An Event-Based Approach for Semantic Metadata Interoperability

Tuukka Ruotsalo and Eero Hyvönen

Semantic Computing Research Group (SeCo)
Helsinki University of Technology (TKK), Laboratory of Media Technology
University of Helsinki, Department of Computer Science
firstname.lastname@tkk.fi
http://www.seco.tkk.fi/

Abstract. This paper presents a method for making metadata conforming to heterogeneous schemas semantically interoperable. The idea is to make the knowledge embedded in the schema structures interoperable and explicit by transforming the schemas into a shared, event-based representation of knowledge about the real world. This enables and simplifies accurate reasoning services such as cross-domain semantic search, browsing, and recommending. A case study of transforming three different schemas and datasets is presented. An implemented knowledge-based recommender system utilizing the results in the semantic portal CULTURESAMPO was found useful in a preliminary user study.

1 Introduction

Different heterogeneous data formats, metadata schemas, and ontologies, such as Dublin Core [2], CIDOC CRM [3], ULAN[1], and ABC [9], are in use for describing resources, such as documents, persons, artifacts, and web pages. The heterogeneity of metadata schemas and vocabularies causes problems when aggregating content for end-users with an integrated view of the data [7].

The problem of schema heterogeneity can be addressed on a *syntactic level* by deriving new schemas as extensions of existing ones, or by aligning metadata elements with each other. For example, VRA[2] extends Dublin Core elements in a compatible way by adding additional elements. CIDOC CRM [3] is an ontology developed as an underlying schema into which other metadata schemas in the cultural domain can be transformed for interoperability. On a *semantic level*, the domain ontologies whose resources can be used as values of metadata schema elements [19] can be used for enhancing interoperability [10]. To deal with problems of incompatible domain ontologies, ontology mapping and alignment or a shared upper domain ontology [4,15] can be used.

In the semantic portal MUSEUMFINLAND [10], a method was presented for transforming heterogeneous database content into a single Dublin Core -like metadata schema for representing metadata about cultural artefacts. By mapping literal metadata element values onto resources of globally shared domain ontologies, semantic interoperability

[1] http://www.getty.edu/research/conducting_research/vocabularies/ulan/
[2] http://www.vraweb.org/projects/vracore4/index.html

K. Aberer et al. (Eds.): ISWC/ASWC 2007, LNCS 4825, pp. 409–422, 2007.
© Springer-Verlag Berlin Heidelberg 2007

between different content sources was achieved, and intelligent services based on the shared metadata schema could be provided to end-users. When applying this approach to publishing cultural contents of various kinds in the semantic portal CULTURESAMPO [11], the following problems were encountered:

1. *Using heterogeneous metadata schemas.* In cross-domain applications the content is described using different kinds of metadata schemas that are already in everyday use in different domains. Enforcing content providers to use one standard is not feasible but rather the portal system has to make the schemas interoperable.
2. *Mismatch between metadata and knowledge representation formats.* The elements used in schemas have been designed from a content indexing and cataloguing point of view. When used for reasoning, other forms of knowledge representation would be more appropriate in many cases. For example, we may know that the *dc:creator* (*dc* refers to the Dublin Core metadata schema namespace) of a *painting* and a *house* is a certain person, say *John Smith*. However, from the knowledge representation viewpoint, *dc:creator* is not an appropriate property [6], because its meaning is relational referring to either a *painting* or a *building* event involving several participants. This knowledge is not available for the computer to reason about unless the different meanings of the binary property *dc:creator* in the different cases are explicated.
3. *Complexity of reasoning with multiple schemas.* Ontologies are developed for reasoning tasks [16]. When using multiple heterogeneous metadata schemas, the number of reasoning rules explodes if a different set of rules has to be specified for each schema separately. For example, the fact that a person is born somewhere at a certain time may be represented in metadata schemas in numerous ways, say with properties *placeOfBirth* and *timeOfBirth*, or with a *birth* event with the properties *time* and *place*. Harmonization of these representations enables simpler reasoning procedures that are independent of the metadata schemas used.

This paper presents an approach to deal with these problems. First, a new method for obtaining semantic interoperability of metadata conforming to *several heterogeneous schemas* is presented. We present a simple generic knowledge representation scheme underlying the metadata schemas based on knowledge about *events taking place in the real world*, such as painting an art work, manufacturing a chair, or being born at a place at a certain time. The idea of event-based knowledge representations has been successfully applied in many fields of artificial intelligence, such as natural language processing [1,22], image content description [19], and knowledge representation [20]. In our case, we employ the idea for obtaining semantic interoperability between heterogeneous metadata schemas by transforming metadata into a shared underlying event-based scheme. Second, it is shown that implicit knowledge embedded in the metadata schema structures exists. During the metadata transformation, this implicit knowledge can be made *explicit* for the machines to reason about by using the shared event-based knowledge representation scheme. It is argued that in this way more "intelligent" services to end-users can be implemented with less complex rules.

