



Aalto University
School of Science
Department of Media Technology



BookSampo - Lessons Learned in Creating a Semantic Portal for Fiction Literature

Eetu Mäkelä, Kaisa Hypén and Eero Hyvönen



Aalto University
School of Science
Department of Media Technology



”Complex referential concepts such as God, hermeneutics or semiotics haven't been adequately dealt with in ontologies.

Thus, ontologies cannot be applied.”

- a Finnish Librarian

Motivation

- According to a 2011 Finnish study
 - 45% of respondents used the library as a source of fiction literature
 - 34% used the library as a source of factual information

Motivation

- According to a 2011 Finnish study
 - 45% of respondents used the library as a source of fiction literature
 - 34% used the library as a source of factual information
- But the indexing tradition (classifying by topic) in libraries is geared toward books containing factual information

Motivation

- According to a 2011 Finnish study
 - 45% of respondents used the library as a source of fiction literature
 - 34% used the library as a source of factual information
- But the indexing tradition (classifying by topic) in libraries is geared toward books containing factual information
- Again, according to studies, the nature of fiction necessitates a move to describe content and context
 - Themes and locations, information about the author, publication history, awards, reviews, ...

Motivation

- According to a 2011 Finnish study
 - 45% of respondents used the library as a source of fiction literature
 - 34% used the library as a source of factual information
- But the indexing tradition (classifying by topic) in libraries is geared toward books containing factual information
- Again, according to studies, the nature of fiction necessitates a move to describe content and context
 - Themes and locations, information about the author, publication history, awards, reviews, ...
- The graph model of the Semantic Web seems a good fit for such a network of interlinked information

BookSampo

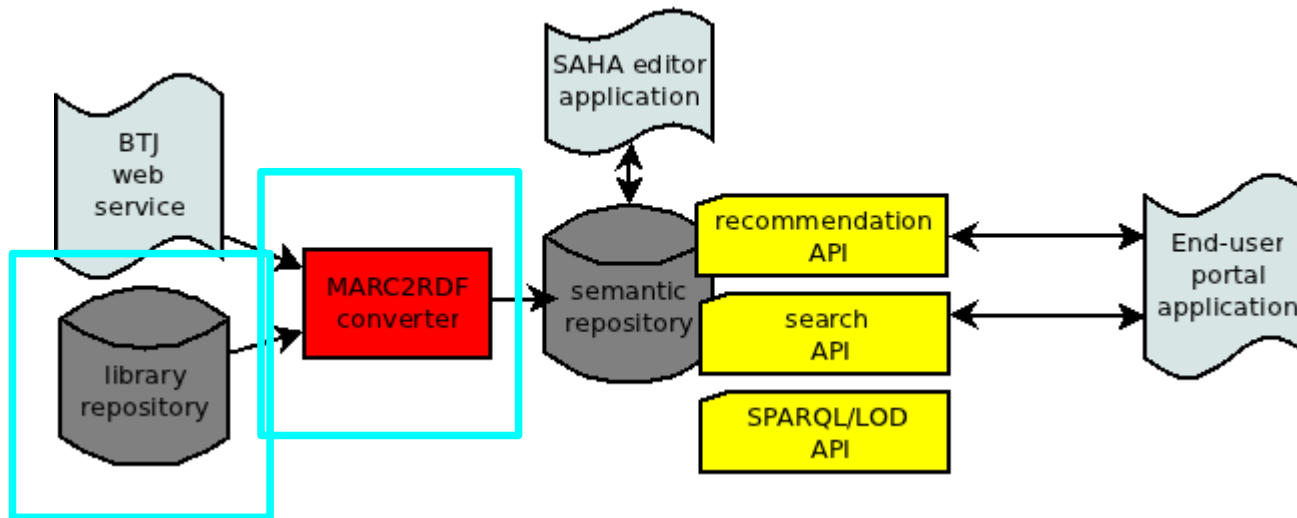
- Joint project with Finnish public libraries on providing new services for fiction literature based on semantic indexing, started in 2008

Outline

- BookSampo system description and functionality
- Lessons learned during the project
 - Insights gained in modelling fiction (as RDF)
 - Librarian experiences about converting a thesaurus into an ontology
 - Challenges faced and benefits gained from applying semantic web technologies to provide fiction literature services
 - Reception of the developed system in library circles

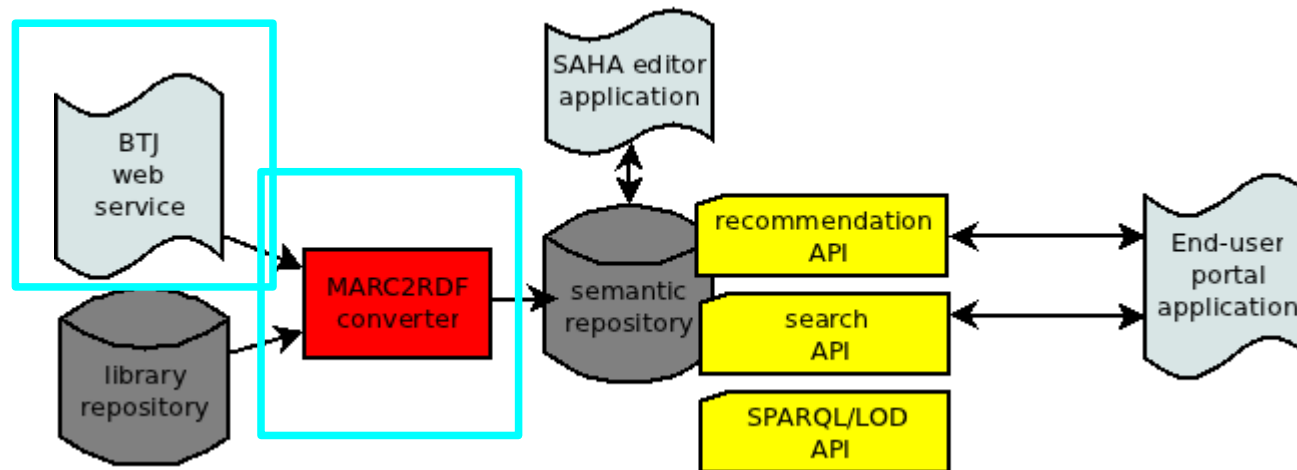
BookSampo

- Data on some 70 000 books and 25 000 authors, originally converted and semantically enriched from four library databases



