Paths and shortcuts in an event-oriented ontology

Mark Fichtner, Vincent Ribaud

Paths and Shortcuts in an Event-Oriented Ontology

Mark Fichtner[1] and Vincent Ribaud[2]

[1] Zoologisches Forschungsmuseum Alexander Koenig, Bonn, Germany
m.fichtner@wiss-ki.eu
[2] Lab-STICC MOCS, Université de Bretagne Occidentale, Brest, France
ribaud@univ-brest.fr

Abstract. The CIDOC CRM is an event-oriented ontology used in cultural heritage documentation. Events are temporal entities that are used as hooks for relating persistent entities. However end-users are relating persistent entities in a direct manner (e.g. J.R.R. Tolkien wrote Bilbo the Hobbit) and skip the path through a temporal entity. Fauconnier and Turner suggest that human conscious thinking tends to compress complex paths into simpler relationships, despite still knowing subconsciously about the complete paths. This paper presents two prototypical approaches yielding compression and decompression to the end-user, shortcuts implementation in Semantic Media Wiki and ontology path features in the WissKI system. Lessons learned yield research perspectives about identification, names, end-user usability, and event pattern heuristics.

Keywords: CIDOC CRM, semantic association, end-user representation.

1 Introduction

"*As told in The Hobbit, there came one day to Bilbo's door the great wizard, Gandalf the Grey, and thirteen dwarves with him [...] With them he set out [...] on a morning of April, it being then the year 1341 [...] on the quest of a great treasure [...] The party was assailed by Orcs in a high pass of the Misty Mountains [...] it happened that Bilbo was lost for a while in the black orcs-mines under the mountains [...] he put his hand on a ring, lying on the floor of a tunnel. He put it in his pocket [...] At the bottom of the tunnel [...] lived Gollum. [...] He possessed a secret treasure [...] a ring of gold that made its wearer invisible.*" In this excerpt from [13], J.R.R. Tolkien relates the circumstances in which Bilbo found the One Ring. Humans compress the topic with a sentence (http://www.thehobbithole.co.uk/bilbo_page.htm) such as "*As Bilbo groped along the dark tunnels [of the Misty Mountains], he found the Ring lying on the ground and slipped it into his pocket. By a subterranean lake Bilbo met Gollum, the creature who had lost the Ring*." We might also sum up the story thus: "Bilbo took the One Ring from Gollum in 1341 under the Misty Mountains".

Fauconnier and Turner [5] refer to the operations of representational contracting and stretching as compression and decompression. They suggest that human conscious thinking tends to compress complex paths of relationships into simpler relationships, despite remaining subconsciously knowledgeable about the complete paths.

J.M. Dodero, M. Palomo-Duarte, P. Karampiperis (Eds.): MTSR 2012, CCIS 343, pp. 214–226, 2012.

Machine processing requires a complete representation of paths. The RDF representation is a graph of nodes and arcs, where nodes are individuals (e.g. Bilbo, the Misty Mountains) or typed values (e.g. the string gold, the date 1341) and arcs are features (e.g. took place at, colour). A special feature relates an individual to its type[1].

The CIDOC Conceptual Reference Model (CRM), which is the ISO 21127 standard, is a core ontology in the cultural domain. The CRM [3] is rooted in the concept of events connecting things, concepts, people, time and place. Fig. 1 is a simplified graph of the circumstances in which Bilbo got the One Ring from Gollum. Individuals (e.g. 1341) are placed within a box labelled by type (e.g. Time-Span).

