ACM Web Science Conference 2011, Koblenz, June 2011

Digital Hermeneutics: Agora and the Online Understanding of Cultural Heritage

Chiel van den Akker*, Susan Legêne
Faculty of Arts - History Department
VU University Amsterdam
De Boelelaan 1105
1091 HV Amsterdam
c.vandenakker@let.vu.nl

Marieke van Erp, Lora Aroyo, Roxane Segers, Lourens van der Meij, Jacco van Ossenbruggen, Guus Schreiber, Bob Wielinga
Computer Science Department
VU University Amsterdam
marieke@cs.vu.nl

Johan Oomen
Netherlands Institute for Sound and Vision
joomen@beeldengeluid.nl

Geertje Jacobs
Rijksmuseum Amsterdam
G.Jacobs@rijksmuseum.nl

ABSTRACT

Cultural heritage institutions are currently rethinking access to their collections to allow the public to interpret and contribute to their collections. In this work, we present the Agora project, an interdisciplinary project in which Web technology and theory of interpretation meet. This we call digital hermeneutics. The Agora project facilitates the understanding of historical events and improves the access to integrated online history collections. In this contribution, we focus on defining and modeling prototypical object-event and event-event relationships that support the interpretation of objects in cultural heritage collections. We present a use case in which we model historical events as well as relations between objects and events for a set of paintings from the Rijksmuseum Amsterdam collection. Our use case shows how Web technology and theory of interpretation meet in the present, and what technological hurdles still need to be taken to fully support digital hermeneutics.

Categories and Subject Descriptors

H.3.3 [**Information Storage and retrieval**]: Information Search and Retrieval; H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces—*Theory and models*

General Terms

Design, Theory

*Corresponding Author

Keywords

Digital hermeneutics, online cultural heritage, collection enrichment

1. INTRODUCTION

The Web has offered cultural heritage institutions and their public a medium rather than a product [8], changing their traditional task from *information interpreters* to that of *information providers* [6]. As information providers, cultural heritage institutions therefore need to think about the way they provide meaningful access to and presentation of their collection. They furthermore need to think about how users may enrich their collection descriptions with their personal perspectives. The information flow from heritage institutions to the public thus needs to be supplemented with an information flow from the public to the heritage institution. This two-way information flow, where the producers and interpreters of Web applications meet, marks the Web as a place of *dialogue* [17].

Digital hermeneutics is the encounter of hermeneutics and Web technology [1]. Traditionally, hermeneutics has been a theory of interpretation in the humanities. With the advent of the Web, this theory needs to be amended to account for the interpretation of information in a digital environment. Digital hermeneutics forms the appropriate context to think about providing access to and interpretation of online collections of cultural heritage institutions. Its main aim is to investigate the relation between the human interpretation process and Web applications supporting that interpretation process.

The main building blocks in the interpretation process of history and collection objects are historical events and the relations between these events and the objects. A collection object on its own, has no meaning; by adding events to the object's description we provide its historical context. This gives the object meaning and makes it possible to interpret it. A single historical event often gives only a part of the whole historical context of an object, therefore narratives are important in the interpretation process. A narrative is formed by a chain of historical events. In principle, any se-

quence of historical events can form a narrative, however, we define a set of event-event-relationships from which narratives follow that are meaningful from a historical perspective. We call these narratives *proto-narratives*, as they are key to the interpretation process of users.

In this paper we aim to answer two questions: *(i)* what object-event-relations and event-event-relations support and enhance access to and interpretation of cultural heritage collections? and *(ii)* how can we model historical event descriptions to facilitate this access and interpretation? This work is carried out in the framework of the Agora project, an interdisciplinary collaboration between the History and Computer Science departments at the VU University Amsterdam, the Rijksmuseum Amsterdam (RMA) and the Netherlands Institute for Sound and Vision (S&V). The project develops methods and techniques to support searching, browsing, interpretation, and representation of historical events online.

