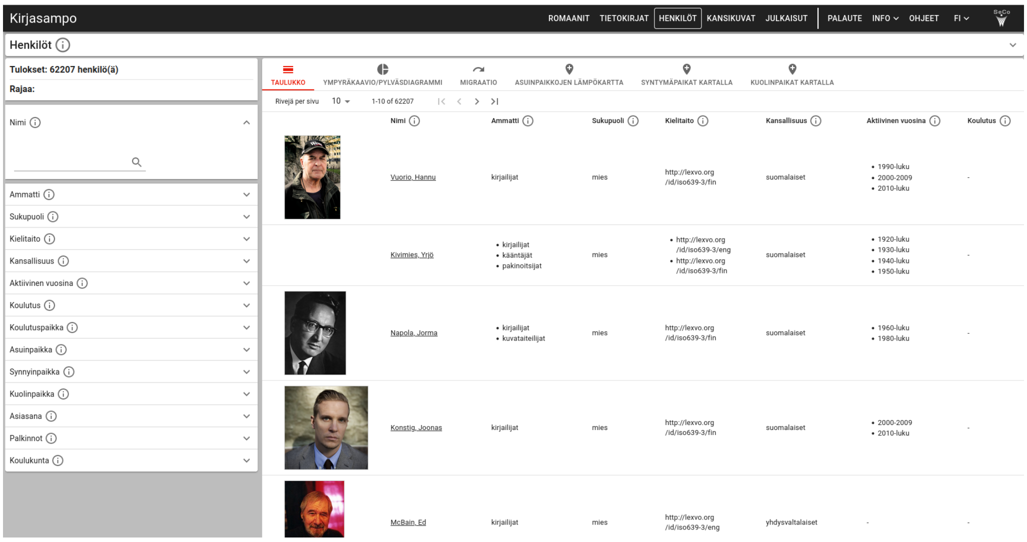
6.5.2 Instance pages

An example instance page for a cover is shown in Figure 6.20. The instance pages for covers include the same information as the faceted search table view with the addition of the object's URI and a link to the SAHA metadata editor page for that instance.



Kansikuva	
URI	http://seco.tkk.fi/saha3/u4c5fac93-2564-46b8-b613-c8d81978f0a0
Nimi	Miami
Asiasana	• aikataulut • työvälineet
Kuvittaja	-
Teos	Miami
Teoksen tyyppi	romaanit
Teoksen kirjallisuudenlaji	• jännityskirjallisuus • poliisikirjallisuus • rikoskirjallisuus
Teoksen teema	• ennakkoluulot • kulttuurierot • muukalaisviha • puukotus • rasismi • tappelut • yhteiskuntakuvaus
Teoksen asiasana	-

Figure 6.20: Covers perspective instance page



Tulokset: 62207 henkilö(ä)

Rajaa:

	Nimi	Ammatti	Sukupuoli	Kielitaito	Kansallisuus	Aktiivinen vuosina	Koulutus
	Vuorio, Hannu	kirjailijat	mies	http://lexvo.org/id/iso639-3/fin	suomalaiset	• 1990-luku • 2000-2009 • 2010-luku	-
	Kivimies, Yrjö	• kirjailijat • kääntäjät • pakinoitsijat	mies	• http://lexvo.org/id/iso639-3/eng • http://lexvo.org/id/iso639-3/fin	suomalaiset	• 1920-luku • 1930-luku • 1940-luku • 1950-luku	-
	Napola, Jorma	• kirjailijat • kuvataiteilijat	mies	http://lexvo.org/id/iso639-3/fin	suomalaiset	• 1960-luku • 1980-luku	-
	Konstig, Joonas	kirjailijat	mies	http://lexvo.org/id/iso639-3/fin	suomalaiset	• 2000-2009 • 2010-luku	-
	McBain, Ed	kirjailijat	mies	http://lexvo.org/id/iso639-3/eng	yhdysvaltalaiset	-	-

Figure 6.21: People perspective faceted search view

6.6 People

This section covers the people perspective which deals with all people related to the literature in the data from authors, translators, and illustrators to reviewers. The facet class of the perspective is set to `foaf:Person` and the configuration for the perspective is done through its own specific JSON configuration file.

6.6.1 Faceted search view

The default tab of the perspective is the table view showcasing the results dataset with various visualizations included in the other tabs as shown in Figure 6.21. Available facets and visualizations are introduced in the next subsections.

Facets

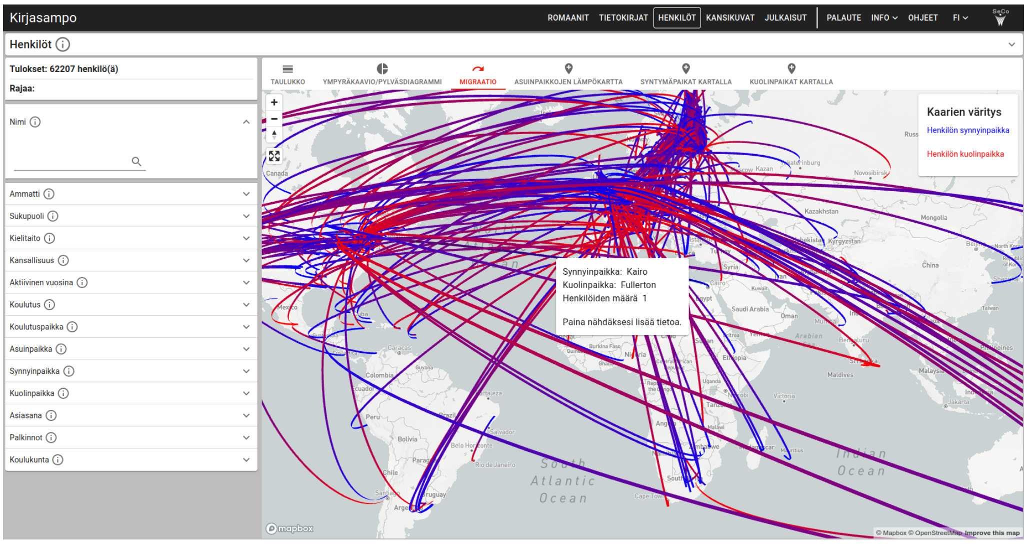
The facets for the perspective are mostly the same properties as the properties in the result table views columns. The table has additional columns for time of birth and time of death that were too specific to be used as facets for filtering people as the time resources in the data do not reliably include hierarchy. The people perspective has a free text search facet for searching for people by name. All the other available facets are checkbox facets with the facet values ordered by their instance counts in a descending order.

The other checkbox facets cover basic information about the person's life like their *occupation*, *gender*, *language abilities*, *nationality*, *active years*, *keywords*, and *associated school (of thought)* as well as things related to the event in the person's life like *education* and *education place*, *places they have live in*, *place of birth*, and *place of death* as well as possible *awards*.

Visualizations

The first visualization tab has the same Sampo-UI pie and bar chart component as used in other perspectives. The available properties for visualizing for people are *gender*, *occupation*, *nationality*, and *genres written*.

```
"peopleMigrations": {
  "tabID": 2,
  "component": "Deck",
  "tabPath": "migrations",
  "tabIcon": "Redo",
  "sparqlQuery": "peopleMigrationsQuery",
  "facetClass": "people",
  "filterTarget": "person",
  "layerType": "arcLayer",
```