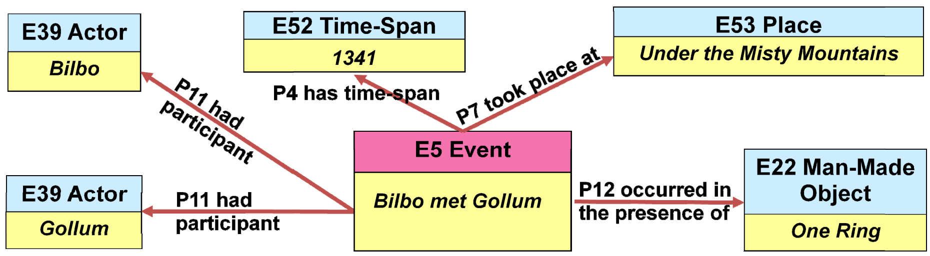

Fig. 1. Representation of the event where Bilbo met Gollum

The E5 Event "Bilbo met Gollum" is a keychain to linked entities. Talking about the One Ring discovery, humans will use a compression sentence "Bilbo found the One Ring under the Misty Mountains" that stands for paths of Fig. 1: Bilbo, an Agent, participated in an Event. The Event took place under the Misty Mountains, a Place. The Event occurred in presence of the One Ring, a Man-Made Object. To keep it simple, we have omitted the need to relate each individual to its identifier (an E41 Appellation) as well as the need to relate each appellation to its value, usually a string.

Forms facilitate data entry and retrieval and can be used to hide schema complexity from the end-user. Different instances of different classes can be presented in the same form, and any visual item can be associated with a schema property and displayed with a label that is meaningful to the end-user, e.g. replacing *P12 occurred in presence of* by *Objects used*. A semantic association [2] is a sequence of individual associations connecting two entities through intermediate entities, e.g. Bilbo and the Misty Mountains through a meeting event. Semantic associations can be discovered through schema processing or data mining techniques. Semantic associations provide the user with representational compression. Decompression aims to produce complex data behind the scenes. This paper presents two proposals for handling CRM-based compression and decompression: an extension based on Semantic MediaWiki; and WissKI, a system dedicated to cultural heritage documentation management.

---

[1] Unfortunately, there is no common agreement on concept names. In RDF, nodes refer to "things" or resources (represented by their URI reference) or constant values - called literals (represented by character strings). Arcs refer to predicates (properties). RDFS refers to types or "kinds of things" as classes, and the property rdf:type is used to indicate that a resource belongs to a class. For object oriented readers, nodes refer to instances and arcs refer to relationships. The instantiation relationship "is-a" links instances to their classes.

2 Background and Related Work

2.1 Introducing the CIDOC CRM

The CIDOC CRM is a formal ontology intended to facilitate the integration, mediation and interchange of heterogeneous cultural heritage information. The current version 5.0.4 [3] consists of a multiple inheritance hierarchy of 86 classes and 138 properties. A class (or entity) is identified by a number preceded by the letter "E" and followed by its name (e.g. E5 Event, E39 Actor). A class is a set of individuals (called class instances) that share common characteristics. A property is a binary relation between classes; the domain is the source class of the property, and range is the target class. A property can be interpreted in both directions, with two distinct, but related interpretations. Properties are identified by numbers preceded by the letter "P," and are named in both directions using verbal phrases in lower case - e.g. *P11 had participant (participated in)*. Property names should be read in their non-parenthetical form for the domain-to-range direction e.g. *P11 had participant*, and in parenthetical form for the range-to-domain direction - e.g. *P11i participated in*. An instance of a property is a link between an instance of its domain and an instance of its range. Classes and properties are specialized with sub-classes and sub-properties.

The CRM is based on a fundamental distinction [10] between persistent entities (endurants, continuants) and temporal entities (perdurants, occurrents). The CRM is rooted in four fundamental principles: endurant participation in an event, part-whole relation, reference information and classification [4] which are the most fundamental relationships connecting things, concepts, people, time and place.