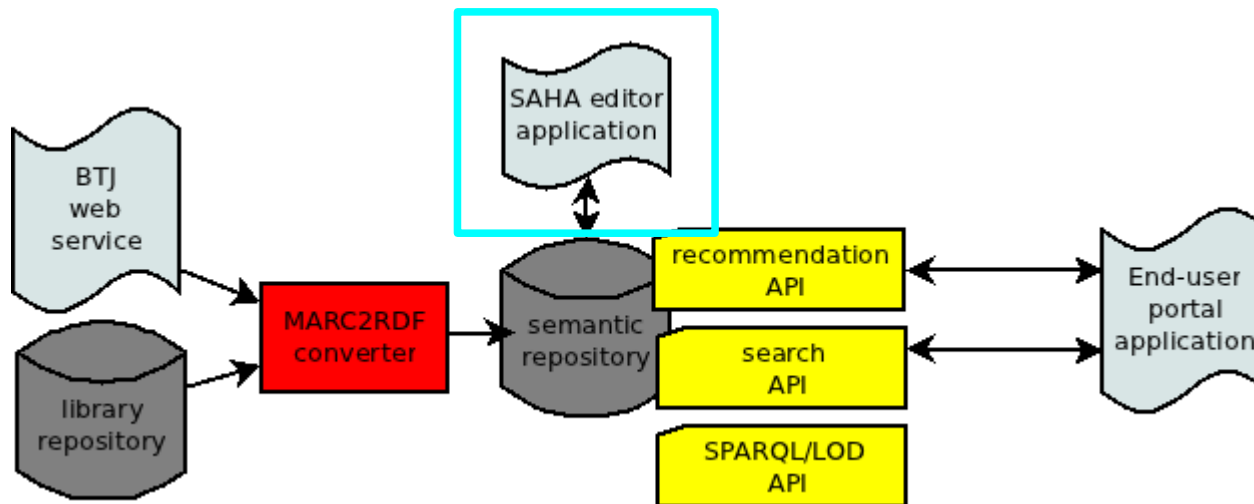
BookSampo

- Information on newly published books are imported nightly and marked for review by the librarians



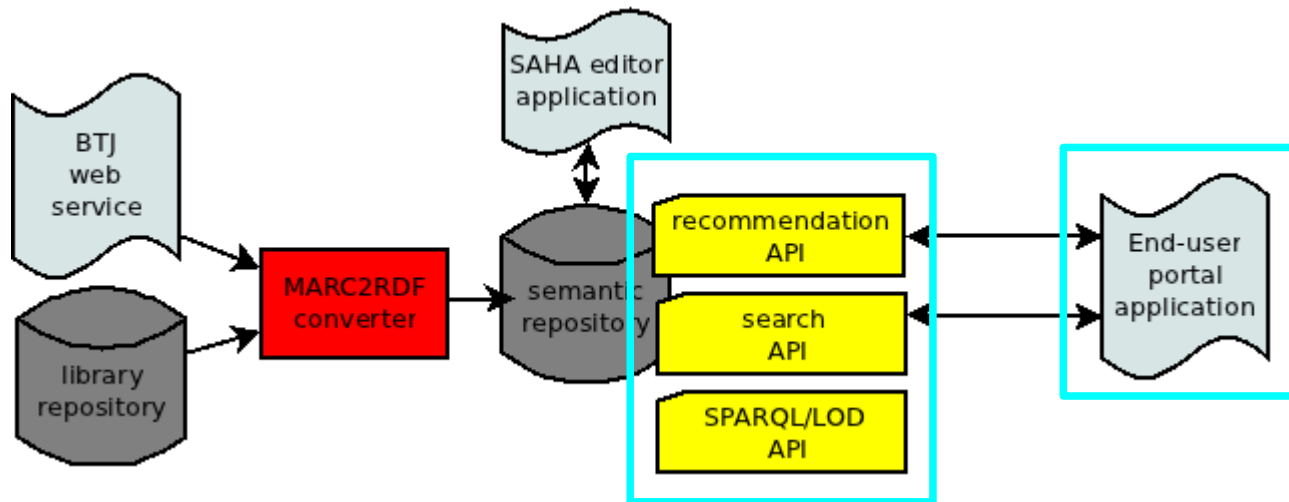
BookSampo

- Tens of librarians around Finland use a web-based metadata editor to add information and correct errors



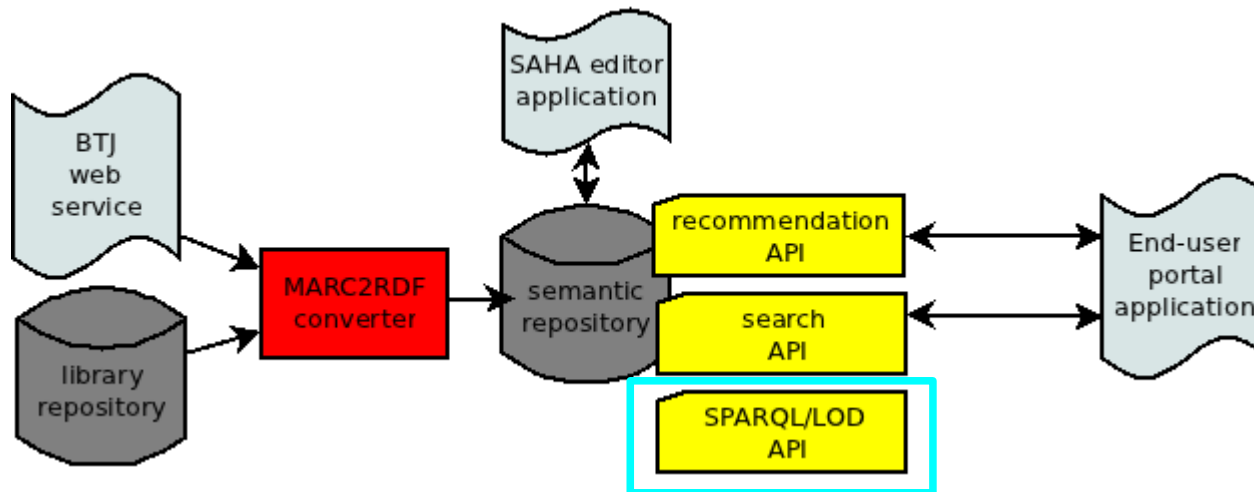
BookSampo

- End-user portal built by libraries on top of Drupal, but primary content stays in the semantic repository, used through APIs for all search and recommendation