In the following, we first present a simple event-based model for representing metadata of the heterogeneous schemas. Second, methodological guidelines are presented

Table 1. Upper-level relations in the event-based knowledge representation schema

Relation	Meaning	Super-relation	Relation category	Domain	Range
agent	Initiates or performs the activity.	participant	thematic role	perdurant	concept
patient	Undergoes some change as a result of the activity.	participant	thematic role	perdurant	concept
instrument	Is used as an instrument in the activity.	participant	thematic role	perdurant	concept
goal	Is a goal of the activity.	participant	thematic role	perdurant	concept
place	Is a place of the activity.	participant	thematic role	perdurant	concept
time	Is a time of the activity.	participant	thematic role	perdurant	concept
participant	Other participant role of the perdurant concept.		thematic role	perdurant	concept
quality	Is a quality / qualifier of the entity		quality relation	concept	concept
partOf	Is a part of the entity		part name	concept	concept

for specifying the transformation from metadata schemas into the event-based model. A case study of transforming three different metadata schemas is presented. The knowledge explication method has been tested and used in practice in the semantic portal CULTURESAMPO [11] to enable metadata schema interoperability and for creating a semantic recommender system to demonstrate benefits of the approach in a real life application.

2 An Event-Based Model for Representing Metadata

In our approach a distinction is made between a *domain ontology* and *event-based metadata* conforming to an *event-based knowledge representation schema* (figure 1). The domain ontology describes the concepts specific to a certain domain, and the ontology can be divided into upper-level concepts and more specific concept hierarchies [9]. The event-based knowledge representation scheme specifies a way to represent heterogeneous metadata schemas using the domain ontology. The metadata is represented by instantiating domain ontology concepts and assigning relations between the instances with respect to the event-based knowledge representation schema.

2.1 Domain Ontology

For the domain ontology we use an ontology, such as DOLCE [4], SUMO [15], ABC [9] or YSO [12], which makes the distinction between major ontological upper categories such as perdurants, endurants, location concepts, and temporal concepts. Our particular interest is the distinction between perduring and enduring concepts' behavior in time [4]. Enduring concepts, such as *person*, *chair* or *car*, preserve their identity in time while perduring concepts refer to things that live in time; they are activities or events, such as running, swimming or raining. These concepts are used for instantiating events with thematic roles in the event-based knowledge representation schema.

2.2 Event-Based Knowledge Representation Schema

Our event-based schema introduces relations enabling representation of the original metadata as events with associated thematic roles and quality roles, an idea proposed in the fields of knowledge representation, natural language processing, and discourse

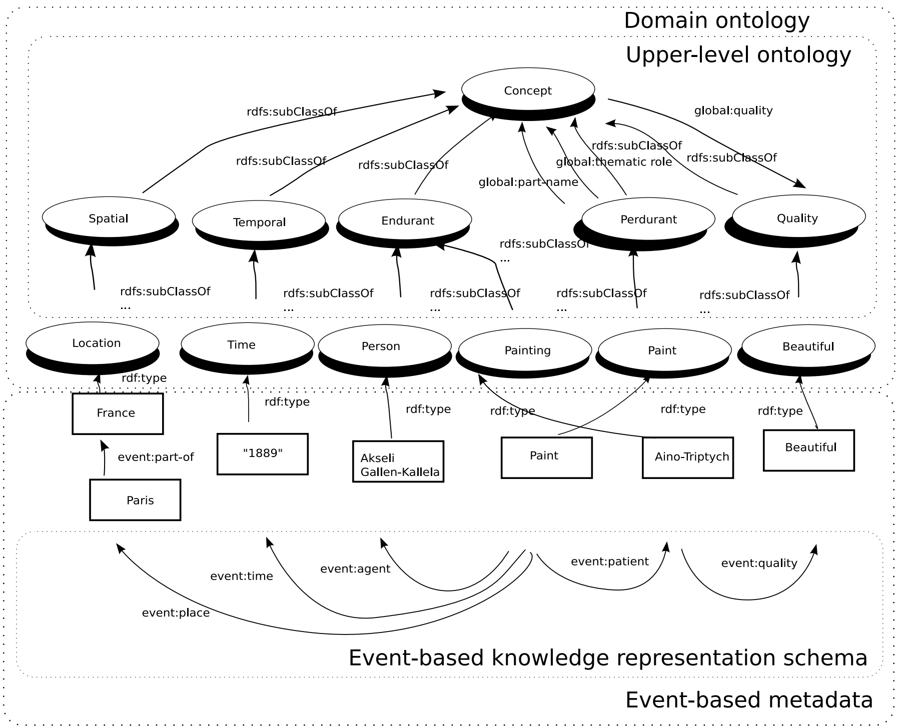

Fig. 1. Event-based model for representing metadata

modeling [1,22,20]. Table 1 presents the nine roles used in our event-based knowledge representation schema, a subset of the thematic role model of Sowa [20]. In addition to the thematic roles of perdurants, we have used the properties *partOf* and *quality* applicable to all concepts in the model.

Masolo et al. [14] propose that a concept based on a relational role is in fact a perduring concept. For example, the relation *manufacturingPlace* actually refers to the perduring concept *manufacturing* and the role *place*. Based on this notion, the relation can be represented as an event frame that consists of an instance of a perduring concept, a set of instances of participating concepts, and a set of relations between these instances. In the following a method for transforming metadata relations into events is presented.

3 Transforming Metadata Schemas to Event-Based Schema

In this section methodological guidelines are given to facilitate the event-based knowledge representation schema of heterogeneous metadata schema representations. First, the criteria for metadata schema classification using a set of meta-properties are given. Second, the method for schema explication is presented. Finally, the steps to perform the explication are shown.

Table 2. Examples of classification of relations

relation type	**meta-properties**	**relation**
non-relational	$-R-F$	person
relational	$-R+F$	teacher
quality	$+R+F$	color
part name	$+R-F$	wheelOf

3.1 Criteria for Relation Classification

To address the problem of semantic heterogeneity in metadata schemas we have followed the classification criteria of Guarino [6] and the closer analysis of relations by Masolo et al. [14]. These criteria are used to define the dependencies of the relations used in the metadata schemas. Guarino defines four different relation types: (1) relational role; (2) non-relational role; (3) quality and (4) part name. Two meta-properties are used to classify the relations: semantic rigidity and foundedness.