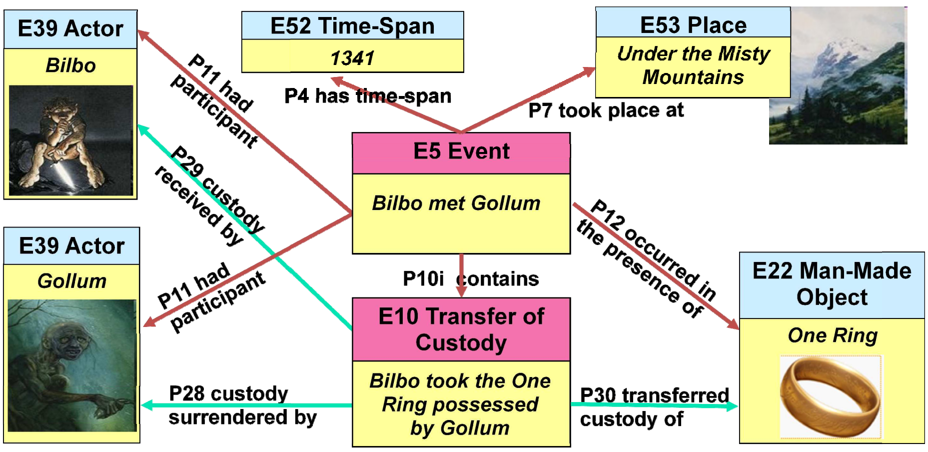

Fig. 2. Representation of a sub-event within the event in which Bilbo met Gollum

In order to provide the reader with comprehensive examples, in Fig. 2 we detail the semantic of the One Ring transfer of custody (a relevant case in cultural heritage management). The whole "Bilbo met Gollum" event *[P10i] contains* a part event, an instance of E10 Transfer of custody. Three properties specify the roles played by participants: *[P28] custody surrendered by* (Gollum's role), *[P29] custody received by* (Bilbo's role) and *[P30] transferred custody of* (One Ring role).

2.2 Semantic Associations: Paths of Knowledge

Paths between entities are semantic associations, a notion introduced by [1, 2]. Two entities e_x and e_y are semantically associated if they are semantically connected or semantically similar. A semantic connection between e_x and e_y is a sequence e_x, P_1, e_2, P_2, e_3, … e_{n-1}, P_{n-1}, e_y where e_i, $2 \leq i < n$, are entities and P_j, $1 \leq j < n$, are properties - whereas the range of P_j is the domain of P_{j+1}. Semantic similarity is based on the existence of two semantic connections having similar properties. In this paper, semantic connectivity alone is used to find semantic associations. Semantic association discovery has been an intensive research area for ten years now. [2] states that two kinds of paths can be discovered. The first kind is paths which are obvious from the schema. The second kind is those paths that exist at data level yet are not evident at schema level. We limit our work to the former, excluding data mining techniques.

As mentioned in the introduction, there is a semantic gap: a lack of coincidence between information one can extract from the knowledge base and user interpretation. Compression/decompression techniques are helpful in reducing this gap.

The compression operation replaces a semantic association (e.g. the individual One Ring, a Man-Made Object, has custody transferred through a Transfer of Custody event, part of an Event that took place under the Misty Mountains, a Place) with a direct relationship between two individuals (e.g. the One Ring, a Man-Made Thing, was found under the Misty Mountains, a Place). The end-user proceeds with the knowledge graph from one individual (e.g. the One Ring) acting as a focal point. Any semantic association starting from this individual is meaningful to the user.

Because the end-user is building and using paths intended to compress the representation, these paths can be used to store decompressed representations. If the compressed path is used to write information (e.g. the One Ring, a Man-Made Thing, was found [by Gollum] in the Gladden Fields, a Place), it requires a decompression and the creation of the full path (e.g. the One Ring, a Man-Made Thing, has custody transferred through a Transfer of Custody event, falling within an Event that took place in The Gladden Fields, a Place). However, there are several possible decompressions, i.e. there are several paths which yield the same compression; e.g. another path might be: the One Ring was present at an Event that took place in The Gladden Fields.

We need to reuse existing paths in larger paths, a grouping feature that helps to factorize shared section of paths. Unfortunately, as stated at [8] "*if both derived and base associations are taken into account, the number of distinct semantic associations (compare to acyclic paths) between two entities may increase drastically.*"

Shortcuts. CRM addresses the problem above through shortcuts. *"A shortcut is a formally-defined single property representing a deduction or join of a data path in the CRM. […] For each shortcut, the CRM contains in its schema the properties of the full data path explaining the shortcut."* [3, p. 11] Shortcuts implement the compression operation, although the CRM is aware that no decompression can be guaranteed: "*An instance of the fully-articulated path always implies an instance of the shortcut property. However, the inverse may not be true; an instance of the fully-articulated path cannot always be inferred from an instance of the shortcut property.*" [3, p.16]

3 Use Cases

In this section we present two approaches for compression and decompression operations and we discuss the advantages, limits and drawbacks of both approaches.