BookSampo

- Usable also as a LOD repository through HTTP and SPARQL
 - Used in a culture hack session organized by a Finnish newspaper



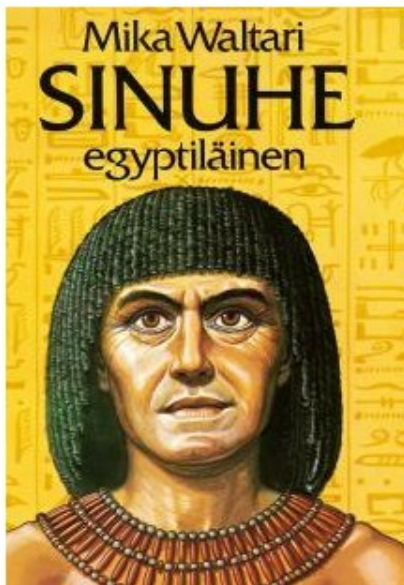
Semantic Recommendation

Etusivu

Hae Kirjasammosta



SINUHE EGYPTILÄINEN



Nimi	Sinuhe egyptiläinen
Tekijä	Waltari, Mika
Tyyppi	romaanit
Kirjallisuudenlaji	historialliset romaanit klassikot
Aiheet ja teemat	elämäntapa (elämänmuoto) matkat orjuus uskonelämä – kriisit vastoin käymiset yhteiskunta
Henkilöt, toimijat	faaraot orjat

Suosittelte

Suosituksset

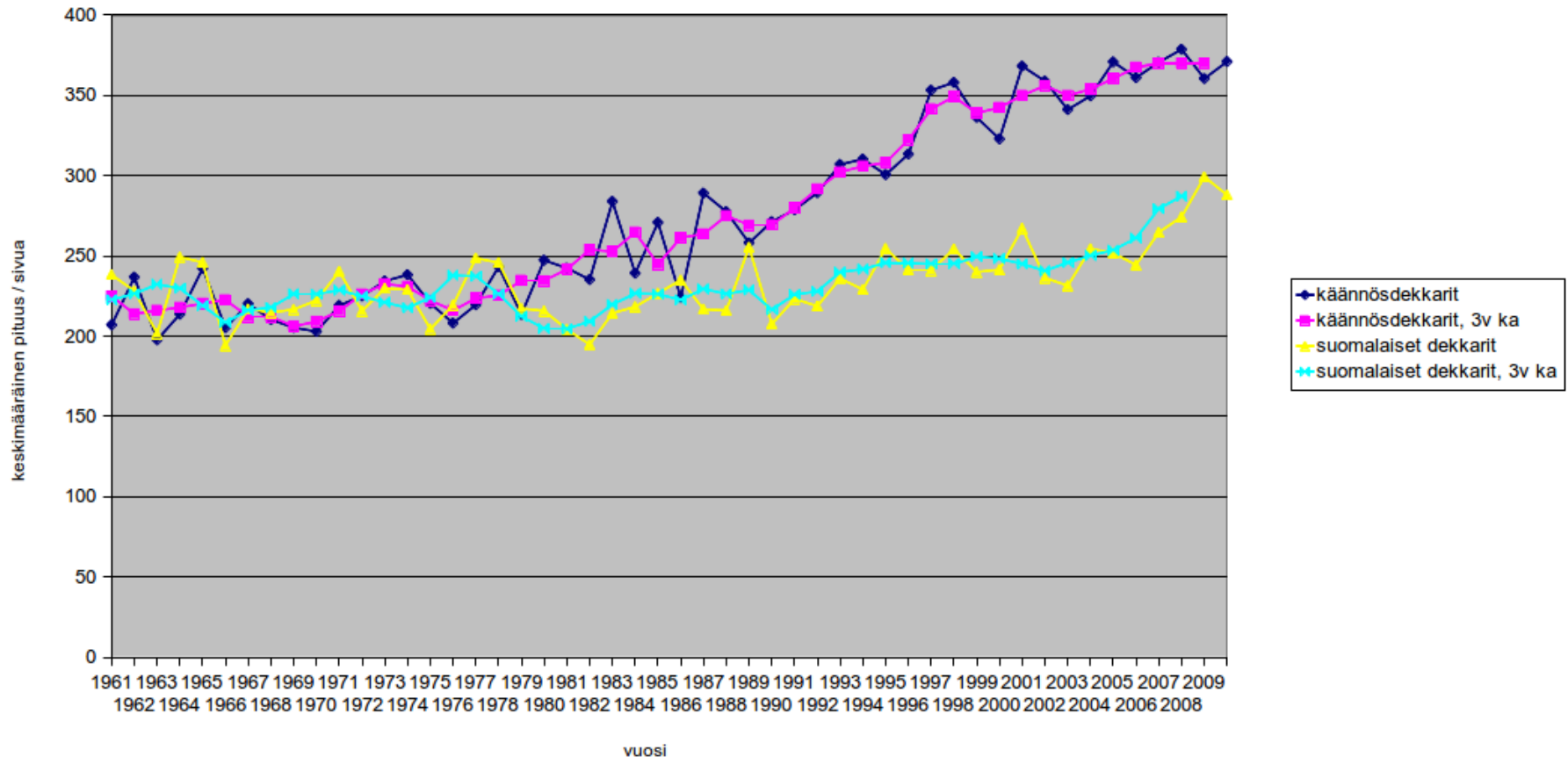
- ▶ sääty-yhteiskunta
- ▶ Nefritite
- ▶ yhteiskuntakehitys
- ▶ orjuus
- ▶ orjat
- ▶ orjakauppa
- ▶ yhteiskuntaluokat
- ▶ Auringon kehrä
- ▶ ubiikkiyhteiskunta
- ▶ tietoyhteiskunta
(yhteiskuntajärjestelmät)