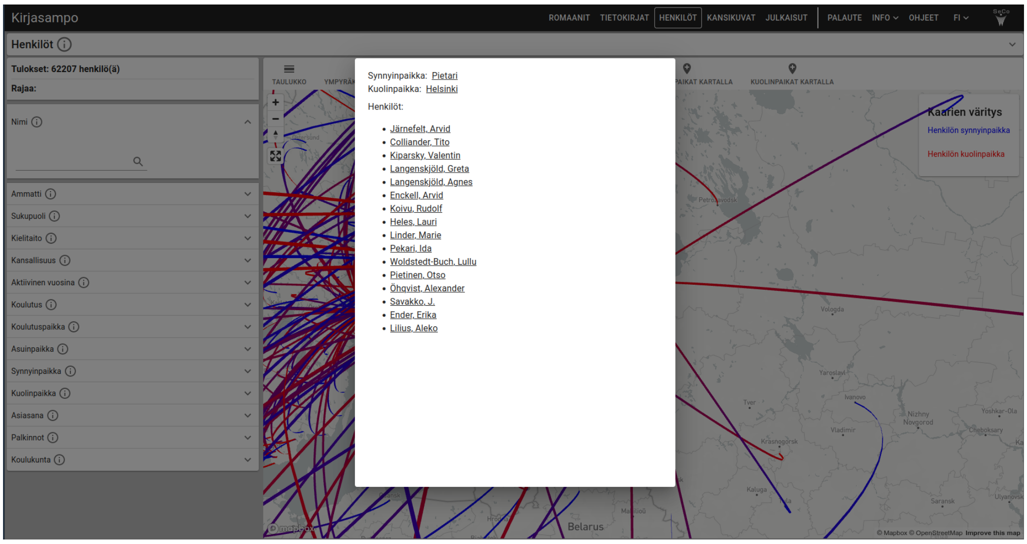
Tulokset: 62207 henkilö(ä)

Rajaa:

Synnyinpaikka: Kairo
Kuolinpaikka: Fullerton
Henkilöiden määrä 1

Paina nähdäksesi lisää tietoa.

Figure 6.22: Migrations visualization for people based on places of birth and death



Tulokset: 62207 henkilö(ä)

Rajaa:

Synnyinpaikka: Pietari
Kuolinpaikka: Helsinki

Henkilöt:

- Järnefelt, Arvid
- Colliander, Tito
- Kiparsky, Valentin
- Langenskjöld, Greta
- Langenskjöld, Agnes
- Enckell, Arvid
- Koivu, Rudolf
- Heles, Lauri
- Linder, Marie
- Pekari, Ida
- Woldstedt-Buch, Lullu
- Pietinen, Otso
- Öhqvist, Alexander
- Savakko, J.
- Ender, Erika
- Lilius, Aleko

Figure 6.23: Tooltip for migrations visualization

```
  "arcWidthVariable": "instanceCountScaled",
  "instanceVariable": "person",
  "showTooltips": true,
  "postprocess": {
    "func": "linearScale",
    "config": {
      "variable": "instanceCount",
      "minAllowed": 3,
      "maxAllowed": 30
    }
  }
},
"peopleMigrationsDialog": {
  "sparqlQuery": "peopleMigrationsDialogQuery",
  "filterTarget": "id",
  "resultMapper": "makeObjectList"
}
```

Listing 6.18: Example migrations map configuration

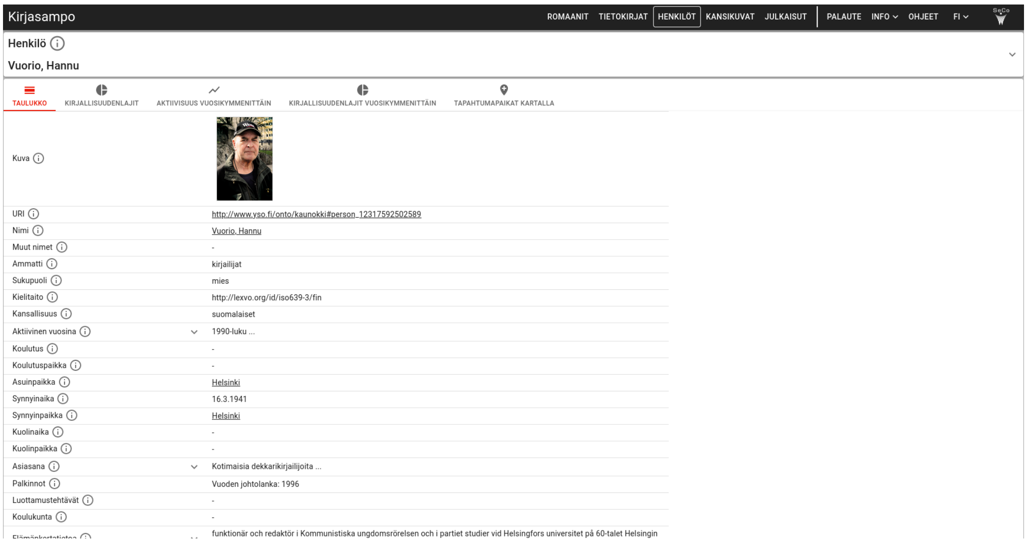
The second visualization (shown in Figure 6.22, configuration in the attribute called `peopleMigrations` in Listing 6.18) is a map visualization showcasing how people have migrated between their birth and death. The component used is a DeckGL map component included in the Sampo-UI framework. The first end of the migration lines (blue) starts at the place of birth of a person and the other end (red) is at the location where the people who were born at the first location have died. The thickness of the arc line indicates the number of people with these places of birth and death.

Hovering over any of the arcs lists the names of the starting and ending place of the arc as well as the count of people included in the arc. Clicking the arc opens up a pop-up (configuration in `peopleMigrationsDialog` in Listing 6.18) that again lists the place names of places of birth and death with links to their instance pages as shown in Figure 6.23. Instead of just showing the count of people who fit the place of birth and death the pop-up lists all of these people with hyperlinks to their instance pages.

The third visualization is a Sampo-UI DeckGL heatmap visualization for the places people have lived in. The coverage for residence places is not very high (54,218 out of 62,207 people have no place of residence listed) and the annotated places heavily favor places in Finland, but the data provides some insight into the major places where people have lived.

The next two visualizations are Sampo-UI Leaflet maps for places of birth and death. The coverage for places of birth has slightly higher coverage than for places of residence (50,123 out of 62,207 people are missing a place of birth) while place of death has a worse coverage (58,464 out of 62,207 people are missing a place of death), so like the heatmap the visualizations do not cover the majority of people but still provide some insights into the origins of people. Places of education were not included in any of these map visualizations due to even lower coverage of less than 2,000 people having any place listed.

6.6.2 Instance pages



Henkilö

Vuorio, Hannu

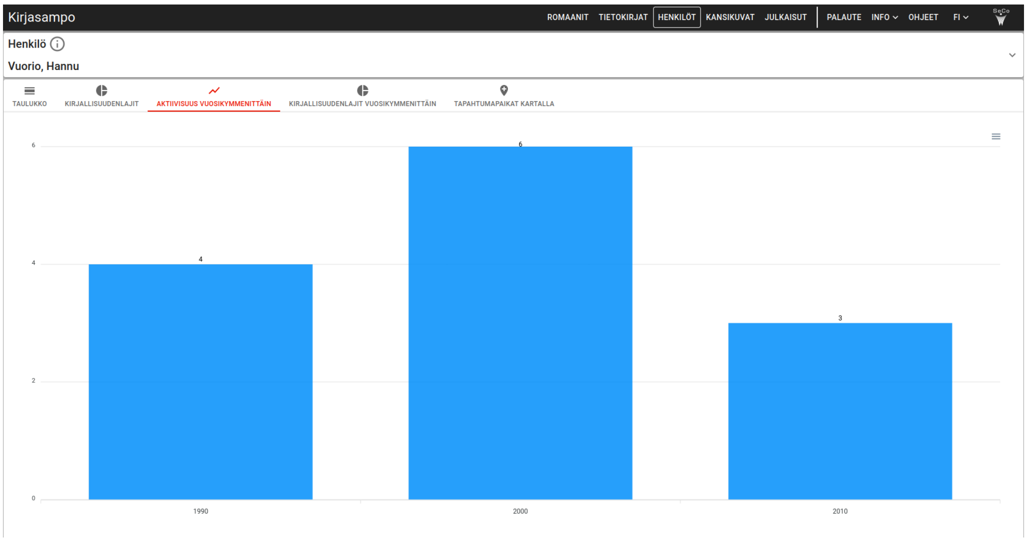
Kuva	
URI	http://www.yso.fi/onto/kaunokki#person_12317592502589
Nimi	Vuorio, Hannu
Muut nimet	-
Ammatti	kirjailijat
Sukupuoli	mies
Kielitaito	http://lexvo.org/id/iso639-3/fin
Kansallisuus	suomalaiset
Aktiivinen vuosina	1990-luku ...
Koulutus	-
Koulutuspaikka	-
Asuinpaikka	Helsinki
Synnyinaika	16.3.1941
Synnyinpaikka	Helsinki
Kuolinaika	-
Kuolinpaikka	-
Asiasana	Kotimaisia dekkarikirjailijoita ...
Palkinnot	Vuoden johtolanka: 1996
Luottamustehtävät	-
Koulukunta	-
Elämänkertatietoa	funktionär och redaktör i Kommunistiska ungdomsrörelsen och i partiet studier vid Helsingfors universitet på 60-talet Helsingin

Figure 6.24: People perspective instance page

The default view in people instance page is the table view as shown in Figure 6.24. The table includes the same information as presented in the faceted search table view with the addition of possible *positions of trust*, *biographical information* string excerpt, hyperlinks to possible other instances of the same person (e.g., under different pseudonyms) and lists of novels, nonfiction books and other works written by the person.

