Paths and shortcuts in an event-oriented ontology

Vincent Ribaud, Mark Fichtner

Some slides are adapted from a presentation by Martin Doerr and Stephen Stead at Imperial College, London, UK, May 22, 2009
Agenda

- Introduction – Bilbo Baggins found the One Ring
- Background: the CIDOC CRM, semantic associations
- Compression and decompression use cases
- Lessons learned
Whom, when and where the Ring came into Bilbo’s custody?

Google “bilbo ring”

http://en.wikipedia.org/wiki/One_Ring
After Gollum hid under the Misty Mountains for hundreds of years, the Ring abandoned him, and Bilbo Baggins picked it up.

http://www.tuckborough.net/bilbo.html
As Bilbo groped along the dark tunnels, 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.

Tolkien’s excerpt

The Lord of the Rings Allen & Unwin 1954
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..

Sum up

Bilbo took the One Ring from Gollum in 1341 under the Misty Mountains.
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, e.g. a string.
Introducing the CIDOC CRM version 5

86 classes

- 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.

138 properties

- A property is a binary relation between classes; the domain is the source class of the property, and the range is the target class.
- Properties are identified by numbers preceded by the letter “P,” and are named in both directions.
  - e.g. E5 Event P11 had participant (participated in) E39 Actor.
- Domain-to-range direction e.g. E5 Event P11 had participant (participated in) E39 Actor.
- Range-to-domain direction e.g. E39 Actor P11i participated in E5 Event.
- An instance of a property is a link between an instance of its domain and an instance of its range.

Distinction between persistent entities (endurants, continuants) and temporal entities (perdurants, occurrents).

4 fundamental principles: endurant participation in an event, part-whole relation, reference information and classification.
Semantic of the One Ring transfer of custody

A semantic connection from \( e_x \) to \( e_y \) is a sequence \( e_x, P_1, e_2, P_2, e_3, \ldots, e_{n-1}, P_{n-1}, e_y \) where \( e_i, 2 \leq i < n \), are entities and \( P_j, 1 \leq j < n \), are properties - \( P_j \) range = \( P_{j+1} \) domain.

A semantic gap:

the One Ring, a Man-Made Thing, was found under the Misty Mountains, a Place vs. the 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.
Paths and shortcuts

Compression

◆ Any individual (e.g. the One Ring) may act as a focal point.
◆ Any semantic association (any path) starting from this individual is meaningful, and the end-user may be aware only from the path extremities.

Decompression

◆ *The One Ring, a Man-Made Thing, was found [by Gollum] in the Gladden Fields, a Place.*
◆ Paths can be used to store decompressed representations.
◆ The One Ring, a Man-Made Thing, was present at an Event that took place in The Gladden Fields, a Place.
◆ *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.*

CRM 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.”
◆ 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."
Two systems for the end-user

- **Motivation**
  - Instruments and usage heritage management -> experience management
  - German State project for a “Scientific Communication Infrastructure”

- **A CRM-based semantic wiki**
  - Wiki (quick) ‘the simplest online database that could possibly work’
  - Page “code” instructs the display and the structure of content. Semantic wikis add semantic annotations to the pages, including embedded links.
  - Semantic wikis can be shaped with a set of classes, properties, queries, templates, forms, based on a CRM-based kernel with a RDF representation
  - Shortcuts are provided and can be defined by end-users, but it requires a post-processing for decompressing the paths

- **WissKI - Wissenschaftliche KommunikationsInfrastruktur**
  - 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 Internet.
  - WissKI uses an OWL-DL compatible ontology suitable for automatic processing by machines, e.g. automatic calculation of inverse properties.
  - Data is acquired either by forms or by text annotation in free texts via a WYSIWYG editor.
Resource Description Language (RDF)

- RDF models (meta)data as 3-uplet (triples): `<subject, property, value>`
- A resource (identified by a URI) has a property which has a value
- The value can be given literally or identified by a URI

- `< J.R.R. Tolkien, Birth date, 3 January 1892 >` or `< The Hobbit, Creator, J.R.R Tolkien >`
Each page is a « triple tank »
- The Factbox depicts triples having the page as a subject
- Most properties have an inverse property, but RDF manages it as 2 separate triples