Kysy kirjallisuudesta

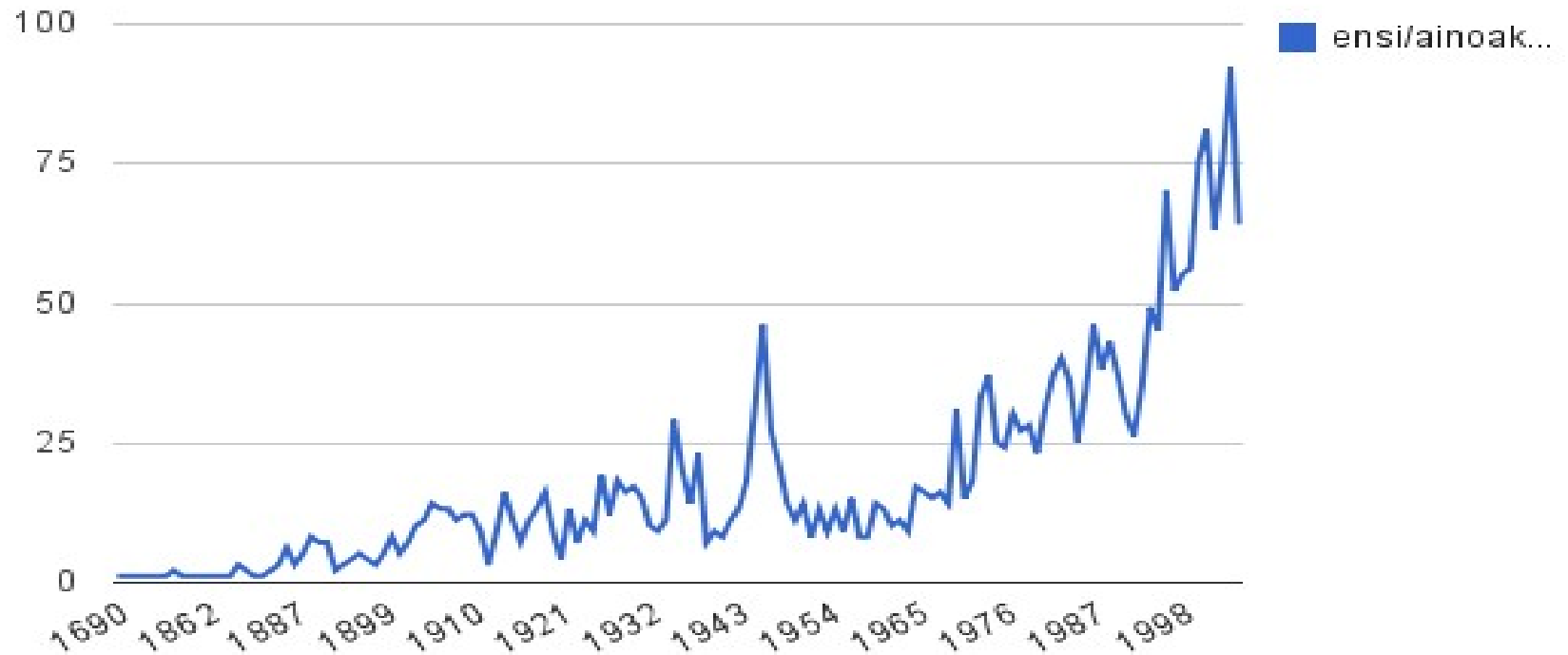
Semantic Search

- Search returns not only objects, but places the search results in their cultural-historical context
- For example, for a query of “Dostojevsky”, BookSampo not only returns his works and biography information, but also lists all authors that say they have been influenced or touched by him, all other works that are compared to his work in their reviews, all kindred works and so on

Length of Finnish and Translated Detective Stories



Number of Writers Publishing Only a Single Book in Their Career



More questions answered based on the BookSampo data (and other linked data)

- Do writers with different backgrounds (place of birth, gender, occupation) write about different themes?
- What themes should I write about if I want to receive government grants or literary awards?
- What are the most popular themes in fiction literature by year?
- Which Finnish authors have won the most awards?
- In fiction, which suburb of Helsinki is most strongly associated with crime?
- What would a most stereotypical Finnish novel look like?

SAHA Web-Based Collaborative Editing Environment

- A generic RDF editing environment for arbitrary schemas
- demo

SAHA3 | kirjasmampo - search [about kirjasmampo](#) [manage project](#) [fi](#) | [sv](#) | [en](#) || [rdf](#) || [HAKO](#) [HELP](#)

http://www.yso.fi/onto/kaunokki#ateos_26600

Novels: Sinuhe the Egyptian

[\[view\]](#) | [\[rdf\]](#) | [\[config\]](#) | [\[ren\]](#)

name	Luxor (place) →name: Luxor
creator päävastuullinen tekijä	results from http://demo Isla Luxor Ku-Luxomo Kwa Luxomo Luxorfjellet administrative region: M →Markaz Luxor inhabited place: Luxomni
character in the narrative select reference from Kaunokki-ontology	inhabited place: Luxor inhabited place: Luxor inhabited place: Luxora peak: Luxor Peak physical feature: Luxor
concrete place concrete places from the real world; select reference from Geo-ontology	luxo [remove] Babylon [edit] [remove] Egypt [edit] [remove] Syria [edit]
time of events general expressions of time; select reference from Kaunokki-ontology	select reference (range unknown) [remove] ancient times [edit]

[\[close\]](#)

name	add new literal
	<input type="text"/>
	[remove] (en) ancient times [remove] (sv) gamla tiden [remove] (fi) vanha aika



Aalto University
School of Science
Department of Media Technology



Modelling Fiction



Aalto University
School of Science
Department of Media Technology



.. or ..



Aalto University
School of Science
Department of Media Technology



**You'd think librarians as professional
metadata indexers would have little
problems with semantic indexing...**

But..