3.1 Ontology Management with SMW

Wiki (quick in Hawaiian) was defined in 1995 by its inventor, Ward Cunningham, as 'the simplest online database that could possibly work.' A wiki page is displayed according to instructions stored in the page "code" (as does HTML). For instance, linking a page to another is accomplished by surrounding the linked page name within double brackets. Semantic wikis add semantic annotations to the pages. Semantic MediaWiki (http://semantic-mediawiki.org/) is a free semantic extension for the free wiki engine MediaWiki (http://mediawiki.org/). We associate a property to a value or to another page by preceding the value or the linked page with the property name.

Fig. 3 presents the SMW code resulting from the semantization of the introductive compression sentence. It uses the CRM as the reference ontology and yields a subgraph of Fig. 1 from a focal point: an E5 Event identified by "Bilbo met Gollum".

```
As [[P11 had participant::Bilbo]] groped along the dark
tunnels of the [[P7 took place at::Misty Mountains]], he
found the [[P12 occurred in the presence of::One Ring]]
lying on the ground and slipped it into his pocket. By a
subterranean lake [[Bilbo]] met [[P11 had partici-
pant::Gollum]], the creature who had lost the [[One
Ring]].
[[Category: E5 Event]]
```

Fig. 3. Semantic MediaWiki code fragment for the E5 Event "Bilbo met Gollum"

Page names identify pages within a wiki. Each page has a URIref (Uniform Resource Identifier reference), built from a base (the wiki URL) and a relative URI (the page name). Each property or class has its own wiki page, used to build its URIref.

Any semantic annotation in the page will produce an RDF triple with the page URIref as subject, the property URIref as predicate, a literal value or an object URIref as object. Each time a page is updated, SMW regenerates that page's RDF triples.

Semantic search allows users to write queries using the same syntax as annotations; e.g. *[[E5 Event]][[P7 took place at::Misty Mountains]]* will retrieve all events located in the Misty Mountains. Semantic queries might be used to implement semantic associations, and are fairly easy for the end-user to understand. For instance, the query *P4i is time-span of.P10i contains.P30 transferred custody of. One Ring* retrieves 1341, a Time-Span that is the time-span of [an Event] that contains [an Event] that transferred custody of [the individual identified by] One Ring.

When a page is the subject of a predicate towards another page (e.g. <Bilbo met Gollum, P11 had participant, Bilbo>, the object should be the subject of an "inverse" triple using the inverse property towards the former subject (e.g. <Bilbo, P11i participated in, Bilbo met Gollum>). It requires a synchronized update of both triples. Inverse triples can be implemented with queries instead of annotations (e.g. in Bilbo's page, the query *P11 had participant::Bilbo* retrieves the page Bilbo met Gollum).

Semantic annotations require a disciplined editing process that can be supported through the use of templates. MediaWiki templates have immense value for normalizing and simplifying display (such as Wikipedia Infoboxes). Semantic forms are a method for including the semantic annotations through MediaWiki templates and generated forms. Each form field is associated with a semantic property and labelled with a meaningful name. A field hosts single or multiple values or object references, according to the cardinality and type of the underlying property. Where the form is used, semantic constructs are used consistently and do not require schema knowledge.

Knowledge representation in SMW is RDF-based and is lacking in OWL characteristics. To benefit from inference and reasoning features, the triple set has to be exported and processed by a post-processor software. The post-processor performs URI alignment to a reference namespace, produces inverse triples and is dedicated to the processing of any operations requested to produce a sound and complete triple set. We developed a simple post-processor using XSLT for validation purposes only.

3.2 Case study: The Use of Shortcuts

Literature reports numerous cases in which semantic wikis for knowledge management are used [9, 11] with some experimentation taking place in the Cultural Heritage domain [15]. For 4 years now, we have been using CRM-based semantic wikis for labs and group projects of a semantic web course for STEM students. Each group is using SMW to build and populate a small ontology about the story of a film, book or

```
The One Ring was created by the Dark Lord [[P14 carried
out by::Sauron]] during the [[P4 has time-span::Second
Age]] in order to gain dominion over the free peoples of
Middle-earth. The Ring seemed simply to be made of [[P9
consists of::gold]], but was impervious to damage. Placed
in fire, the ring displayed an
[[P102 has title::inscription]] in the Black Speech.
Translated, the words mean:
 One ring to rule them all, one ring to find them,
 One ring to bring them all and in the darkness bind
them.
[[Category: E22 Man-Made Object]]
```

Fig. 4. End-user SMW code fragment about the One Ring

biography. Although stories are naturally event-oriented, we have observed that (i) students tend to produce Wikipedia-like pages depicting endurants rather events; (ii) students frequently misuse properties, and especially property domains. Fig. 4 shows a typical SMW code excerpt that students might produce depicting the One Ring.