1. *Foundedness.* In order for a concept *x* to be founded on another concept *y*, any instance *a* of *x* has to be necessarily associated to an instance *b* of *y* which is not related to *a* by any *partOf* relation. In other words, the instances of *x* cannot exist as such except in a more comprehensive unity where they are associated to some other object. For example, *son* is founded since sons exist only within the framework of a family, where they are associated to their parents. On the other hand, the existence of *person* is essentially independent.
2. *Rigidity.* A concept is semantically rigid if it contributes to the very identity of its instances, in such a way that, if *a* is an *x* in a particular situation, it has to keep to be an *x* in any possible situation in order to keep its identity. For instance, an animal can cease to be a pup while still being a dog: *animal* and *dog* are semantically rigid, *pup* is not.

The relation types for relations are based on rigidity and foundedness of the relation. We denote rigidity with $+R$, anti-rigidity with $-R$, foundedness with $+F$ and anti-foundedness with $-F$.

Table 2 shows different relation types with examples. According to Guarino [6], an entity is considered to be a non-relational role when it is a unary predicate that does not have a natural relational interpretation. More formally, a non-relational role is a relation that is anti-rigid and anti-founded. For example, the entity *person* is a non-relational role, because it is a unary predicate that does not have an extension to any other concept in its natural interpretation.

An entity is a relational role when it is a unary predicate that has natural relational interpretation. More formally, a relation is a relational role if it is founded and anti-rigid. For example, the entity *teacher* actually refers to a *teaching* activity having the person (teacher) as an *agent* and a person (student) as a *patient*, but is represented as a binary role between the two entities. A relation is considered to be a quality if it is rigid and founded and if an instance of the entity is a predicable entity [6].

A clear distinction between qualities and other types of relations is that the interpretation of a quality is that they are predicable by themselves (i.e. may be names of predicates), but the same does not apply to other roles [6]. For example, a quality *color* can be name of a predicate and the value of the predicable instances are also qualities, such as *red*, *blue* or *green*.

part names are relations that are not founded, but are rigid. For example, a *wheel* of a *car* can exist independently of a *car*, but may be a relevant feature of a *car* in particular cases. Part names are described with a simple *partOf* relation. For a more complex meronymy we refer to [17].

3.2 A Method for Explicating Schema Knowledge

To enable the interoperability between the heterogeneous metadata schemas they have to be explicated using an event-based schema. The novel idea in our work is to use the domain ontology as a basis for describing—at the same time—the semantics of the metadata schema elements and the content descriptions of the resources, i.e. the values of the metadata schema slots. This approach provides interoperability between schema and domain semantics.

The method is based on what we call *explication* of metadata schemas. The input for applying the method is a set of metadata schemas MS, a domain ontology DO, and metadata MD conforming to MS. The output is event-based metadata EM that is metadata MD represented in a event-based knowledge representation scheme KS that is more suitable for reasoning tasks than MS. The method (for our case study schemas) consists of the following steps:

1. Classify each relation $e(x, y)$ in a metadata schema ms in MS according to the foundedness $(+/-F)$ and rigidity $(+/-R)$ criteria.
2. Explication rules for each metadata schema relation $e(x, y)$ in ms are:
   (a) If $e(x, y)$ is a non-relational role $(-R - F)$, then define $rdf : type(x, y)$ relation such that y is a concept in DO.
   (b) If $e(x, y)$ is a relational role $(-R + F)$, then create an instance p of a selected perduring concept in DO, and create a set of thematic roles $tr(p, y)$ or $tr(p, x)$ or quality roles $qr(p, y)$ such that y is an instance of a concept in DO. Add $event : hasEvent(x, p)$, which ensures that the description is connected to an original annotation source, e.g. a document. (In our case schemas the meaning of each $e(x, y)$ can be explicated with one event.)
   (c) If $e(x, y)$ is a quality role $(+R + F)$ (e.g., property "colour"), and e does not exist in DO, then explicate its meaning by selecting a concept q in DO such that $rdf : isDefinedBy(e, q)$ (e.g., class "colour").
   (d) If $e(x, y)$ is a part name relation $(+R - F)$, then define $partOf(x, y)$ relation and create statement $rdfs : subPropetyOf(e(x, y), partOf(x, y))$.
3. Transform metadata MD (conforming to MS) into EM (conforming to KS) by using the transformation rules.

Figure 2 illustrates an example of the metadata schema explication. The left side of the figure shows a part of an original metadata description from the ULAN dataset of the Getty Foundation. The relation *birthPlace* is first classified using the rigid and foundedness criteria and resolved to be a relational role. The explication against the

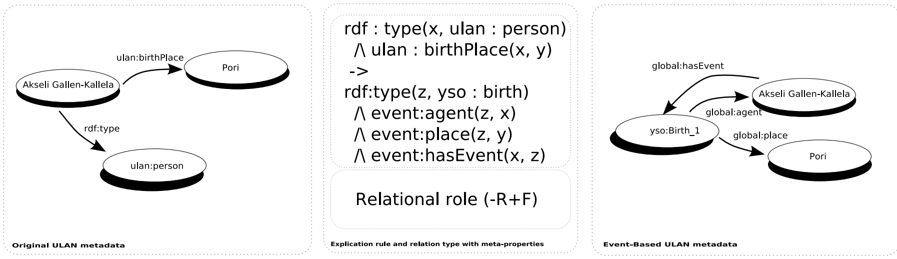

Fig. 2. An example of a metadata explication

event-based knowledge representation schema is made using the YSO [12] domain ontology. An explication rule where the instance of a perduring concept *birth* is related to the place of the birth using thematic role *place*, is derived. Finally, the right side of the figure shows the resulting event-based metadata.