- Librarians are used to MARC indexing, which is mostly a flat form of indexing concentrating on bibliographical information of a particular edition of a book
- In BookSampo, also the references exist as primary objects (so not only books, but authors, series, awards, fictional characters, years, ..)
- This is already quite a leap for the indexers

For example:

- Instead of changing the referenced time object from the 1970's to 1960's, an indexer just opened the 1970's resource inline and changed its label to 1960's
- Once, the Novels class resource was completely removed from the project. Someone had seen a short story wrongly indexed as a novel

However,

- After sufficient training in modelling and editing, the librarians have lauded this mode of indexing as revolutionary
 - In the Linked Data environment of BookSampo, newly added authors or publishers, along with all their detailed information are immediately available and usable for all the dozens of voluntary BookSampo indexers across Finland.
 - Once entered, such information need not be repeated again for all new books, which adds an incentive to provide richer detail about these secondary resources
 - Adding a detail to any node in the graph immediately adds value also to all items linked to that node

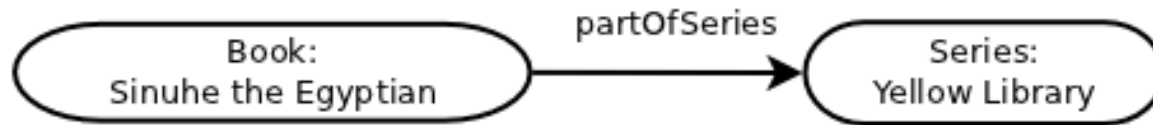
Also,

- The flexibility of the model allows for finer-grained use of the librarians expertise
 - New properties to capture information not originally thought of can be added easily
 - More finer-grained keywords can be used (and new ones created to the shared pool) without losing the benefits of a constrained vocabulary, as long as new keywords are indexed as extending the taxonomy of the constrained vocabulary

- To the librarians, this object-sharing -based Linked Data approach is in sharp positive contrast with their older MARC-record sharing schemes, where data moved in bulk batches between dark siloes

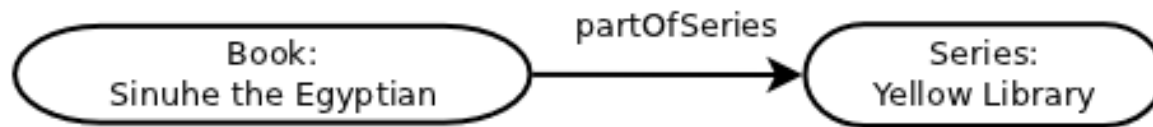
- However, there is one shortcoming in the RDF data model that, when worked around, caused much confusion

Problem With the RDF Data Model



RDF is a nice graph of independently existant objects and their relations, except ..

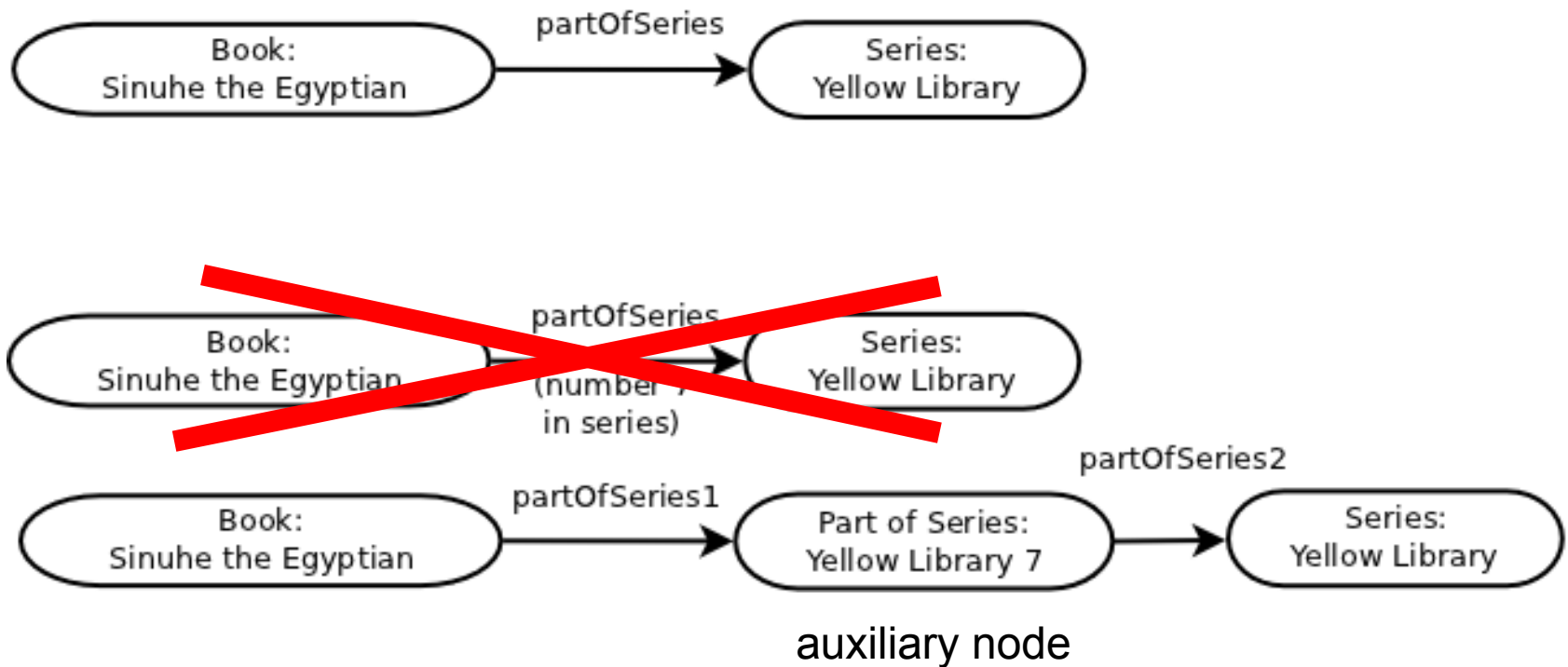
Problem With the RDF Data Model



RDF is a nice graph of independently existant objects and their relations, except when there is additional metadata related to the relation



Problem With the RDF Data Model: Need for Auxiliary Nodes



Problem With the RDF Data Model: Need for Auxiliary Nodes

- Further developments in RDF browsing and editing environments should do well to provide support for special handling of auxiliary resources, so that such inconsistencies arising from technical limitations can be hidden behind user interfaces.