The instance pages for people also have four different visualization tabs in addition to the default table view. The first visualization has a Sampo-UI pie chart visualization for showcasing the genres of the books that the person has written if (s)he is an author.

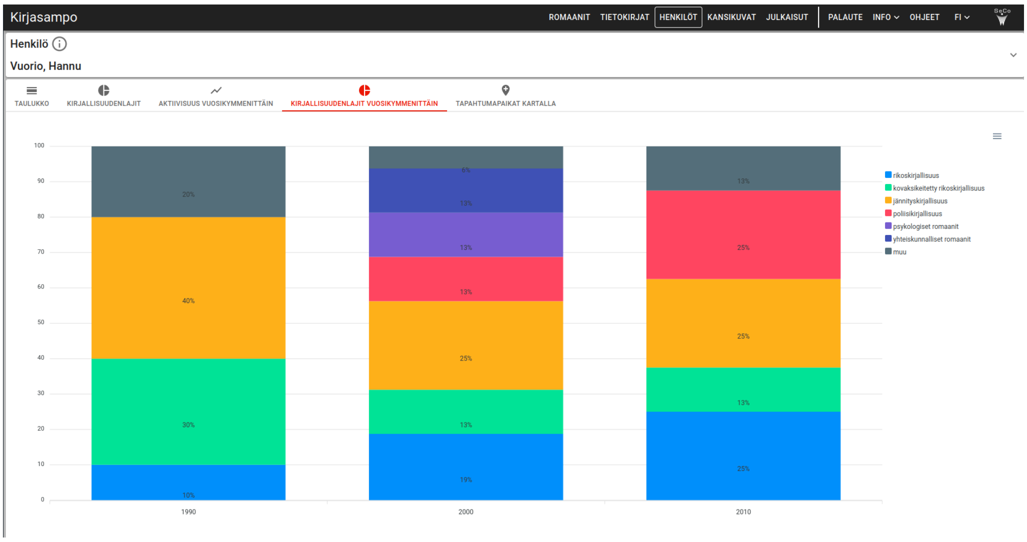
```
"instacePageWorks": {
  "tabID": 2,
  "component": "ApexCharts",
  "doNotRenderOnMount": true,
  "tabPath": "activity_chart",
  "tabIcon": "ShowChart",
  "facetClass": "people",
  "sparqlQuery": "worksByDecadeQuery",
  "resultMapper": "mapPieChart",
  "filterTarget": "work",
  "createChartData": "createApexBarChartData",
```



Henkilö

Vuorio, Hannu

Figure 6.25: Column chart of a person's activity throughout decades



Henkilö

Vuorio, Hannu

Figure 6.26: Stacked column chart of genres written by a person

```
  "pageType": "instancePage",
  "sliceVisibilityThreshold": -1,
  "dropdownForResultClasses": false,
  "resultMapperConfig": {
    "fillEmptyValues": true,
    "emptyValueInterval": 10
  }
}
```

Listing 6.19: Example configuration for a column chart

The next two visualization tabs have custom column chart components for showing the person's activity throughout decades. These two components are both built with the ApexCharts library. The first of these (shown in Figure 6.25, configuration in Listing 6.19) shows the amount of works written during each of the decades they have been active in. Decades with no works written in them are filled in with zeroes if they are between active decades.

```
"instacePageGenresByDecade": {
  "tabID": 3,
  "component": "ApexCharts",
  "doNotRenderOnMount": true,
  "tabPath": "genre_decade_chart",
  "tabIcon": "PieChart",
  "facetClass": "people",
  "sparqlQuery": "novelGenresByDecadeQuery",
  "filterTarget": "novel",
  "resultMapper": "mapStackedColumnChart",
  "sliceVisibilityThreshold": -1,
  "createChartData": "createStackedColumnChartData",
  "pageType": "instancePage",
  "dropdownForResultClasses": false,
  "resultMapperConfig": {
    "fillEmptyValues": true,
    "emptyValueInterval": 10
  }
}
```

Listing 6.20: Example configuration for a 100% stacked column chart

The second column chart visualization (shown in Figure 6.26, configuration in Listing 6.20) has 100% stacked columns for showing the ratio of genres of the works written in those decades. Empty decades are handled the same way as in the last visualization.

The last visualization tab has a Sampo-UI Leaflet map for showing the concrete settings of the novels the person has written as shown in Figure 6.27. Clicking on a location node with more than one novel taking place there expands the node to show each novel individually. Clicking on one of these individual nodes brings up a small custom tooltip which shows the cover image of the novel as well as the name of the novel with a hyperlink to that novel's instance page.

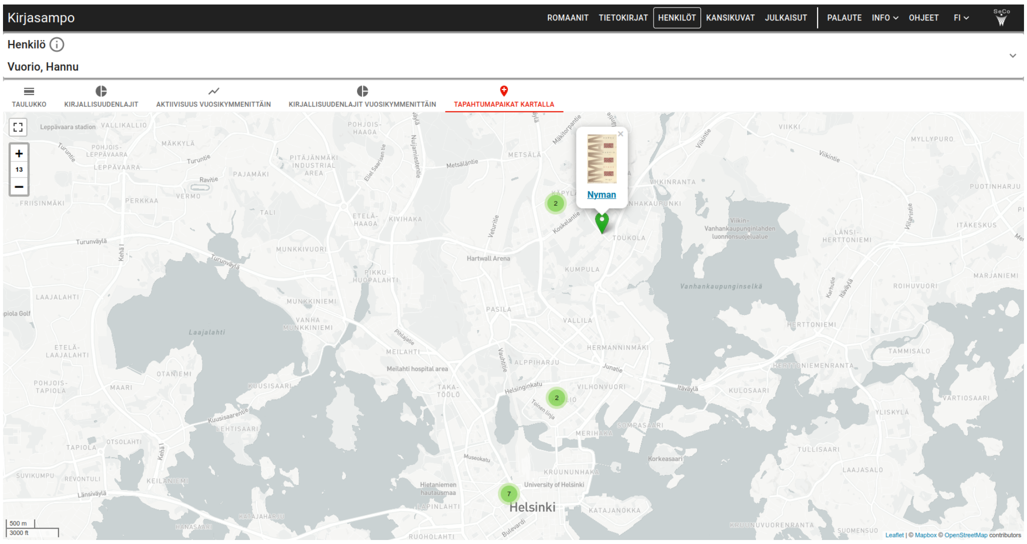


Figure 6.27: Custom Leaflet map tooltip component for a person's novels taking place at a setting