3.3 Benefits of the Schema Explication

We argue that schema explication leads to the following benefits. (1) Semantic interoperability of syntactically different schemas can be obtained by defining the meaning of metadata schemas in terms of the underlying domain ontology concepts. This enables the usage of the transitive subsumption hierarchies of the domain ontology in reasoning. (2) It is possible to exploit additional semantic reasoning by explicating the hidden implicit semantics of metadata schemas. This is achieved by more explicit descriptions of the relational roles in terms of domain ontologies. For example, the relation *manufacturingPlace* can be explicated using the concept *manufacturing* and relation *place*. (3) Knowledge representation at a more foundational level reduces the number of different properties to be dealt with, which leads to simpler and more general reasoning. The number of relational roles in original schemas can be exponential, e.g. any perduring concept and role pair is possible. (4) The problem of aligning different metadata schemas onto each other becomes easier by using a canonical representation model. The number of pairwise mappings between n schemas is $O(n * (n - 1)/2)$, but there are only $O(n)$ mappings between the schemas and the event-based knowledge representation model. To test our hypotheses, we next discuss a case study of applying the metadata explication method for three different schemas used in the semantic portal CULTURESAMPO.

4 Three Case Studies

A case study using three different metadata schemas and metadata was conducted: (1) Descriptions of artifacts conforming to the Dublin Core -like metadata schema of MUSEUMFINLAND, (2) descriptions of paintings conforming to the CIDOC Conceptual Reference Model (CRM) [3] used in the Finnish National Gallery and (3) descriptions of artists conforming to the ULAN. The domain ontology used was the

Table 3. Representative relation types and explication rules in Finnish museum dataset

row	relation	relation type	classification criteria	explication rules
1	mf:museumName(x,y)	quality	$+R+F$	$mf : museumName(x, y) \rightarrow$ $rdf : isDefinedBy(mf : museumName(x, y), yso : name)$
2	mf:museumUrl(x,y)	quality	$+R+F$	$mf : museumUrl(x, y) \rightarrow$ $rdf : isDefinedBy(mf : museumUrl(x, y), yso : identifier)$
3	mf:objectType(x,y)	non-relational	$-R-F$	$mf : objectType(x, y) \rightarrow rdf : type(x, y)$
4	mf:name(x,y)	quality	$+R+F$	$mf : name(x, y) \rightarrow rdf : isDefinedBy(mf : name(x, y), yso : name)$
5	mf:manufacturingPlace(x,y)	relational	$-R+F$	$rdf : type(x, mf : museumItem) \wedge mf : manufacturingPlace(x, y) \rightarrow$ $rdf : type(z, yso : manufacturing) \wedge event : place(z, y) \wedge rdf : type(y, yso : place) \wedge event : patient(z, x) \wedge event : hasEvent(x, z)$; $rdf : type(x, mf : painting) \wedge mf : manufacturingPlace(x, y) \rightarrow$ $rdf : type(z, yso : paint) \wedge event : place(z, y) \wedge rdf : type(y, yso : place) \wedge event : patient(z, x) \wedge event : hasEvent(x, z)$
6	mf:creator(x,y)	relational	$-R+F$	$rdf : type(x, mf : museumItem) \wedge mf : creator(x, y) \rightarrow$ $rdf : type(z, yso : manufacturing) \wedge event : agent(z, y) \wedge event : patient(z, x) \wedge event : hasEvent(x, z)$
7	mf:creator(x,y)	relational	$-R+F$	$rdf : type(x, mf : painting) \wedge mf : creator(x, y) \rightarrow$ $rdf : type(z, yso : paint) \wedge event : agent(z, y) \wedge event : patient(z, x) \wedge event : hasEvent(x, z)$
8	mf:manufacturing-StartTime(x,y)	relational	$-R+F$	$rdf : type(x, mf : museumItem)$ $\wedge mf : manufacturingStarTime(x, y) \rightarrow$ $event : time(z, k) \wedge event : startTime(k, y)$
9	mf:manufacturing-EndTime(x,y)	relational	$-R+F$	$rdf : type(x, mf : museumItem)$ $\wedge mf : manufacturingEndTime(x, y) \rightarrow$ $event : time(z, k) \wedge event : endTime(k, y)$
10	mf:material(x,y)	relational	$-R+F$	$rdf : type(x, mf : museumItem) \wedge event : material(x, y) \rightarrow$ $rdf : type(z, yso : manufacturing) \wedge event : material(z, y) \wedge event : hasEvent(x, z)$
11	mf:keyword(x,y)	relational	$-R+F$	$rdf : type(x, mf : museumItem) \wedge mf : keyword(y) \wedge y \in yso : perduring \rightarrow k = y \wedge event : hasEvent(x, k)$; $k \notin yso : perduring \rightarrow rdf : type(k, yso : perduring) \wedge event : hasEvent(x, k)$; $rdf : type(x, mf : museumItem) \wedge mf : keyword(y) \wedge y \in yso : enduring \rightarrow event : participant(k, x)$
12	mf:stylePeriod (x,y)	quality	$+R+F$	$mf : stylePeriod(x, y) \rightarrow rdf : isDefinedBy(mf : stylePeriod(x, y), yso : stylePeriod)$
13	mf:inCollection (x,y)	part name	$+R-F$	$mf : inCollection(x, y) \rightarrow event : partOf(x, y) \wedge rdf : type(yso : museumCollection, y) \wedge rdfs : subPropertyOf(mf : inCollection(x, y), event : partOf(x, y))$
14	mf:part(x,y)	part name	$+R-F$	$mf : part(x, y) \rightarrow event : partOf(x, y) \wedge rdfs : subPropertyOf(mf : part(x, y), event : partOf(x, y))$