Problem With the RDF Data Model: Need for Auxiliary Nodes

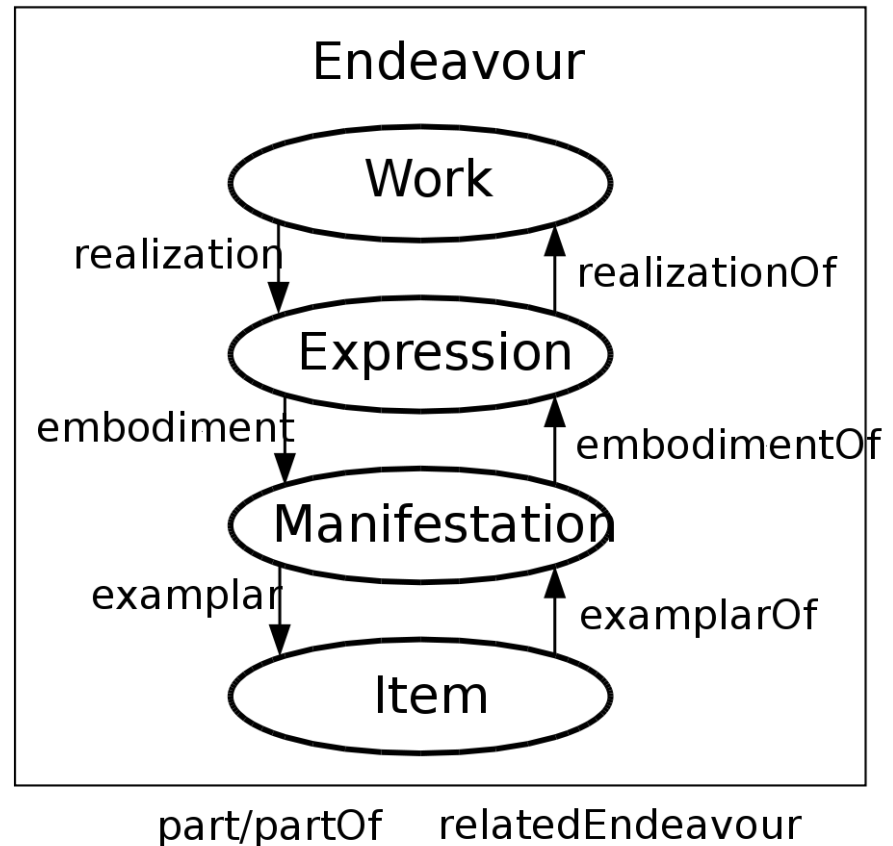
- Further developments in RDF browsing and editing environments should do well to provide support for special handling of auxiliary resources, so that such inconsistencies arising from technical limitations can be hidden behind user interfaces.
 - (We haven't solved this in our own environment, because we originally identified the wrong problem to solve)

Modelling Fiction Literature

- Using RDF (and RDF tools) made it possible change the model mid-flight.
- For example, first, biographical information on authors was encoded as events.
- User research, as well as interviewing library indexers however revealed that these were not easily understood.
- People wanted to see information on authors' birth and death dates and places as simply attribute-object values of the author, instead of as events where the author was involved in.