End-User Representation. Typically, an end-user will relate a physical or conceptual object directly to its creator. Indeed, in CRM event-oriented ontology, going through an E12 Production or E65 Creation event is required.

As several part-whole relationships exist in the CRM, end-users might use a wrong property (e.g. linking Things with *P9 consists of*, intended for temporal entities). A user might be faced with subtle differences in part-whole relations. For instance, an E18 Physical Thing may use *P46 is composed of* to relate to its components, if they are E18 Physical Thing; but should use *P45 consists of* to relate to its E57 Material. From our experience, end-users rarely choose the right part-whole relation.

The end-user might perceive parts indissociable from the whole. She will include parts description in the page depicting the whole, where this inclusion denotes the part-whole relationship. Indeed it requires the instantiation of parts, inside their proper class, and the instantiation of the part-whole and whole-part (inverse) relationships.

CRM-Compliant Representation. Fig. 5 depicts the correct SMW code compliant with the CRM ontology. The semantic association between an object and its creator should be replaced by a path denoting the necessity of going through an E12 Production event. Indeed, the semantic association between an object and its creation date should be replaced by a path starting with the same E12 Production event, but ending with an E52 Time-span rather with an E39 Actor. This illustrates the necessity of sharing paths, so as to avoid the instantiation of two Events denoting a single event.

```
The One Ring was created by the Dark Lord [[P108B was
produced by.E12 Production.P14 carried out by::Sauron]]
during the [[P108B was produced by.E12 Production.P4 has
time-span::Second Age]] in order to gain dominion over
the free peoples of Middle-earth. The Ring seemed simply
to be made of [[P45 consists of::gold]], but was impervi-
ous to damage. Placed in fire, the ring displayed an
[[P128 carries::inscription]].

[[Category: E22 Man-Made Object]]
```

```
The inscription appearing on the [[P128i is carried
by::One Ring]] is in [[P72 has language::the Black
Speech]]...
[[Category: E33 Linguistic Object]]
```

Fig. 5. CRM-compliant SMW code fragment about instances involved in Fig. 4

Because the inscription is an E33 Linguistic Object[2], distinct from the One Ring, it should be treated as a separate part with its distinct type (due to SMW limits, the only way of doing so is to create a separate page for the inscription), and the part should be related to the whole with a special part-whole relation *P128 carries (is carried by)*. The part might have its own properties (e.g. *P72 has language*::the Black Speech).

Discussion. Clearly, semantically correct annotations cannot be produced without solid knowledge of the CRM. Contributors can be provided with well-designed forms, where paths are replaced by shortcuts and displayed with meaningful labels (e.g. creator, creation date). SMW does not offer path group management.

A drawback of this approach is that shortcuts are not handled inside SMW. When a semantic connection between entities is required, ontology managers have to create the SMW property corresponding to the shortcut with the path as a name (e.g. *P108B was produced by*.E12 Production.*P4 has time-span* to connect a Man-Made Object to its production date). Shortcuts will be processed later on by the post-processor, which will decode the full path and produce the required triples. Hence, a strong limitation of using shortcuts in SMW is that the underlying path does not exist in the triple set and that searches cannot, therefore, use the missing triples.

3.3 The WissKI Approach

WissKI [7] was a research project in the cultural heritage domain, funded by the German Research Council (DFG) from 2009 to 2012. The name "WissKI" is a German acronym for "Scientific Communication Infrastructure" (Wissenschaftliche KommunikationsInfrastruktur). The project developed a software infrastructure which enhances the Drupal CMS (http://drupal.org/) for handling ontologies and semantic data. The software is open source and available via Github (https://github.com/wisski) or the project website (http://wiss-ki.eu/). The WissKI system requires an OWL-DL compatible ontology based on description logics and therefore suitable for automatic processing by machines. Thus the system benefits from calculations performed by reasoning mechanisms, e.g. automatic calculation of inverse properties. We use an OWL-DL implementation of the CRM, available at http://erlangen-crm.org/.