General Finnish Ontology YSO [13]. It contains some 20,000 general concepts in ten major facets including perduring objects (e.g. events and activities), enduring objects (e.g. physical things), properties, time, and locations. This lightweight ontology was created based on the General Finnish Thesaurus YSA[3]. The namespace *mf* is used to refer to the MuseumFinland system, *crm* to CIDOC CRM, *ulan* to ULAN, *event* conforming to our event-based knowledge representation schema and RDF(S) to Resource Description Framework [4]. We use logic programming syntax to express the rules [5].

Case Study 1: Finnish Museum Data. The Finnish museum dataset contains 4453 descriptions of museum items. We analyzed the superset of the relations occurring in the dataset and used the method to explicate the relations. Table 3 describes a selection of typical relations, the classification of the relations, and the rules defined for explication of the relations in the MuseumFinland metadata schema.

[3] http://vesa.lib.helsinki.fi
[4] http://www.w3.org/RDF/
[5] Dot (.) is used to indicate chained relations.

All relations in the dataset were explicated. On row 1, the relation *mf:museumName* was aligned to *yso:name* in the domain ontology. Another option would have been to extend the domain ontology to contain a sub-class of *yso:name* and align *mf:museumName* to this additional property. On rows 6 and 7 the relational role *mf:creator* is founded by the type of the object. Two separate rules were written. First, the objects typed as paintings were explicated by *yso:paint* activity while the objects typed as *mf:museumItems* were explicated by *yso:manufacturing* activity. On rows 8 and 9 the representation of interval time forced to introduce a simple time object that was able to represent *event:startTime* and *event:endTime* values. On row 11 a more complex rule was written to handle the *mf:keyword* relation. The relation was relational with respect to its values, i.e. the thematic roles were missing. A simple rule was written to predict the missing thematic roles. First, if the *mf:keyword* contained an instance that was in the sub-class hierarchy of *yso:perduring* concepts, then the instance was set to be the perduring object in the event-based description. Otherwise a superclass of this hierarchy, i.e. *yso:perdurant* was instantiated. Other values that were in the sub-class hierarchy of enduring objects were set as the value of the *event:participant* role to the perduring concept instance.

Case Study 2: CIDOC CRM from the National Gallery of Finland. The National gallery of Finland dataset conforms to the CIDOC CRM model and contains 553 descriptions of fine arts items. The content descriptions (*crm:isAbout* relation) were originally annotated using the ICONCLASS[6] vocabulary. A pre-processing stage was conducted and the descriptions were transformed to use the YSO ontology using a simple string matching alignment based on lemmatized labels of the concepts. Table 4 describes the partial but representative relations and the explication rules of the dataset.

All of the relations in the dataset were explicated. On row 7 *crm:productionEvent* is directly the perduring concept. In this case a separate alignment to YSO ontology was required. On row 8 the representation of time is again an interval and represented with a CIDOC CRM specific representation schema. On row 11 a new thematic role *technique* was introduced as a sub-property for *participant* to enable more specific correspondence with the original metadata schema. On row 15 the type property of the *crm:depicts* event has a value from CIDOC CRM ontology and therefore requires a separate alignment to the YSO ontology. As noted before, the domain ontology level alignment was performed before the explication.

Case Study 3: ULAN of Getty Foundation. A subset of Finnish Artists in the ULAN dataset contains 429 metadata descriptions. A preprocessing stage was conducted and the descriptions were transformed to use the YSO ontology as in case study 2. Table 5 describes partial, but representative set of relations in the dataset and the explication rules. All of the relations in Finnish Artists in ULAN dataset were explicated using the method. Some relations such as *ulan:nationalities* on row 3, *ulan:role* on row 4 and *ulan:gender* on row 5 required domain ontology alignment. This means the values of the relations were from the ULAN domain ontology and were separately aligned to YSO ontology concepts.

[6] http://www.iconclass.nl/

Table 4. Representative relation types and explication rules in National Gallery of Finland dataset