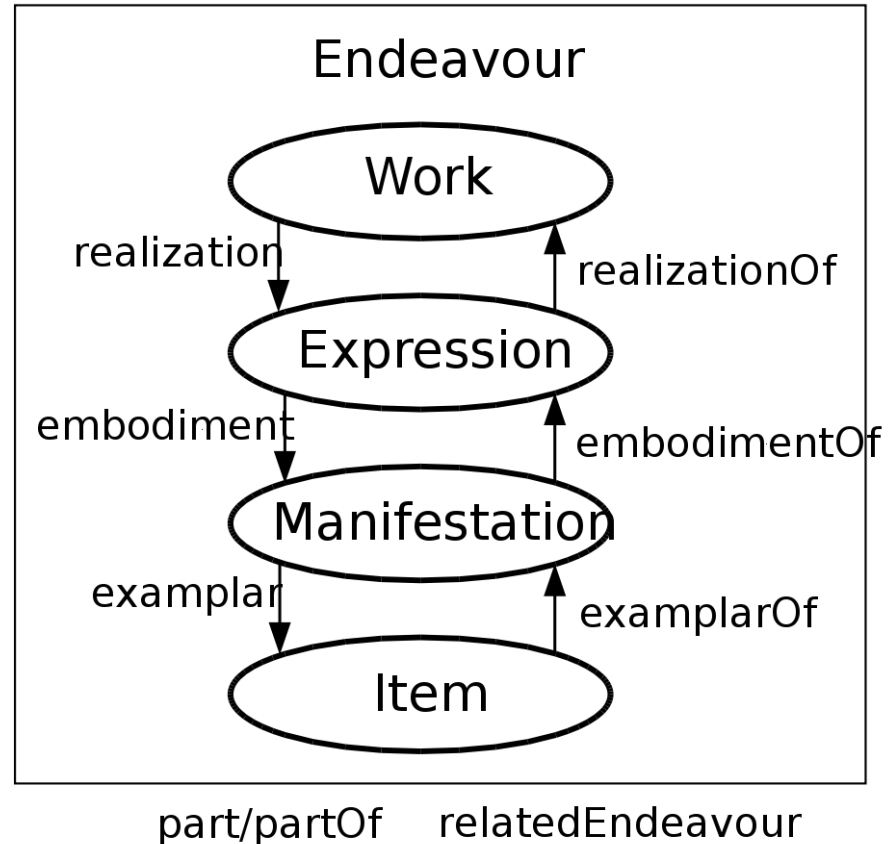
Modelling Fiction Literature

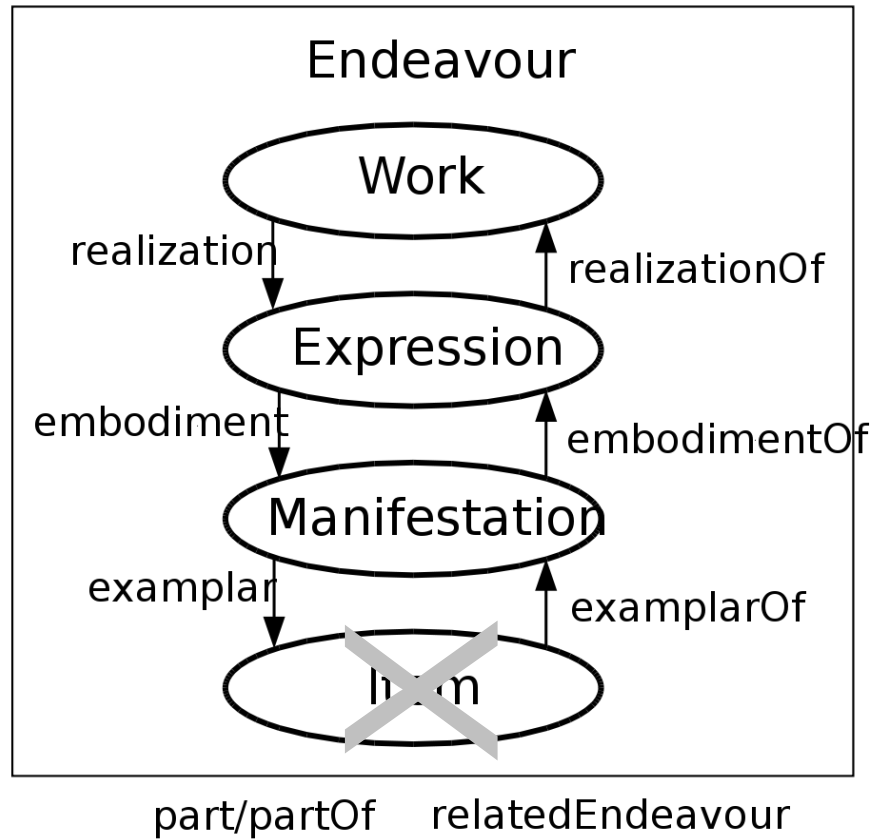
- At the beginning, BookSampo focused only on the abstract work – content, creator etc.
- However, edition-specific data crept in, such as publisher and publication date
- Then, also Swedish language editions were included
- A move was needed to a more full FRBRoo representation