In the WissKI system, data is acquired either by forms or by text annotation in free texts via a WYSIWYG editor. Both input methods store the data in the same way in a triple store backend. The system aims to provide the user with concepts and relationships close to her representation of reality. At first concepts are identified, e.g. actors, time-spans, places, events and man-made objects. Then relationships are identified. Here WissKI suggests a new approach: focusing on semantic associations between endurants and providing the user with automatic generation of event-oriented paths.

If we consider Fig. 2, from the point of view of the "One Ring" there are four basic facts related to the custody transferred through an E10 Transfer of custody event: Gollum *[P28i] surrendered custody through* [an Event] while Bilbo *[P29i] received custody through* [an Event]. The E10 Transfer of custody event *[P10] falls within* an

[2] Formally, the text of an E33 Linguistic Object is documented with P3 has note: E62 String.

E5 Event that *[P7] took place at* under the Misty Mountains and *[P4] has time-span* 1341. Each fact can be represented by a semantic association, e.g. relating One Ring with 1341 is the path One Ring.*P30i custody transferred through.* Bilbo took the One Ring possessed by Gollum.*P10 falls within.* Gollum met Bilbo.*P4 has time-span.*aTime-span.*P1 identified by.*anAppellation.*P3 has note.*1341. Paths are instances of constructs (called "ontology paths" in the WissKI system) which act as stencils for the data. For instance, the ontology path between One Ring and 1341 is E22.*P30i*.E10.*P10*.E5.*P4*.E53.*P1*.E41.*P3*.E62.

Fig. 6. The WYSIWYG editor in the WissKI system

Each ontology path can be attributed a name, which makes it easier for humans to read; e.g. the ontology path above could be called "Time of Transfer". A complete set of concepts and ontology paths can be defined prior to using the system, or new concepts and paths can be added on-the-fly. The semantic annotation of free text is performed within the WYSIWYG editor (see a snapshot in Fig. 6) in a two-step process. Individuals referred to in the text (e.g. One Ring) have to be tagged by the user with the right concept (e.g. Object). Then, ontology paths can be instantiated: whenever the end-user selects any tagged individual, the system offers potential ontology paths and the user is able to select the right one (e.g. One Ring had a time of transfer 1341). Ontology path use is immediately processed by the WissKI system and the triple chain is generated. Ontology paths can be used in forms instead of ontology properties, thus enabling the user to display or enter data in an endurant-centric point of view while in the backend, the full power of semantic data slumbers.

3.4 Case study: The Use of the Pathbuilder

The WissKI component called Pathbuilder is a core utility for the creation, deletion and management of ontology paths. Once an ontology (e.g. the CRM) is loaded in the system, the administrator can navigate using dropdown fields through the ontology and is thus assisted in the construction of ontology paths. Paths have attributes: names, I/O look and feel, mandatory input and disambiguation.

Individuals are identified by appellations. The CRM uses the property *P1 is identified by* (and its sub-properties) to connect an individual to one or several instances of E41 Appellation (and its sub-classes) and an appellation instance is related to its primitive datatype value: a string, a date, etc. There might therefore be several individuals associated with a single value. When referring to an individual having that value, the user might want to use an existing individual or create a homonym. This is the purpose of the disambiguation feature: the system uses the backmost part of the ontology path as a search pattern in the triple store. User input yields the search condition. Whenever a match is found in the triple store, the user is prompted as to whether she wants to use the existing instance from the triple store or whether she wants to create a new one.

The endurant-centric approach uses a set of ontology paths starting from a single concept and leading on to different concepts. The set shall be divided among subsets sharing a common path, e.g. all paths mentioned in section 3.3 share E22 Man-Made Object.*P30i custody transferred through.*E10 Transfer of Custody. The system can be instructed to use the same instances for this section of the path when it generates the path instantiation. This feature is called "grouping of ontology paths". All ontology paths which are part of a group share the same root. Path groups have attributes, e.g. names. Forms and text input in the WissKI system are automatically enabled and generated for top level groups - those which are not part of other groups.