row	relation	relation type	classification criteria	explication rules
1	crm:hasType(x,y)	non-relational	$-R-F$	$crm : hasType(x, y) \rightarrow rdf : type(x, y)$
2	crm:consistsOf(x,y)	relational	$-R+F$	$rdf : type(x, crm : painting) \wedge crm : consistsOf(y) \rightarrow rdf : type(z, yso : manufacturing) \wedge event : material(z, y) \wedge event : patient(z, x) \wedge rdf : type(y, yso : material) \wedge event : hasEvent(x, z)$
3	crm:hasTitle(x,y)	quality	$+R+F$	$crm : hasTitle(x, y) \rightarrow rdf : isDefinedBy(crm : hasTitle(x, y), yso : title)$
4	crm:hasDimension(x,y)	quality	$+R+F$	$crm : hasDimension(x, y) \rightarrow rdf : isDefinedBy(crm : hasDimension(x, y), yso : dimension)$
5	crm:isAbout. conceptualObject(x,y)	relational	$-R+F$	$rdf : type(x, crm : painting) \wedge crm : isAbout.conceptualObject(y) \wedge y \in yso : perduring \rightarrow k = y \wedge event : hasEvent(x, k);$ $k \notin yso : perduring) \rightarrow rdf : type(k, yso : perduring) \wedge event : hasEvent(x, k);$ $rdf : type(x, crm : painting) \wedge crm : isAbout.conceptualObject(y) \wedge y \in yso : enduring \rightarrow event : participant(k, x)$
6	crm:isAbout. actor(x,y)	relational	$-R+F$	$rdf : type(x, crm : painting) \wedge crm : isAbout.actor(x, y) \rightarrow event : agent(k, y) \wedge event : hasEvent(x, z)$
7	crm:isAbout. productionEvent(x,z)	relational	$-R+F$	$crm : isAbout.productionEvent(x, z) \wedge event : patient(z, x) \wedge event : hasEvent(x, z)$
8	crm:isAbout. productionEvent. hasTimeSpan. atSomeTimeWithin(x,y)	quality	$+R+F$	$crm : isAbout.productionEvent.hasTimeSpan. atSomeTimeWithin(x, y) \rightarrow event : time(z, y)$
9	crm:isAbout. productionEvent. tookPlaceAt(x,y)	relational	$-R+F$	$rdf : type(x, crm : painting) \wedge crm : isAbout.productionEvent. tookPlaceAt(x, y) \rightarrow rdf : type(z, yso : paint) \wedge event : place(z, y) \wedge event : hasEvent(x, z)$
10	crm:isAbout. productionEvent. carriedOutBy(x,y)	relational	$-R+F$	$rdf : type(x, crm : painting) \wedge isAbout.productionEvent. carriedOutBy(x, y) \rightarrow event : agent(z, y) \wedge event : hasEvent(x, z)$
11	crm:isAbout. productionEvent. usedGeneralTechnique(x,y)	quality	$+R+F$	$rdf : type(x, crm : painting) \wedge crm : isAbout.productionEvent. usedGeneralTechnique(x, y) \rightarrow event : technique(z, y) \wedge event : hasEvent(x, z)$
12	crm:wasUsedFor. activity. generalPurpose(x,y)	relational	$-R+F$	$rdf : type(x, crm : painting) \wedge crm : isAbout. productionEvent.generalPurpose(x, y) \rightarrow event : goal(z, y) \wedge event : hasEvent(x, z)$
13	crm:depicts(x,y)	relational	$-R+F$	$rdf : type(x, crm : painting) \wedge crm : depicts(x, y) \rightarrow rdf : type(z2, yso : depict) \wedge event : patient(z2, y) \wedge event : hasEvent(x, z2)$
14	crm:depicts. informationCarrier. about(x,y)	relational	$-R+F$	$crm : depicts.informationCarrier.about(x, y) \rightarrow event : effector(z2, y) \wedge event : hasEvent(x, z2)$
15	crm:depicts. informationCarrier. type(x,y)	non-relational	$-R-F$	$crm : depicts.informationCarrier.type(x, k) \rightarrow rdf : type(m, k)$
16	crm:isDocumentedIn. document(x,y)	relational	$-R+F$	$rdf : type(crm : painting, x) \wedge crm : isDocumentedIn. document(x, y) \rightarrow rdf : type(z3, yso : documenting) \wedge event : patient(z3, x) \wedge event : hasEvent(x, z3)$

Table 5. Representative relation types and explication rules in ULAN dataset

row	relation	relation type	classification criteria	explication rules
1	ulan:name(x,y)	quality	$+R+F$	$ulan : name(x) \rightarrow rdf : isDefinedBy(ulan : name(x, y), yso : name)$
2	ulan:alternativeName(x,y)	quality	$+R+F$	$ulan : alternativeName(x, y) \rightarrow rdf : isDefinedBy(ulan : alternativeName(x, y), yso : additionalName)$
3	ulan:nationalities(x,y)	quality	$+R+F$	$ulan : nationalities(x, y) \rightarrow rdf : isDefinedBy(ulan : nationalities(x, y), yso : nationalities)$
4	ulan:role(x,y)	non-relational	$-R-F$	$ulan : role(x, y) \rightarrow rdf : type(x, y)$
5	ulan:gender(x,y)	quality	$+R+F$	$ulan : gender(x, y) \rightarrow rdf : isDefinedBy(ulan : gender(x, y), yso : gender)$
6	ulan:birthPlace(x,y)	relational	$-R+F$	$rdf : type(x, ulan : person) \wedge ulan : birthPlace(x, y) \rightarrow rdf : type(z, yso : birth) \wedge event : agent(z, x) \wedge event : location(z, y) \wedge event : hasEvent(x, z)$
7	ulan:deathPlace(x,y)	relational	$-R+F$	$rdf : type(x, ulan : person) \wedge ulan : deathPlace(x, y) \rightarrow rdf : type(z2, yso : death) \wedge event : agent(z2, x) \wedge event : location(z2, y) \wedge event : hasEvent(x, z2)$
8	ulan:studentOf(x,y)	relational	$-R+F$	$rdf : type(x, ulan : person) \wedge ulan : studentOf(x, y) \rightarrow rdf : type(z3, yso : teaching) \wedge event : agent(z3, y) \wedge event : patient(z3, x) \wedge event : hasEvent(x, z3)$