In this paper, we illustrate the notion and implementation of digital hermeneutics in a real use case from the Rijksmuseum Amsterdam. In the remainder of this paper, we first introduce our use case (Section 2). Then, in Section 3, we present our theory of digital hermeneutics and show how its components can be represented in an event model. The use case is revisited in Section 4 where we describe how historical descriptions can be modeled and what hurdles this puts up for automatic methods. In Section 5, we show how collection objects can be presented within their historical context in a prototype of an event-driven collection access system. We conclude with an outlook on future work (Section 6).

2. USE CASE: A CHILD IN WAR

To illustrate our vision of digital hermeneutics, we have used the Dutch version of the book "A Child in War: Mohammed Toha paints Yogyakarta 1948-1949"[1]. This book shows 44 paintings of an Indonesian child about the Dutch occupation he witnessed in Yogyakarta in 1948 and 1949, along with descriptions of the paintings, commentary about the events in the colonial war and background information about the people mentioned and the political situation.

Each of the paintings in the book is represented by a record in the Rijksmuseum Amsterdam collection database. This database is modeled using VRA-Core (ver. 3)[2] and contains the usual meta-data information about a museum object: its object-id, what type of object is it, when was it made, by whom, during which exhibitions was it displayed, and a short textual description of what can be seen in the object. In some cases, the textual descriptions in the database provide information about the historical context of the object. However, in most cases, this information is minimal at best, and furthermore, it is not modeled in such a way that it is easily accessible. In the current database view, for instance, one cannot select all objects that depict a certain actor that was involved in a particular event or series of events. For example, in our "A Child in War" use case, it would be interesting to be able to select all objects that depict an event in which "the Dutch army" participated, if one is interested in the historical relation between the Dutch military and Yogyakarta.

As information about the historical context of the object is currently not present in the museum database, we automatically enrich the database records with historical events from the "A Child in War" book by using information extraction techniques. Details of our approach can be found in [15, 10].

The historical event information is modeled according to the Simple Event Model (SEM) [16]. To maximize interoperability, SEM is aimed at minimal modelling of events (similar to the Event Ontology [9] and Lode [12]) but it has the added benefit of being compatible with other external, and complex vocabularies and event models. To bring to light to what extent SEM on its own can facilitate digital hermeneutics, and what external vocabularies are necessary, next we describe our conceptualisation of digital hermeneutics.

3. DIGITAL HERMENEUTICS

In this section, we detail our notion of digital hermeneutics. We describe its main components, illustrate them with examples and show how they can be modeled using SEM. In SEM each event is defined by four elements, i.e. actor, location, period and type. The core of digital hermeneutics is formed by two components: *object-event relationships* and *event-event relationships*. By making explicit relationships between the objects and events, and between the events themselves we can facilitate users in their access and interpretation processes of objects in online cultural heritage collections. Below, we detail the two types of relationships.

3.1 Object-Event Relationships

Object-event relationships are defined as relationships between an object and an event, that are meaningful from a historical perspective. By making explicit the relationships between objects to events we provide users with a means to understand those objects.

We distinguish three types of object-event-relationships defining the *event dimension of objects*. The first two types of object-event relationships are given by the historical context as the majority of the general public will perceive it; what historical event, such as, a war or important invention is related to this object?

i The object depicts an event (e.g., Attack on Yogyakarta) or aspect thereof (e.g., an actor, a place)

ii The object is used or functions in an event (e.g., a sword is used in a war)

In addition to the object-event relations that are related to the content of the object, there is a third type of object-event relationships that tells us more about the history of the object itself, such as, the creation of the object or its inclusion in a museum collection. Although this may not be the first thing most museum (website) visitors would think of when they start interpreting historical events in relation to collection objects, the events related to the provenance of collection objects are key elements to interpretation, as

[1]Kind in de Oorlog: Mohammed Toha schildert Yogyakarta 1948-1949 [4]