A CRM-Compliant End-User Representation. Recall the facts about the One Ring depicted in Fig. 4 and Fig. 5. An administrator defines a set of ontology paths from which for end-users can select their annotations. Table 1 gives some examples. Each path ends with a primitive datatype to handle primitive values, which is skipped here for the sake of simplicity.

Table 1. Examples of ontology paths related to E22 Man-Made Object instances

Path name	Domain / Range	Path structure
inscription	E22 Man-Made Object / E33 Linguistic Object	E22.*P128 carries*.E33
creator	E22 Man-Made Object / E82 Actor Appellation	E22.*P108i was produced by*.E12.*P14 carried out by*.E39.*P131 is identified by*.E82
creation date	E22 Man-Made Object / E49 Time Appellation	E22.*P108i was produced by*.E12.*P4 has time-span*.E52.*P78 is identified by*.E49
language inscription	E22 Man-Made Object / E56 Language	E22.*P128 carries*.E33. *P72 has language*.E56

In Fig. 7, we use these paths to produce an annotated text corresponding to Fig. 5. CRM properties (e.g. *P128 carries*) are very simple paths (from the property domain to its range), but can be customized to have another name (e.g. inscription) or other path attributes. Paths named *creator* and *creation date* start with the same sub-path (E22.*P108i*.E12) and shall be grouped (to share the same E12 Production instance). Handling a whole and its part is performed (as in SMW) in two steps: (i) writing, annotation and storage of the text related to the whole (e.g. One Ring) that will contain part-whole instantiations to its parts (e.g. the reference to an inscription) then attributing the right types to parts (which will instantiate the parts); (ii) clicking on each part and writing, annotating and storing the text related to each part.

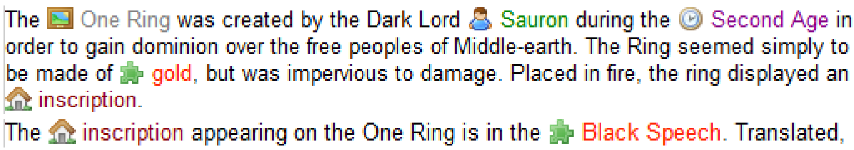

Fig. 7. WissKI annotations corresponding to Fig. 5

4 Lessons Learned and Perspectives

4.1 Identifiers and Names

Typically, a (semantic) wiki uses page names as individuals' identifiers and there cannot be two pages with the same name. When several pages use the same name, a disambiguation word should be added to the name. However the CRM recognizes the fact that a distinction must be made between objects and their appellation. In WissKI, individual URIS are built from the individual's class and a generated unique identifier, e.g. E22_123 might be the relative URI of the individual identified by the appellation "One Ring", that obviously enables the use of the same name for different individuals.

This approach can be applied in a wiki, e.g. using E22_123 as a page name and the string "One Ring" as the display name. If we want to avoid each knowledge management system using its own isolated knowledge silo, we must recognize different URIs as being related to the same individual. This co-reference issue, is an active research topic. Exchanging or sharing triple sets between systems will require post-processing of the set in order to align URIs in a common agreed namespace.

4.2 End-User Usability of Instantiation, Compression and Decompression

Within SMW, the user has to create a wiki page for every individual she wants to instantiate and she has to locate the page in the right category (class). Instantiating the association between individuals is performed in the subject's page, and its inverse in the other page. In the WissKI free text editor the user has to select the type of the individual which is created if the automatic annotation process cannot find it. The user then asserts an ontology path to an individual which the text describes or refers to.

SMW templates using shortcuts or WissKI ontology paths provide the end-user with a friendly-representation of compressed paths of knowledge. WissKI uses SPARQL queries to implement ontology paths, so that the fully-articulated path and the ontology path are always synchronised. The same feature can be implemented in SMW when shortcuts are implemented using SMW queries.

When the user selects an ontology path in the WissKI editor or uses it within a form, the fully-articulated path is generated on-the-fly and stored in the triple store - whereas in SMW, decompression is performed by the post-processor after an export of the triple set (which must be re-imported if it is to be accurate).