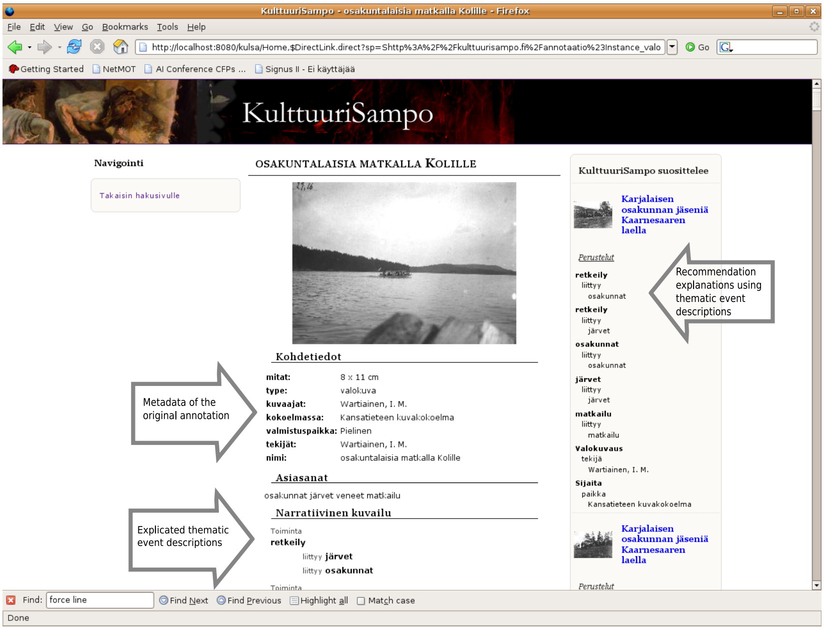

Fig. 3. User interface of CULTURESAMPO recommendation system

4.1 Implemented Use Case

The method and the case studies presented above have been implemented in the CULTURESAMPO prototype portal [11]. Explication rules were written for each schema using the Java-based Prolog system Prova[7].

Figure 3 illustrates the user interface of the portal showing a page about a photograph concerning a student union traveling to the Koli mountain in Karelia. On the right side the system gives recommendation links to other content items with explanations such as “hiking related to a student association” and “traveling related to a student association”. The event-based system gives these links because the image describes a “hiking” event with a “student association” and “lake” in participant roles. The method also gives links to content items that are “stored” in same collection, “photographed” by the same person, etc.

The recommendation system has been empirically evaluated by seven users and in total seventy metadata description pairs. The precision of the method using the event-based knowledge-base was 82 per cent. For a complete description of the recommendation method and the empirical evaluation and results see [18].

5 Discussion

Recent work on schema matching using ontologies suggests that no common or a minimal ontological commitment is needed [5] and that it is unrealistic to assume that there

[7] http://www.prova.ws/

will be an agreement of one or even a small set of ontologies [16]. As a result, ontologies and metadata schemas will be developed by communities without global standardization. To overcome the interoperability problem, additional representation formalisms defining the inter-ontology or inter-metadata schema mappings have been proposed, as reviewed in [21].

In this paper we have proposed an approach that utilizes domain ontologies and an event-based knowledge representation schema to enable heterogeneous metadata interoperability. Methodological guidelines to explicate schema and metadata content in terms of events were presented and applied successfully to three highly heterogeneous metadata schemas. To utilize the resulting event-based knowledge representation, a semantic recommender system in the semantic portal CULTURESAMPO was implemented and tested [18]. In this practical use case the usefulness of the event-based approach was shown in the form of an intuitive user interface, a standardized reasoning procedure, and enhanced relevance precision.

While the case study presented in this paper confirmed that the event-based knowledge representation schema was able to represent all of the needed implicit metadata, some difficulties were encountered when using the explication method. Some of the relations referred to local domain ontology resources that had to be mapped separately onto YSO concepts. For example, the *ulan:gender* relation in the ULAN dataset referred to *ulan:female* or *ulan:male* and was mapped to the corresponding concepts in YSO. A major problem was how to enrich the metadata with new thematic roles. For example, in the National Gallery of Finland and Finnish museums datasets the content descriptions of the values contained values such as *yso:horse*, *yso:ride*, and *yso:man* without any relation to each other. Thematic roles can easily be resolved by a human annotator, e.g. that a man rides a horse, and not that a horse rides a man. However, selecting the fillers of the roles often requires tacit human knowledge and is difficult for fully automated methods. This problem is a topic of ongoing research (cf. e.g. [1]) and requires further development in the heterogeneous schema integration field.

The idea of using event-based frames for representing knowledge has been explored in many areas of research [22,1,20,19]. There are a number of metadata models that recognize the importance of events or actions in unambiguously describing resources and facilitating interoperability across the domains [9,3,4]. Many of these ontologies propose an upper-level class hierarchy that can be extended by the domain ontologies. Wache et al. [21] give an extensive survey of current approaches including single, global and hybrid ontology approaches. Semi-automatic methods based on statistical matching of ontologies have been studied [8,16].

An event-based canonical model for metadata in cultural heritage domain is proposed in the CIDOC CRM model [3]. It "provides the definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation"[8]. The framework includes 81 classes, such as *crm:Man-Made Object*, *crm:Place*, and *crm:Time-Span*, and a large set of 132 properties relating the entities with each other, such as *crm:Has Time-Span* and *crm:IsIdentifiedBy*. Our approach is different in that our underlying knowledge representation does not concentrate on documentation but describes the underlying real world. The idea is to use existing ontologies

[8] http://cidoc.ics.forth.gr/

of thousands of classes describing the world in the annotations. In contrast to our model, CIDOC CRM contains many very specific properties, such as *crm:is_documented_in* and *crm:was_destroyed_by*. In our approach they are considered highly relational, and are described using events such as "documenting" and "'destroying". In our case study, CIDOC CRM was therefore considered as an example of a heterogeneous metadata schema to be made semantically interoperable with the other metadata schemas.