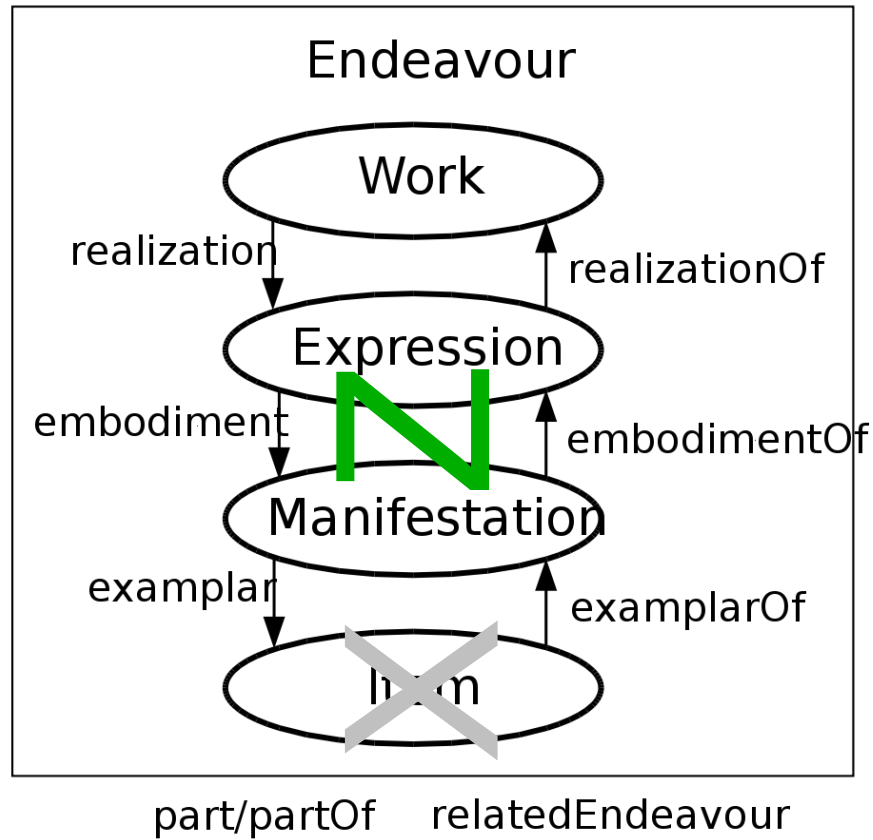


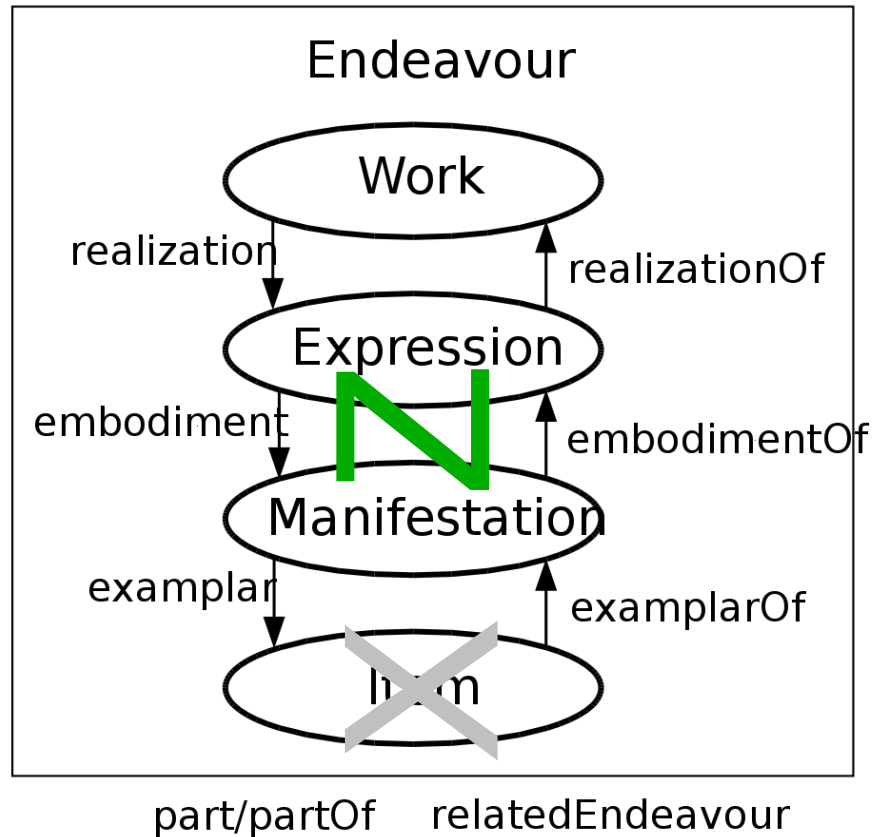
Modelling Fiction Literature

- Except the library professionals didn't want the complexity of their own full model









- Later, it turned out that the combination of expression and manifestation levels is a common choice
 - Open library and LibraryThing do it on a model level, The Australian Music Centre and the OCLC WorldCat system on the user interface level



Aalto University
School of Science
Department of Media Technology



Ontology Creation

Ontology Creation

- The experience of the librarians who ontologized the Finnish fiction thesaurus Kaunokki was that it brought in a very welcome structuring to the vocabulary
 - Organizing the keywords in subsumption hierarchies ensures that no keyword floats around in isolation, but is always surrounded by concepts that help define and relate it
 - This also beneficially forces the vocabulary keeper to narrow down their definition of the keyword, which in turn helps in ensuring uniform use by indexers



Aalto University
School of Science
Department of Media Technology



Applying SW Technologies

SAHA Web-Based Collaborative RDF Editing Environment

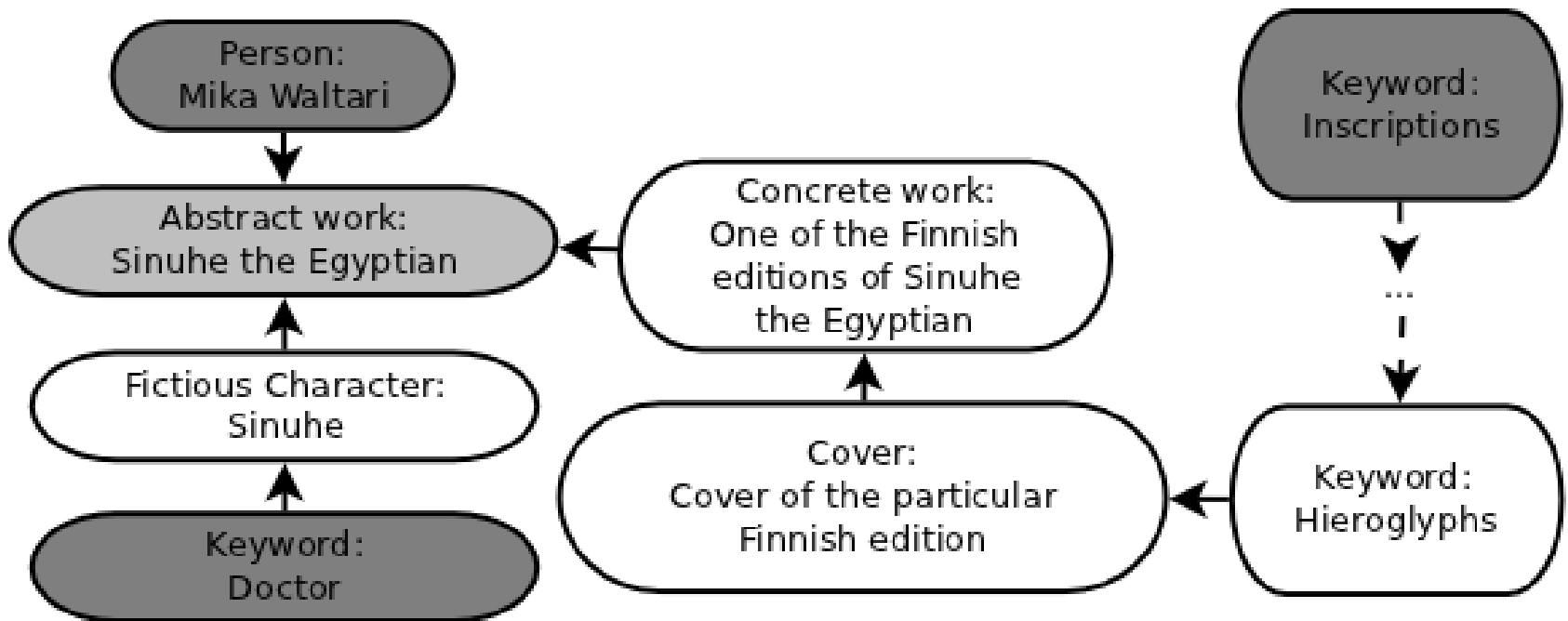
- Three years of experience from librarians using it
- The librarians particularly enjoyed
 - the linked data model in general
 - the integrated semantic autocompletion functionality with previews
 - the inbuilt chat functionality
- Inline editing of linked resources was implemented based on feedback (particularly for creating auxiliary resources)

End-User Querying

- The complexity of the BookSampo data model led to some problems in our search and recommendation services

End-User Querying

Text query: "Waltari Doctor Inscriptions" → Sinuhe the Egyptian
Need to efficiently map and combine resource matches to final results





Aalto University
School of Science
Department of Media Technology



Reception by Librarians

Reception by Librarians

- Libraries have a large amount of institutional friction in applying new forms of indexing
- Before publishing the end-user portal, the benefits of BookSampo have remained elusive to the library professionals
 - Particularly, there was a noted scepticism with regard to the added value of ontologies versus the cost of their maintenance
- However, after the end-user portal was published and the analyses made using the data from the culture hack evening were published, the search and recommendation functionalities afforded by the semantic engine have been lauded as revolutionary



Aalto University
School of Science
Department of Media Technology



Thank you

eetu.makela@aalto.fi