6.7 Instance page only perspectives

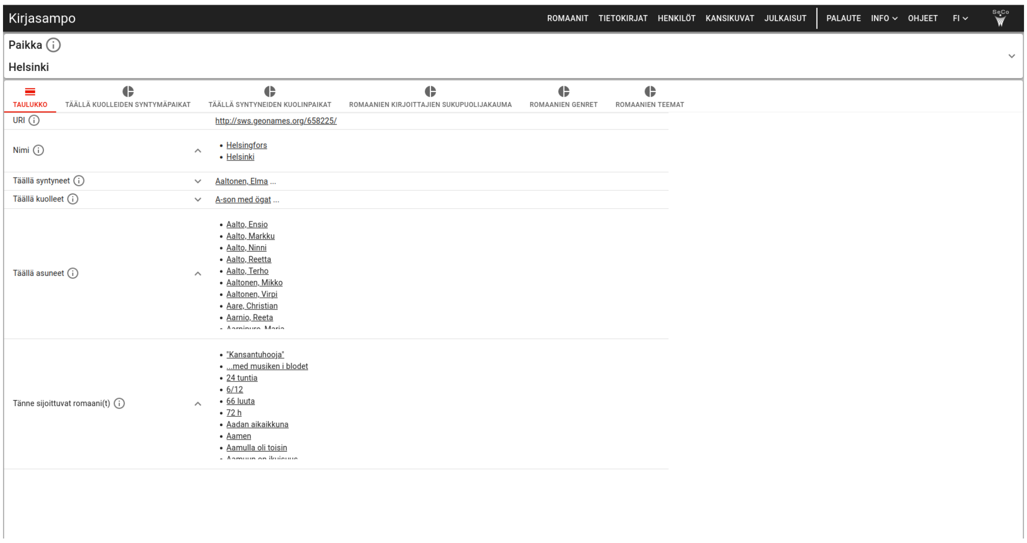
This section covers the three perspectives with only instance pages and no faceted search view. The only way to reach these pages is through hyperlinks from other perspectives; the pages cannot be directly searched.

6.7.1 Places

The instance page (shown in Figure 6.28) includes some basic information regarding the people and novels associated with the place: people who were born, lived, or died at the current place as well as all novels with their setting marked as that place.

Place instance pages also have visualizations associated them in addition to the basic instance page table. All these visualizations use Sampo-UI's ready-to-use pie chart component. The first two visualizations are pie charts of places of birth of people who died at the specified place and places of death of people who were born at the specified place, respectively.

The last three visualizations specifically deal with novels connected to the place through their concrete setting. The first of the three is a gender ratio visualization of the authors' genders for novels with the specified place as the setting. The next two visualizations visualize the top novel genres and



Paikka

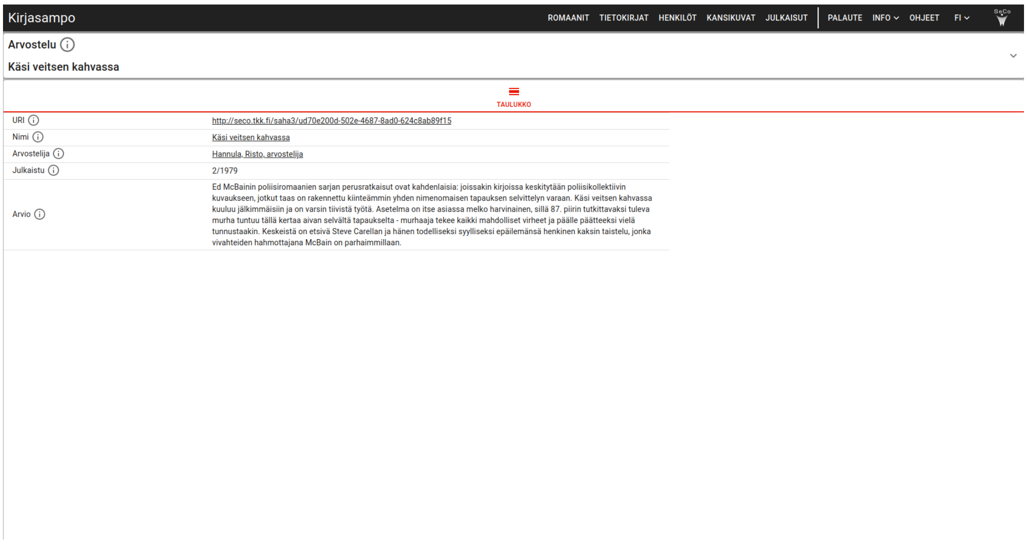
Helsinki

URI	http://sws.geonames.org/658225/
Nimi	Helsingfors Helsinki
Täällä syntyneet	Aaltonen, Elma ...
Täällä kuolleet	A-son med ögat ...
Täällä asuneet	Aalto, Ensio Aalto, Markku Aalto, Ninni Aalto, Reetta Aalto, Terho Aaltonen, Mikko Aaltonen, Virpi Aare, Christian Aarnio, Reeta Aarnipuro, Maria
Tänne sijoittuvat romaani(t)	"Kansantuhooja" ...med musiken i blodet 24 tuntia 6/12 66 luuta 72 h Aadan aikaikkuna Aamen Aamulla oli toisin Aamuun on ikuisuus

Figure 6.28: Places perspective instance page

themes for novels taking place in this setting.

6.7.2 Reviews



Arvostelu

Käsi veitsen kahvassa

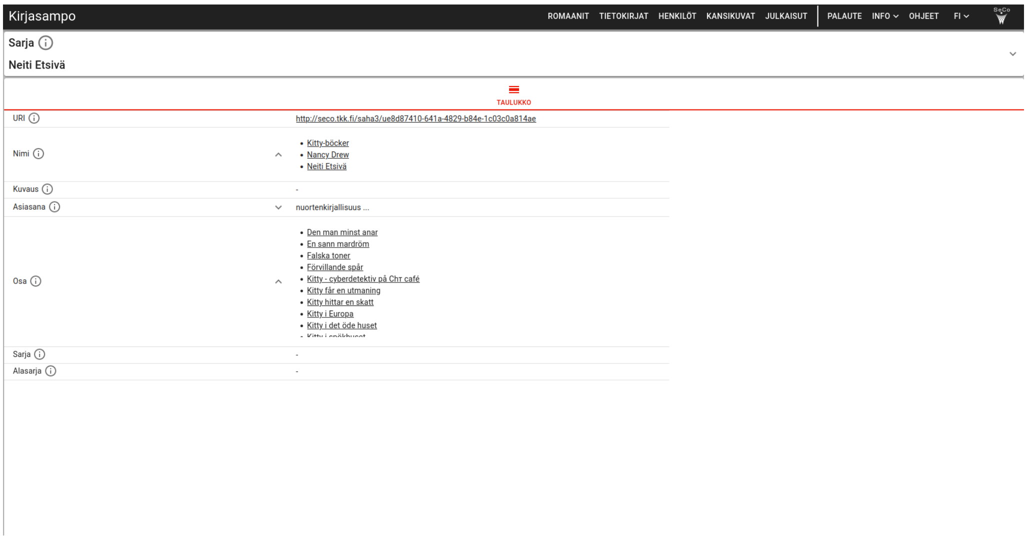
URI	http://seco.tkk.fi/saha3/ud70e200d-502e-4687-8ad0-624c8ab89f15
Nimi	Käsi veitsen kahvassa
Arvostelija	Hannula, Risto, arvostelija
Julkaistu	2/1979
Arvio	Ed McBainin poliisiromaanien sarjan perusratkaisut ovat kahdenlaisia: joissakin kirjoissa keskitytään poliisikollektiivin kuvaukseen, jotkut taas on rakennettu kiinteämmin yhden nimenomaisen tapauksen selvittelyn varaan. Käsi veitsen kahvassa kuuluu jälkimmäisiin ja on varsin tiivistä työtä. Asetelma on itse asiassa melko harvinainen, sillä 87. piirin tutkittavaksi tuleva murha tuntuu tällä kertaa aivan selvältä tapaukselta - murhaaja tekee kaikki mahdolliset virheet ja päälle päätteeksi vielä tunnustaakin. Keskeistä on etsivä Steve Carellan ja hänen todelliseksi syylliseksi epäilemänsä henkinen kaksin taistelu, jonka vivahteiden hahmottajana McBain on parhaimmillaan.

Figure 6.29: Reviews perspective instance page

The instance page for a review is shown in Figure 6.29. Reviews in the

BookSampo data were very scarcely annotated, so in addition to the URI and the name of the review, only the *reviewer*, *publication time* and the *review* itself are included. The reviewer field includes a hyperlink to the person's own instance page.