4.3 Perspectives: Looking for Event Patterns

Almost all the shortcuts or ontology paths we encountered in our practice involve events. Following analysis of existing event models, [12] mention an agreement for six aspects and discuss a pattern for each aspect. [7] uses semantic documentation templates and an XML-based query language to create a documentation model and build flexible user interfaces for accessing and editing the documentation. In our view, whenever the end-user focuses her representation on an endurant participating in an event, this focus should be reflected by a set of semantic associations toward other endurants. When mereological relationships (part-whole) exist between events, semantic associations starting either from the part or from the whole are useful. And we could suppose that useful semantic associations can be deduced from other patterns. One perspective of this work is to use patterns and the work in [14] to generate a set of shortcuts/ontology paths for intuitive querying of CRM based repositories.

5 Conclusion

We present two prototypical approaches, SMW and WissKI, both of which aim to reduce the semantic gap between end-user representation and a complex event-based ontology. We focus on the compression and decompression of semantic associations through the use of shortcuts in the former approach, and ontology paths in the latter. Representational contracting is based on queries, and works fairly well in both approaches. Stretching end-user representation to generate fully-articulated paths is performed on-the-fly with WissKI whereas a shortcut post-processor has to be added to SMW. Considering base associations as well as derived ones is a difficult issue, as the number of possibilities may increase drastically. Thus, a skilled shortcut/ontology path design is required, to ensure end-user usability. An interesting perspective is to associate shortcut/ontology path generation with the recognition of event patterns.

References

1. Aleman-Meza, B., Halaschek, C., Arpinar, I.B., Sheth, A.: Context-aware semantic association ranking. In: Proc. of 1st Int. WS. on Semantic Web and Databases, pp. 33–50 (2003)
2. Anyanwu, K., Sheth, A.: The ρ operator: Discovering and ranking associations on the semantic web. SIGMOD Record 31, 42–47 (2002)
3. CIDOC CRM, `http://www.cidoc-crm.org/official_release_cidoc.html`
4. Doerr, M., Kritsotaki, A.: Documenting Events in Metadata. In: Proc. of 7th International Symposium on Virtual Reality, Archaeology and Cultural Heritage, pp. 1–5 (2006)
5. Fauconnier, G., Turner, M.: The Way We Think: Conceptual Blending and the Mind's Hidden Complexities. Basic Books, New York (2002)
6. Häyrinen, A.: Towards Semantic Modelling of Cultural Historical Data. In: Proc. of 12th Int. Conf. Information Modelling and Knowledge Bases, pp. 312–320 (2011)

7. Hohmann, G., Schiemann, B.: An ontology-based communication system for cultural heritage: Approach and progress of the WissKI project. In: Proc. of Scientific Computing & Cultural Heritage (2009)
8. Jämsen, J., Niemi, T., Järvelin, K.: Derived types in semantic association discovery. Journal of Intelligent Information Systems 35(2), 213–244 (2010)
9. Krötzsch, M., Vrandecic, D., Völkel, M., Haller, H., Studer, R.: Semantic Wikipedia. Journal of Web Semantics 5, 251–261 (2007)
10. Masolo, C., et al.: WonderWeb Deliverable D17. The WonderWeb Library of Foundational Ontologies and the DOLCE ontology (2002)
11. Millard, I., Jaffri, A., Glaser, H., Rodriguez-Castro, B.: Using a Semantic MediaWiki to Interact with a Knowledge Based Infrastructure. In: Proc. of 15th Int. Conf. on Knowledge Engineering and Knowledge Management (2006)
12. Scherp, A., Frantz, T., Saathoff, C., Staab, S.: A core ontology on events for representing occurrences in the real world. Multimed. Tools Appl. 58, 293–331 (2012)
13. Tolkien, J.R.R.: The Lord of the Rings - The Fellowship of the Ring. George Allen & Unwin (1954)
14. Tzompanaki, K., Doerr, M.: Fundamental Categories and Relationships for intuitive querying CIDOC-CRM based repositories. ICS-FORTH/TR-429 (2012)
15. Witte, R., Gitzinger, T., Kappler, T., Krestel, R.: A Semantic Wiki Approach to Cultural Heritage Data Management. Language Technology for Cultural Heritage Data (2008)