Semantic search
- Specify 1. Which pages to select 2. What information to display about those pages
- An interface that assists users with creating and executing semantic queries
- Inline queries dynamically include query results into pages
A semantic wiki

- **Constructs**
  - **Categories ↔ Classes**
    - A category is a set of individuals
    - A category groups pages sharing characteristics
  - **Pages ↔ Instances**
    - A page describes « something »
    - « something » is an individual, an occurrence, an instance
    - « something » belongs to one or several categories
  - **Properties**
    - Attributes: properties having a data type type String, Date, …
    - Typed links: properties linking to a page E53 Place, E39 Actor, … type Page

- **Textual part**
  - **Annotated and hyperlinked text**
    - Bilbo is called [[P1 is identified by::Elf-friend]]
    - Bilbo Baggins is a [[P2 has type::Hobbits]]
  - **May or may not be structured**

- **Facts about**
  - **Triple RDF < subject, property, value >**
    - `< Bilbo Baggins, P2 has type, Hobbits >`
Case study: a semantic web course for STEM students

Goal: to populate a small ontology about the story of a film, a book, or a bio

- Each year, around 20 “groups” of students from 1 to 4 people
- Students tend to produce Wikipedia-like pages depicting endurants rather events
- Students frequently misuse properties, and especially property domains

End-user representation

- Endurant-endurant relationships
  *The One Ring was created by [[P14 carried out by::Sauron]]*
  Typically, an end-user will relate a physical or conceptual object directly to its creator.
  Indeed, in CRM, going through an E12 Production or E65 Creation event is required.

- Complexity of the part-whole relationship
  *The Ring displayed an [[P102 has title::inscription]] in the Black Speech*
  End-users rarely choose the right part-whole relation.
  The end-user might perceive parts inseparable from the whole.
  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

- Semantically correct annotations is requiring a solid knowledge of the CRM
  *The One Ring was created by [[P108B was produced by.E12 Production.P14 carried out by::Sauron]]. created by is a shortcut of the path P108B.E12 Production.P14*

- Defining, sharing and processing the paths
  Shortcuts will be processed later on by the post-processor, which will decode the full path and produce the required triples.
Features of the WissKI system

- Shaping the system: a two-step approach
  - Concepts definition: actors, time-spans, places, events and MM-objects.
  - Ontology path definition: relate endurants with a semantic association, e.g. relating a Man-Made Object (One Ring) with a Place (Misty Mountains) is the path One Ring. P30i custody transferred through. Bilbo took the One Ring possessed by Gollum. P10 falls within. Gollum met Bilbo. P7 took place at a Place. P1 identified by an Appellation. P3 has note Misty Mountains.
  - Group: a set of ontology paths from the same concept to different concepts

- Endurant-centric view: a two-step annotating process
  - Tagging names with concepts: actors, time-spans, places, events, ...
  - Ontology path instantiation: computer-aided and immediately processed
  - Disambiguation: several individuals might be associated with a single value. The system assists the user to decide to use an existing individual or create a homonym.

As Bilbo groped along the dark tunnels of the Misty Mountains, he found the One Ring lying on the ground and slipped it into his pocket. By a subterranean lake Bilbo met Gollum, the creature who had lost the One Ring.
Features of the WissKI system (2)

Path management

- Paths attributes: names, I/O look and feel, manual filling, disambiguation
  - e.g. I/O look and feel defines different input and output behavior.

Handling the part-whole relationship

- 1 - Write, annotate and store the text related to the whole (e.g. One Ring) that will contain part-whole instantiations to its parts (e.g. towards an inscription) then attribute the proper types to parts (parts instantiation)
  - 2 - Click on each part and write, annotate, store the text of each part.
Lessons learned and perspectives

Identifiers and names

 A wiki uses page names as identifiers. 2 pages cannot have the same name. But a distinction must be made between objects and their appellation.
 In WissKI, URIs are built from the individual's class and a generated unique identifier, that enables the use of the same name for different individuals.
 If we want to avoid isolated knowledge silos, we must recognize different URIs as being related to the same individual. This is called co-reference issue.

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 proper category.
 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. Then the user only has to take one action to assert a property or an ontology path to an individual which the text describes or refers to.

Perspectives: looking for event patterns

 When mereological relationships (part-whole) exist between events, semantic associations starting either from the part or from the whole are useful.
 Use patterns to generate a set of shortcuts for intuitive querying CRM based repositories.
Conclusion

- Reduce the semantic gap between end-users representation and a complex event-based ontology
  - 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 is working pretty-well in both approaches.