6.7.3 Series



Sarja

Neiti Etsivä

URI	http://seco.tkk.fi/saha3/ue8d87410-641a-4829-b84e-1c03c0a814ae
Nimi	Kitty-böcker Nancy Drew Neiti Etsivä
Kuvaus	-
Asiasana	nuortenkirjallisuus ...
Osa	Den man minst anar En sann mardröm Falska toner Förvillande spår Kitty - cyberdetektiv på Chт café Kitty får en utmaning Kitty hittar en skatt Kitty i Europa Kitty i det öde huset Kitty i spökhuset
Sarja	-
Alasarja	-

Figure 6.30: Series perspective instance page

The instance page of a series is shown in Figure 6.30. In addition to the basic information of the series resource itself (URI, preferred label), possible *description* and *keywords* as well as the list of publications in the series are listed with hyperlinks to the publication instance pages. Since series can be a part of another series in the data, *series* and *subseries* fields are included for the current series' super- and subseries.

Chapter 7

Discussion

This chapter first examines the results of the master's thesis through the research questions posed in Chapter 1. After that, the chapter discusses the possibilities for the future in regard to the developed semantic portal as well as the BookSampo dataset.

7.1 Results

The goal of this master's thesis was to build a tool for using the BookSampo data for information retrieval and research purposes by building a new semantic portal user interface for it and to answer the research questions posed back in Chapter 1.

Based on the implementation of the new user interface the research questions posed at the start of the thesis were met with following answers:

1. *How can the user utilize BookSampo data for intelligent information retrieval and research?*

    (a) *How should the BookSampo knowledge graph be visualized and what kind of visualizations would be the best for these purposes?*

    Time-based visualizations like the time series charts and column charts as well as visualizations for showing most common property values like pie and bar charts were included to account for the basic visualizations relevant to bibliographic data that were presented in articles, such as the articles of Klink et al. [39] and Börner et al. [12]. Since the BookSampo data also has location information for things like settings and events in people's lives, map-based visualizations were also included.

The crucial factor for visualizations was that they should be able to be affected by applying different facets in the faceted search view. Outside of the bar chart race visualization, which offers no way of filtering the data set used, all other visualizations aim to be as customizable with the different facets as possible to let the user have the freedom to limit what kind of data they want to visualize.

The choice of properties that can be used as the basis for visualizations was made based on the annotation coverage of those properties. For components like pie charts the goal was that the 'Unknown' value for the property should not completely dominate the chart, even if it is the top value due to lower coverage. Location related annotation coverage tended to be on the lower side but was included in cases where even the smaller data set could provide some insights into Finnish literature, such as the concrete settings for novels as well as places related to events in people's lives. Even in these cases some location facets with extremely low annotation coverage, such as the place of education for people were not turned into visualizations.

(b) *What kind of searches/information retrieval should the user be able to do using BookSampo data and the developed user interface?*

The initial perspectives for the portal were chosen based on providing a comprehensive coverage for all different aspects in the data: the works on an abstract level (novels and nonfiction books perspectives) and a physical level (publications and covers perspectives) as well as the people in the literature field (people perspective). Novels and nonfiction books were chosen as the initial perspectives for the abstract level to cover a wide variety of both fiction and nonfiction books in the portal.

The principle behind the choosing of facets and data included in the perspectives itself was to provide the user as much freedom in their filtering as possible. While the visualizations discussed in the last question 1.a prioritized visualizing things with high annotation coverage, the facets in the perspectives include things with lower coverage as well to provide the user with more tools for searching. For facets with values that likely could have some values to be annotated there, that is something that likely exists for all objects (e.g., characters for novels in comparison to things like awards which only exist for novels which have won something), the facet shows the instance count for objects with nothing annotated in the field as 'Unknown'. This way the user can quickly somewhat evaluate how high or low the annotation coverage is and how reliable the results become if that facet is used.

As for the data shown to the user in the faceted search table view, every property included in the facets is included as columns as well. The columns may include some information not covered by facets (e.g., cover images in the novels perspective), but the opposite does not hold. In order to avoid the table from becoming too wide and cumbersome to navigate for the user, additional information not deemed relevant enough to be shown in the table view is shown in the instance pages of the entities. These may include things like string excerpts with biographical information that isn't something you can search using the facet menu and would just take up too much space to be included in one of the columns otherwise. In principle, the instance pages for objects include all properties and their values that have at least some annotation coverage in the data.

2. *How to configure a new semantic portal using the Sampo-UI framework and a knowledge graph?*

   As long as the knowledge graph can easily be stored at a reachable endpoint and the data structure itself is workable, setting up a new semantic portal using the Sampo-UI framework is extremely fast. Sampo-UI offers tons of ready-to-use tools that can be easily configured for different types of data through just the configuration and the actual query files alone. A whole semantic portal could be built with just these ready-to-use tools and components without needing web component development experience. However, with the ready-to-use tools and configuration files there is some extra leftover code and configurations left in the files from other portals using Sampo-UI, which complicates the initial set-up process unless the developer has access to either documentation outlining the creation of a new portal or a person who has previously used the framework for the purpose.

   If one wants to create more customized ways of viewing the data, Sampo-UI framework supports this. Developed custom components can be configured to be usable the same way the ready-to-use components are through the configuration files alone. Due to this Sampo-UI framework offers a good way of building a simple semantic portal that can later be expanded and customized further based on possible new arising needs instead of being constrained to just what the base framework offers.

In this case of the new BookSampo portal, this lead to the addition of several custom components—both new as well as expanded and altered versions of already existing components. Novel and nonfiction book instance pages were complemented with a new component consisting of multiple tables in a list (see Figure 6.10). This made it possible to showcase publication-specific information for each publication separately in the same tab. Map components received custom tooltips as well as a new DeckGL map layer option for showing scatterplot maps with variable circle colors (see Figure 6.8). The components based on the ApexCharts library were expanded as well. Two types of custom column components (see Figures 6.25 and 6.26) were developed for the instance pages in the people perspective. In addition, new mappers and data processing functions had to be developed for ApexCharts-based time series components (see Figures 6.14–6.17) to account for the dynamic nature of the data being visualized in the publications perspective.

3. *How to deal with problematic (e.g., missing labels, hierarchy) and/or incomplete data when developing portals like these?*

   Instead of trying to hide the problematic or incomplete data, showing the problems openly in the new user interface helped better visualize the problems. Especially if the data is managed by another company or institution, being able to directly show both the problems itself and their ramifications in terms of the implementation makes it easier to make the point that data cleaning or fixing is necessary and/or worthwhile. In addition, seeing the severity of each problem helps prioritize the order for the possible future fixes.

   For problems that made it impossible to directly showcase their problems, some workarounds were employed to show what kind of things could be done with the data if corrected. Though these workarounds are functional, their performance (e.g., query time) clearly suffers from it. To guarantee the best possible user experience these workarounds should eventually be eliminated and therefore still work as motivators towards fixing the data.

4. *What is the quality of the BookSampo data?*

While the new user interface uncovered problems with the data, the overall quality of data is good considering the sheer size and age of the dataset. The major uncovered problems were due to missing or incorrect values, but not with the structure or properties of the data model itself, and can thus easily be solved without needing to overhaul anything. In addition, none of the problems like the low label coverage or missing hierarchy actually prevented the data from being used in the new portal with just some time-based visualizations requiring slight workarounds for grouping the data.

A lot of the label issues would likely be able to be automatically corrected by sampling either preferred or other labels that either have language tags other than Finnish or ones without any language tags specified at all. These changes would require some slight modifications to the new implemented portal's configuration and queries before improvement could be seen on the user interface itself. The problems requiring manual correction (e.g., incorrect concrete setting mappings) are fortunately not a systematic problem and, as they did not directly affect the implementation of the user interface, correcting the data in this aspect would not require any further changes to the user interface itself.

7.2 Future work

At the time of writing this thesis the initial work for cleaning up and correcting some of the BookSampo data has been started. This offers up the possibility of cleaning up and uniformizing the user interface in the future if the issues with labels and hierarchy are fixed. Depending on the extent of this work on the data, the visualizations in the user interface could be expanded upon by utilizing the possibilities that could open up from the addition of hierarchy to things like time resources and location resources. When this cleaning up work is done, the portal can be officially opened to the public.