Our research is part of the National Finnish Ontology Project (FinnONTO) 2003-2007[9], funded mainly by the National Technology Agency (Tekes) and a consortium of 37 companies and public organizations.

References

1. Baker, C.F., Fillmore, C.J., Lowe, J.B.: The Berkeley Framenet project. In: Proceedings of the 17th international conference on Computational linguistics, Proceedings of the 36th annual meeting on Association for Computational Linguistics, vol. 1 (1998)
2. Baker, T.: A grammar of Dublin Core. D-Lib Magazine 6(10) (2000)
3. Doerr, M.: The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. AI Magazine 24(3), 75–92 (2003)
4. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L.: Sweetening ontologies with DOLCE. In: Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web, pp. 166–181 (2002)
5. Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies 43(5-6), 907–928 (1995)
6. Guarino, N.: Concepts, attributes and arbitrary relations: some linguistic and ontological criteria for structuring knowledge bases. Data and Knowledge Engineering 8(3) (1992)
7. Hakimpour, F., Geppert, A.: Resolving semantic heterogeneity in schema integration: an ontology based approach. In: Proceedings of FOIS 2001 (2001)
8. Huma, Z., Jaffar-Ur, M., Iftikhar, N.: An ontology-based framework for semi-automatic schema integration. Journal of Computer Science and Technology 20(6), 788–796 (2005)
9. Hunter, J.: Enhancing the semantic interoperability of multimedia through a core ontology. IEEE Transactions on Circuits and Systems for Video Technology 13(1), 49–58 (2003)
10. Hyvönen, E., Mäkelä, E., Salminen, M., Valo, A., Viljanen, K., Saarela, S., Junnila, M., Kettula, S.: MuseumFinland – Finnish museums on the semantic web. Journal of Web Semantics 3(2), 25 (2005)
11. Hyvönen, E., Ruotsalo, T., Häggström, T., Salminen, M., Junnila, M., Virkkilä, M., Haaramo, M., Mäkelä, E., Kauppinen, T., Viljanen, K.: CultureSampo–Finnish culture on the semantic web: The vision and first results. In: Developments in Artificial Intelligence and the Semantic Web - Proceedings of the 12th Finnish AI Conference STeP 2006, October 26-27 2006 (2006)
12. Hyvönen, E., Valo, A., Komulainen, V., Seppälä, K., Kauppinen, T., Ruotsalo, T., Salminen, M., Ylisalmi, A.: Finnish national ontologies for the semantic web - towards a content and service infrastructure. In: DC 2005. Proceedings of International Conference on Dublin Core and Metadata Applications (November 2005)
13. Hyvönen, E., Viljanen, K., Mäkelä, E., Kauppinen, T., Ruotsalo, T., Valkeapää, O., Seppälä, K., Suominen, O., Alm, O., Lindroos, R., Känsälä, T., Henriksson, R., Frosterus, M., Tuominen, J., Sinkkilä, R., Kurki, J.: Elements of a national semantic web infrastructure - case study finland on the semantic web (invited paper). In: IEEE ICSC 2007. Proceedings of the First International Semantic Computing Conference, Irvine, California, IEEE Computer Society Press, Los Alamitos (forth-coming, 2007)

[9] http://www.seco.tkk.fi/projects/finnonto/

14. Masolo, C., Guizzardi, G., Vieu, L., Bottazzi, E., Ferrario, R.: Relational roles and qua-individuals. In: Proceedings of the AAAI Fall Symposium on Roles, an interdisciplinary perspective, Hyatt Crystal City, Arlington, Virginia. (2005)
15. Niles, I., Pease, A.: Towards a standard upper ontology. In: FOIS 2001. Proceedings of the international conference on Formal Ontology in Information Systems, Ogunquit, Maine, pp. 2–9. ACM Press, New York, NY, USA (2001)
16. Noy, N.F.: Semantic integration: a survey of ontology-based approaches. SIGMOD Rec. 33(4), 65–70 (2004)
17. Odell, J.J.: Six sifferent kinds of composition. Journal Of Object-Oriented Programming 5(8) (1994)
18. Ruotsalo, T., Hyvönen, E.: A method for determining ontology-based semantic relevance. In: DEXA 2007. Proceedings of the 18th International Conference on Database and Expert Systems Applications, Springer, Heidelberg (2007)
19. Schreiber, A.T., Dubbeldam, B., Wielemaker, J., Wielinga, B.: Ontology-based photo annotation. IEEE Intelligent Systems 16(3) (2001)
20. Sowa, J.F.: Knowledge Representation: Logical, Philosophical and Computational Foundations. Brooks/Cole, Pacific Grove, CA (2000)
21. Wache, H., Vögele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., Hübner, S.: Ontology-based integration of information — a survey of existing approaches. In: Stuckenschmidt, H. (ed.) IJCAI 2001. Workshop: Ontologies and Information Sharing, pp. 108–117 (2001)
22. Zarri, G.P.: NKRL, a knowledge representation tool for encoding the meaning of complex narrative texts. Natural Language Engineering 3(2) (1988)