- Stretching the end-user representation
  - Generate fully-articulated paths is performed on-the-fly with WissKI whereas SMW has to be completed with a shortcut post-processor.
  - Semantic wikis require too complex actions for creating and linking instances whereas the WissKI free text editor offers one-click feature, much more convenient.
Limits of implementing the CRM in a semantic wiki

- CRM fully objet-oriented
  - The CRM is using properties between objects
  - To simplify, we need to have datatype properties for literal values
  - A S before a property: a shortcut, mostly converting to datatype
    - P7 took place at: E53 Place ↔ SP7 took place at: String

- Inverse properties
  - Example of P14 carried out by (performed)
    - Domain: E7 Activity P14 carried out by Range: E39 Actor
    - Domain: E39 Actor P14B performed Range: E7 Activity
  - Locality principle
    - We have to manually update former or future pages related with inverse property

- Type control for domain and range
  - For object property, the wiki provides only the type Page
  - We present the property in the category associated with the domain
  - And the inverse property in the category associated with the range
  - No automatic control within the wiki
Question: whom, where and when did Biblo get the One Ring?

- Google “bilbo find ring”:
  - [http://www.archivesdegondor.net/hobbit/chap5.php](http://www.archivesdegondor.net/hobbit/chap5.php)
  - [http://membres.multimania.fr/kingconan/tolkien-livre_raconte_01.htm](http://membres.multimania.fr/kingconan/tolkien-livre_raconte_01.htm)

- Display the place of (that witnessed) activities that used the One Ring?
  ```
  {{#ask:[[Category:E7 Activity]]
  [[SP16 used specific object ::One R|ng]]
  |?SP7 took place at}}
  ```

- Display the date of (that is time-span of) activities that used the One Ring?
  ```
  {{#ask:[[Category:E7 Activity]]
  [[SP16 used specific object ::One R|ng]]
  |?SP4 has time-span }}
  ```

- Display the participants to (that performed) activities that used the One Ring and with Bilbo as a participant?
  ```
  {{#ask:[[Category:E7 Activity]]
  [[SP16 used specific object ::One R|ng]]
  [[P14 carried out by::Bilbo]]
  |?P14 carried out by }}
  ```
A semantic wiki (2)

```plaintext
{{Tolkien character
| character_name = Bilbo Baggins
| character_alias = Ring-bearer, Elf-friend ...
| image_character =
| character_race = Hobbits
| Book(s) = The Hobbit, The Lord of the Rings }}
```

- **Structured part**
  - Based on a semantic template
  - Field are related to semantic properties

- **Category**
  - Associated with template and form
  - Defining a set of properties that may be used to annotate pages of the category

- **Facts about**
  - Coming from textual and structured parts
  - Each page is a small set of triples
  - All sets together are a semantic network
  - Searches work on the network

Bilbo Baggins (in comfortable middle age at 50 years old) was hired in spite of himself as a burglar by Gandalf and thirteen dwarves led by their king Thorin Oakenshield on a quest to reclaim the Lonely Mountain and its treasure from the dragon Smaug. The adventure takes Bilbo and the companions through the wilderness, to the haven of Rivendell, across the Misty Mountains and the black forest of Mirkwood, to Esgaroth in the middle of Long Lake, and eventually to the Mountain itself. Here, after the dragon is killed and the Mountain reclaimed, the Battle of Five Armies takes place.

Category: E39 Actor

**Facts about Bilbo Baggins**

- P107B is current or former member of Company of dwarves
- P11B participated in Quest of Erebor, and Battle of Five Armies
- P1 is identified by Ring-bearer, Elf-friend, and Bilba Labingi
- P2 has type Hobbits
- P48 has preferred identifier Bilbo Baggins
- P70B is documented in The Hobbit, and The Lord of the Rings
... based on an ontology

- An ontology with classes and properties
  - A class is a set of items sharing common characteristics
  - A property is used to define a relationship between classes
    Domain: departure set
    Range: arrival set

- CIDOC CRM, international standard ISO 21127:2006
  - event-centered
  - around 90 classes and 150 properties
  - ontological kernel with 40 classes and 30 properties
  - extensible