The future publication of the new BookSampo portal offers up the possibility of gathering feedback from users on the usability of the service as well as what could be improved in the future. This feedback from external users could be used to properly evaluate the portal in both information retrieval and research use contexts. Previous evaluations of other Sampo-UI-based UIs as well as of the search paradigms behind the Sampo Model suggest good usability and scalability of the UI for the end-user [14, 17, 38]. The current performance of the queries is fast enough for the purposes of the portal, though some of the more time-intensive queries could be sped up with improvements to the underlying data as well as by adding additional links between entities to allow for shorter predicate paths in queries. At the moment of writing this, it is still too early to say how well-suited the portal will be for literary research purposes. However, research utilizing the BookSampo data has been started and some initial results have already been obtained, which is promising for the future [49].

The current version of the portal covers the five perspectives presented in this thesis. With time the portal could easily be expanded to include other types of works present in the BookSampo data, such as short stories and poems. The data available could also possibly lend itself to new kinds of visualizations based on the needs, wants and ideas of the users using the interface.

Lastly, the data could possibly be enriched with data from other cultural heritage sources such as other Sampo portals. In addition to enriching the data within the portal itself, the portal could provide links to other portals or sources of information available where the user could seek more information on the object present in the data outside the scope of the BookSampo data.

Chapter 8

Conclusions

This thesis went over the design and implementation of a new semantic portal using BookSampo data while also assessing some of the problems arising from using already existing data for portals like these.

The thesis first briefly covered some background on Semantic Web and Linked Data in both general and bibliographical contexts. Afterwards the tools and materials used for this thesis, the Sampo-UI framework and the BookSampo data respectively, were introduced. The section for the data also covered quality assessment on the data and what issues it has.

The latter part of the thesis covered the design and implementation of the new portal itself. The design section covered the basic design of what the portal should include and how the different perspectives should be split. The implementation section introduced the actual implementation of the portal and discussed some of the choices made during the development and the reasoning behind them.

The current implementation offers the user various ways of filtering and searching the data for both information retrieval as well as research purposes. The user can both look for works fulfilling specific criteria as well as visualize the whole result set to see the most typical features of Finnish literature in general or even of a very specific subset. The visualizations offered can also be used to visualize how Finnish literature has evolved throughout the years and what kind of trends exist through annotation trends present in the data.

Though the current implementation is done in the scope of this thesis work, the implementation offers ways of expanding it further in the future. Setting up new components is extremely easy due to the configurable nature of the Sampo-UI framework behind the implementation. Visualizations can easily be added based on user needs and the coverage of the data set can be increased with further perspectives covering other types of works not included in the initial perspective selection.

The importance of cleaning up and fixing the original data set cannot be understated as well. While the overall quality of the data is good when taking into account the age and sheer volume of the data, there were problems uncovered during the implementation of the new portal that do negatively affect the user experience if not fixed. Fixing incomplete data like missing hierarchy links also opens up the possibility of new kinds of visualizations and search possibilities to complement the existing ones. So while the scope of work in thesis is now finished, the BookSampo data set still offers a lot of further possibilities for developing tools for both research and information retrieval purposes.

Bibliography

[1] Aitonurmi, T. Kirjasampo-palvelun kehitys ja uudet ominaisuudet. *Informaatiotutkimus 34*, 1-2 (2015).

[2] Aitonurmi, T. Kirjasammossa yli 1,6 miljoonaa käyntiä 2021. *Kirjasampo* (2021). Available at: `https://www.kirjasampo.fi/fi/kirjasammon-tilastot-2021` (Accessed 3.1.2023).

[3] Alemu, G., Stevens, B., Ross, P., and Chandler, J. Linked data for libraries: benefits of a conceptual shift from library-specific record structures to rdf-based data models. *New library world 113*, 11/12 (2012), 549–570.

[4] Antoniou, G., Franconi, E., and Harmelen, F. v. Introduction to semantic web ontology languages. In *Reasoning web.* Springer, 2005, pp. 1–21.

[5] Baker, T., Bermès, E., Coyle, K., Dunsire, G., Isaac, A., Murray, P., Panzer, M., Schneider, J., Singer, R., Summers, E., et al. Library linked data incubator group final report, 2011.

[6] Beckett, D. RDF 1.1 N-Triples. Tech. rep., W3C, 2001.

[7] Beckett, D., and Berners-Lee, T. Turtle - terse RDF triple language, W3C team submission. Tech. rep., W3C, 2008.

[8] Berners-Lee, T. Linked Data - Design Issues. *W3C*, 09/20 (2006).

[9] Berners-Lee, T., and Connolly, D. Notation3 (N3): A readable RDF syntax. Tech. rep., W3C, 2008.

[10] Berners-Lee, T., and Fischetti, M. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*, 1st ed. Harper San Francisco, 1999.

[11] Berners-Lee, T., Hendler, J., and Lassila, O. The semantic web. *Scientific american 284*, 5 (2001), 34–43.

[12] Börner, K., Chen, C., and Boyack, K. W. Visualizing knowledge domains. *Annual review of information science and technology 37*, 1 (2003), 179–255.

[13] Brickley, D., and Guha, R. RDF Schema 1.1 - W3C Recommendation 25 February 2014. Tech. rep., W3C, 2014.

[14] Burrows, T., Pinto, N. B., Cazals, M., Gaudin, A., and Wijsman, H. Evaluating a semantic portal for the “Mapping Manuscript Migrations” project. *DigItalia 15*, 2 (2020), 178–185.

[15] Cardoso, J. The semantic web vision: Where are we? *IEEE Intelligent systems 22*, 5 (2007), 84–88.

[16] Charles, V., Freire, N., and Isaac, A. Links, languages and semantics: linked data approaches in The European Library and Europeana. *IFLA, Lyon* (2014).

[17] English, J., Hearst, M., Sinha, R., Swearingen, K., and Lee, K. Flexible search and navigation using faceted metadata. Tech. rep., Technical report, University of Berkeley, School of Information Management . . . , 2002.

[18] Feigenbaum, L., Herman, I., Hongsermeier, T., Neumann, E., and Stephens, S. The semantic web in action. *Scientific American 297*, 6 (2007), 90–97.

[19] Fernández, J. D., and Martínez-Prieto, M. A. *RDF Serialization and Archival.* Springer International Publishing, 2018, pp. 1–11.

[20] Frosterus, M., Tuominen, J., Pessala, S., Seppälä, K., and Hyvönen, E. Linked Open Ontology Cloud KOKO—managing a system of cross-domain lightweight ontologies. In *Extended Semantic Web Conference* (2013), Springer, pp. 296–297.

[21] Gómez-Pérez, A., and Corcho, O. Ontology languages for the semantic web. *IEEE Intelligent systems 17*, 1 (2002), 54–60.

[22] Hallo, M., Luján-Mora, S., Maté, A., and Trujillo, J. Current state of linked data in digital libraries. *Journal of Information Science 42*, 2 (2016), 117–127.

[23] Hannemann, J., and Kett, J. Linked data for libraries. In *Proc of the world library and information congress of the Int'l Federation of Library Associations and Institutions (IFLA)* (2010).

[24] Haslhofer, B., and Isaac, A. data. europeana. eu: The europeana linked open data pilot. In *International Conference on Dublin Core and Metadata Applications* (2011), pp. 94–104.

[25] Hitzler, P. A review of the semantic web field. *Communications of the ACM 64*, 2 (2021), 76–83.

[26] Hogan, A. The semantic web: Two decades on. *Semantic Web 11*, 1 (2020), 169–185.

[27] Horrocks, I., Patel-Schneider, P. F., and Van Harmelen, F. From SHIQ and RDF to OWL: The making of a web ontology language. *Journal of web semantics 1*, 1 (2003), 7–26.

[28] Hypén, K. Kirjasampo: rethinking metadata. *Cataloging & classification quarterly 52*, 2 (2014), 156–180.

[29] Hyvönen, E. FinnONTO—Building the Basis for a National Semantic Web Infrastructure in Finland. In *Proceedings of the 12th Finnish AI Conference STeP* (2006), vol. 6.

[30] Hyvönen, E. Cultural Heritage Linked Data on the Semantic Web: Three Case Studies Using the Sampo Model. *VIII Encounter of documentation centres of contemporary art: open linked data and integral management of information in cultural centres. Artium, Vitoria-Gasteiz, Spain, October* (2016), 19–20.

[31] Hyvönen, E. Building and Using a National Linked Open Data Infrastructure for Digital Humanities: The Finnish Approach. In *Proceedings: Data for History 2020. Modelling Time, Places, Agents.* Humboldt-Universität zu Berlin, 2020.

[32] Hyvönen, E. Digital Humanities on the Semantic Web: Sampo Model and Portal Series. In *Semantic Web journal* (2022), IOS PRESS.

[33] Hyvönen, E., et al. Linked open data infrastructure for digital humanities in Finland. In *DHN 2020 Digital Humanities in the Nordic Countries. Proceedings of the Digital Humanities in the Nordic Countries 5th Conference* (2020), CEUR-WS.org.

[34] Hyvönen, E., et al. Sammon taontaa semanttisessa webissä (Forging Sampos on the Semantic Web). *Tekniikan Waiheita* (2021).

[35] 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 *European Semantic Web Conference* (2014), Springer, pp. 226–230.

[36] Hyvönen, E., Viljanen, K., Tuominen, J., and Seppälä, K. Building a national semantic web ontology and ontology service infrastructure–the FinnONTO approach. In *European Semantic Web Conference* (2008), Springer, pp. 95–109.

[37] Hyvönen, E. "Sampo" Model and Semantic Portals for Digital Humanities on the Semantic Web. In *Proceedings of the Digital Humanities in the Nordic Countries 5th Conference (DHN 2020)* (Germany, 2020), no. 2612 in CEUR workshop proceedings, CEUR-WS.org, pp. 373–378.

[38] Ikkala, E., Hyvönen, E., Rantala, H., and Koho, M. Sampo-UI: A Full Stack JavaScript Framework for Developing Semantic Portal User Interfaces. *Semantic Web - Interoperability, Usability, Applicability Volume 13, issue 1* (2022), 16.

[39] Klink, S., Ley, M., Rabbidge, E., Reuther, P., Walter, B., and Weber, A. Browsing and Visualizing Digital Bibliographic Data. In *VisSym* (2004), vol. 2004, pp. 237–242.

[40] Lassila, O., and Swick, R. R. Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendation, W3C, 1999.

[41] Mäkelä, E., Hypén, K., and Hyvönen, E. BookSampo—lessons learned in creating a semantic portal for fiction literature. In *International Semantic Web Conference* (2011), Springer, pp. 173–188.

[42] Mäkelä, E., Hypén, K., and Hyvönen, E. Improving fiction literature access by linked open data-based collaborative knowledge storage-the BookSampo project. In *78th IFLA General Conference and Assembly, Helsinki* (2012).

[43] Mäkelä, E., Hypén, K., and Hyvönen, E. Fiction literature as linked open data—The BookSampo dataset. *Semantic Web 4*, 3 (2013), 299–306.

[44] Maniraj, V., and Sivakumar, R. Ontology languages-a review. *International Journal of Computer Theory and Engineering 2*, 6 (2010), 887.

[45] Miles, A., and Bechhofer, S. SKOS Simple Knowledge Organization System Reference. *W3C recommendation* (2009).

[46] Mäkinen, E. Näin keräät kirjaasi kaikki kliseet. *Helsingin Sanomat* (2011). Available at: `https://www.hs.fi/kulttuuri/art-2000004809978.html` (Accessed 2.1.2023).

[47] Mäkinen, E. Romaanit pidentyneet. *Helsingin Sanomat* (2011). Available at: `https://www.hs.fi/kulttuuri/art-2000004809975.html` (Accessed 2.1.2023).

[48] Mäkinen, E. Yhden kirjan kirjoittajien määrä kasvaa. *Helsingin Sanomat* (2011). Available at: `https://www.hs.fi/kulttuuri/art-2000004809977.html` (Accessed 2.1.2023).

[49] Peura, T. Suomeksi yli rajojen. Kvantitatiivinen tutkimus suomenkielisten romaanien monimuotoisuudesta 1970-2020. Master's thesis, University of Helsinki, Department of Digital Humanities, Helsinki Centre for Digital Humanities (HELDIG), January 2023. Forthcoming.

[50] Prud'hommeaux, E., Harris, S., and Seaborne, A. SPARQL 1.1 Query Language. Tech. rep., W3C, 2013.

[51] Prud'hommeaux, E., and Seaborne, A. SPARQL Query Language for RDF. Tech. rep., W3C, 2008.

[52] Raza, Z., Mahmood, K., and Warraich, N. F. Application of linked data technologies in digital libraries: a review of literature. *Library Hi Tech News* (2019).

[53] Riva, P., Doerr, M., and Zumer, M. FRBRoo: enabling a common view of information from memory institutions. In *World Library and Information Congress: 74th IFLA General Confrence and Council* (2008).

[54] Riva, P., Le Boeuf, P., Žumer, M., et al. *IFLA library reference model: A conceptual model for bibliographic information.* International Federation of Library Associations and Institutions (IFLA), 2018.

[55] Shadbolt, N., Berners-Lee, T., and Hall, W. The semantic web revisited. *IEEE intelligent systems 21*, 3 (2006), 96–101.

[56] Shahzad, K., and Khan, S. A. Factors affecting the adoption of integrated semantic digital libraries (sdls): A systematic review. *Library Hi Tech*, ahead-of-print (2022).

[57] Smith-Yoshimura, K. Analysis of 2018 international linked data survey for implementers. *Code4Lib journal*, 42 (2018).

[58] Sporny, M., Longley, D., Kellogg, G., Lanthaler, M., and Lindström, N. JSON-LD 1.0. *W3C recommendation 16* (2014), 41.

[59] 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 *ISWC (Industry Track)* (2014).

[60] Sure, Y., and Studer, R. Semantic Web technologies for digital libraries. *Library Management* (2005).

[61] Viljanen, K., Tuominen, J., and Hyvönen, E. Ontology libraries for production use: The Finnish ontology library service ONKI. In *European Semantic Web Conference* (2009), Springer, pp. 781–795.

[62] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data 3*, 1 (2016), 1–9.

[63] Žumer, M. IFLA library reference model (IFLA LRM)—harmonisation of the FRBR family. *KO Knowledge Organization 45*, 4 (2018), 310–318.

Appendix A

Example portal configuration file

Below is an example Sampo-UI framework portal configuration file. The file describes a portal with the ID booksampo that has eight perspectives, five full search perspectives and three perspectives with only instance pages. The perspective has two locales, Finnish and English, with Finnish being the default locale.

```
{
    "portalID": "booksampo",
    "rootUrl": "",
    "perspectives": {
        "searchPerspectives": [
            "novels",
            "nonfictionBooks",
            "people",
            "covers",
            "publications"
        ],
        "onlyInstancePages": [
            "places",
            "series",
            "reviews"
        ]
    },
    "localeConfig": {
        "defaultLocale": "fi",
        "readTranslationsFromGoogleSheets": false,
        "availableLocales": [
            {
                "id": "en",
                "label": "English",
                "filename": "localeEN.json"
            },
            {
                "id": "fi",
                "label": "Finnish",
                "filename": "localeFI.json"
            }
        ]
    },
    "sitemapConfig": {
        "baseUrl": "https://sampo-ui.demo.seco.cs.aalto.fi",
        "langPrimary": "en",
        "langSecondary": "fi",
        "outputDir": "./src/server/sitemap_generator",
        "sitemapUrl": "https://sampo-ui.demo.seco.cs.aalto.fi/sitemap",
        "sitemapInstancePageQuery": "sitemapInstancePageQuery"
    },
    "knowledgeGraphMetadataConfig": {
        "showTable": false,
        "perspective": "novels"
    },
    "layoutConfig": {
        "colorPalette": {
            "primary": {
                "main": "#212121"
            },
            "secondary": {
                "main": "#EB1806"
            }
        },
        "hundredPercentHeightBreakPoint": 900,
        "reducedHeightBreakpoint": 1920,
        "tabHeight": 58,
        "paginationToolbarHeight": 37,
        "tableFontSize": "0.8rem",
        "topBar": {
            "logoTextTransform": "none",
            "hideLogoTextOnMobile": true,
            "showLanguageButton": true,
            "showSearchField": false,
            "feedbackLink": "https://link.webropolsurveys.com/...",
            "externalInstructions": false,
            "externalAboutPage": false,
            "reducedHeight": 48,
            "defaultHeight": 64,
            "mobileMenuBreakpoint": 1360,
            "infoDropdown": [
                {
                    "id": "about",
                    "translatedText": "aboutThePortal",
                    "internalLink": "/about"
                },
                {
                    "id": "blog",
                    "externalLink": true,
                    "translatedUrl": "blogUrl",
                    "translatedText": "blog"
                }
            ]
        },
        "mainPage": {
            "bannerImage": "main_page/mmm-banner.jpg",
            "bannerBackround": "linear-gradient( rgba(0, 0, 0, 0.45),
                                rgba(0, 0, 0, 0.45) ),
                                url(<BANNER_IMAGE_URL)",
            "bannerMobileHeight": 150,
            "bannerReducedHeight": 220,
            "bannerDefaultHeight": 300,
            "wrapSubheading": true
        },
        "infoHeader": {
            "default": {
                "height": 49,
                "expandedContentHeight": 160,
                "headingVariant": "h4",
                "infoIconFontSize": 40
            },
            "reducedHeight": {
                "height": 40,
                "expandedContentHeight": 100,
                "headingVariant": "h6",
                "infoIconFontSize": 32
            }
        },
        "footer": {
            "reducedHeight": 44,
            "defaultHeight": 64
        }
    },
    "mapboxConfig": {
        "mapboxStyle": "light-v10"
    },
    "yasguiConfig": {
        "yasguiBaseURL": "https://yasgui.triply.cc",
        "yasguiParams": {
            "contentTypeConstruct": "text/turtle",
            "contentTypeSelect": "application/sparql-results+json",
            "endpoint": "https://ldf.fi/booksampo-2022/sparql",
            "requestMethod": "POST",
            "tabTitle": "Exported query"
        }
    },
    "documentFinderConfig": {
        "apiURL": "https://data.finlex.fi/document-finder-backend"
    }
}
```

Appendix B

Example perspective configuration file

Below is an example Sampo-UI framework perspective configuration file. The file describes a perspective with the ID `covers`. The perspective has a results table view (`paginatedResultsConfig`) as well as a visualization tab including a pie/bar chart (`coversByProperty`). The objects present in the perspective are queried for five different properties, *image*, *URI*, *preferred label*, *keyword*, and *work type*. There are three facets present in the perspective for the preferred labels, keywords, and work types.

```
{
    "id": "covers",
    "endpoint": {
        "url": "https://ldf.fi/booksampo-2022/sparql",
        "useAuth": true,
        "prefixesFile": "SparqlQueriesPrefixes.js"
    },
    "sparqlQueriesFile": "SparqlQueriesCovers.js",
    "facetClass": "kaunokki:kansi",
    "langTag": "fi",
    "frontPageImage": "main_page/works-452x262.jpg",
    "searchMode": "faceted-search",
    "defaultActiveFacets": [
        "prefLabel"
    ],
    "defaultTab": "table",
    "defaultInstancePageTab": "table",
    "resultClasses": {
        "covers": {
            "paginatedResultsConfig": {
                "tabID": 0,
                "component": "ResultTable",
                "tabPath": "table",
                "tabIcon": "CalendarViewDay",
                "propertiesQueryBlock": "coverProperties",
                "pagesize": 10,
                "sortBy": null,
                "sortDirection": null,
                "paginatedResultsAlwaysExpandRows": true,
                "paginatedResultsRowContentMaxHeight": 160
            },
            "instanceConfig": {
                "propertiesQueryBlock": "coverProperties",
                "instancePageResultClasses": {
                    "instancePageTable": {
                        "tabID": 0,
                        "component": "InstancePageTable",
                        "tabPath": "table",
                        "tabIcon": "CalendarViewDay"
                    }
                },
                "localIDAsURI": true
            }
        },
        "coversByProperty": {
            "tabID": 1,
            "component": "ApexCharts",
            "doNotRenderOnMount": true,
            "tabPath": "pie_chart",
            "tabIcon": "PieChart",
            "facetClass": "covers",
            "dropdownForResultClasses": true,
            "defaultResultClass": "coversByKeyword",
            "resultClasses": {
                "coversByKeyword": {
                    "sparqlQuery": "coversByKeywordQuery",
                    "filterTarget": "cover",
                    "resultMapper": "mapPieChart",
                    "sliceVisibilityThreshold": 0.01,
                    "dropdownForChartTypes": true,
                    "resultMapperConfig": {
                        "fillEmptyValues": false
                    },
                    "chartTypes": [
                        {
                            "id": "pie",
                            "createChartData": "createApexPieChartData"
                        },
                        {
                            "id": "bar",
                            "createChartData": "createApexBarChartData"
                        }
                    ]
                },
                "coversByWorkType": {
                    "sparqlQuery": "coversByWorkTypeQuery",
                    "filterTarget": "cover",
                    "resultMapper": "mapPieChart",
                    "sliceVisibilityThreshold": 0.01,
                    "dropdownForChartTypes": true,
                    "resultMapperConfig": {
                        "fillEmptyValues": false
                    },
                    "chartTypes": [
                        {
                            "id": "pie",
                            "createChartData": "createApexPieChartData"
                        },
                        {
                            "id": "bar",
                            "createChartData": "createApexBarChartData"
                        }
                    ]
                }
            }
        }
    },
    "properties": [
        {
            "id": "image",
            "valueType": "image",
            "previewImageHeight": 150,
            "makeLink": true,
            "externalLink": true,
            "sortValues": true,
            "numberedList": false,
            "hideHeader": true
        },
        {
            "id": "uri",
            "valueType": "object",
            "makeLink": true,
            "externalLink": true,
            "sortValues": true,
            "numberedList": false,
            "onlyOnInstancePage": true
        },
        {
            "id": "prefLabel",
            "valueType": "object",
            "makeLink": true,
            "externalLink": false,
            "sortValues": true,
            "numberedList": false,
            "minWidth": 200
        },
        {
            "id": "keyword",
            "valueType": "object",
            "makeLink": false,
            "externalLink": false,
            "sortValues": true,
            "numberedList": false,
            "minWidth": 150
        },
        {
            "id": "workType",
            "valueType": "object",
            "makeLink": false,
            "externalLink": false,
            "sortValues": true,
            "numberedList": false,
            "minWidth": 150
        }
    ],
    "facets": {
        "prefLabel": {
            "containerClass": "one",
            "facetType": "text",
            "filterType": "textFilter",
            "sortBy": "prefLabel",
            "sortByPredicate": "skos:prefLabel",
            "textQueryProperty": "skos:prefLabel"
        },
        "keyword": {
            "containerClass": "ten",
            "facetType": "list",
            "facetLabelFilter": "FILTER(LANG(?prefLabel_) = '<LANG>')",
            "filterType": "uriFilter",
            "predicate": "kaunokki:asiasana",
            "searchField": true,
            "sortButton": true,
            "sortBy": "instanceCount",
            "sortByPredicate": "kaunokki:asiasana/skos:prefLabel",
            "sortDirection": "desc"
        },
        "workType": {
            "containerClass": "ten",
            "facetType": "list",
            "facetLabelFilter": "FILTER(LANG(?prefLabel_) = '<LANG>')",
            "filterType": "uriFilter",
            "predicate": "^kaunokki:kansikuva/^kaunokki:manifests_in/rdf:type",
            "searchField": true,
            "sortButton": true,
            "sortBy": "instanceCount",
            "sortByPredicate": "^kaunokki:kansikuva/^kaunokki:manifests_in/
                                rdf:type/skos:prefLabel",
            "sortDirection": "desc"
        }
    }
}
```