[2]`http://www.vraweb.org/projects/vracore3/categories.html`

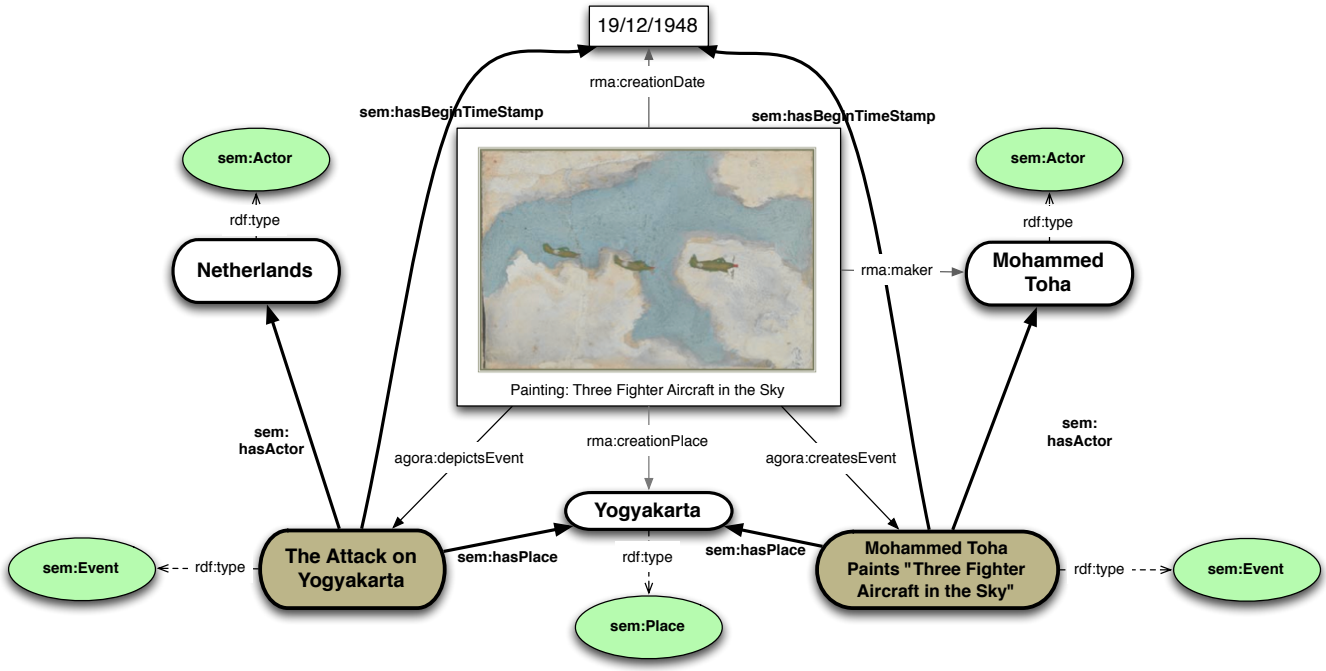

Figure 1: Event dimensions of a depicted event and a creation event of a painting about the Attack on Yogyakarta from the Rijksmuseum Amsterdam collection

they show visitors the meaning why the object is part of the museum collection. This information is in most cases already present in collection databases, and is thus not part of the research in Agora.

iii The making/collecting/exhibiting is in itself an event (e.g., Toha painted attack on Yogyakarta)

An event is something that happens at a certain time and place involving an actor as either agent or patient. From this definition, and following SEM, we discern four *event properties* that are central to our modeling of events: actor, place, time and type. Consult [3, 14] for a more detailed description of event definition and properties used here.

When we relate an event to an object, we essentially provide the object with its *event dimension*. The event dimension consists of the event, that is described through the event properties, and the type of object-event relation. We show an example of an object with two events in its event dimension in Figure 1.

In Figure 1 we show an object and its event dimension. This object is a painting by Mohammed Toha on the attack on Yogyakarta. The collection database tells us that that this object is linked to a 'creation event' *Mohammed Toha Paints "Three Fighter Aircraft in the Sky"*, in which *Mohammed Toha* was the actor, it took place in *Yogyakarta* and the time of the event was *19 December 1948*. What is lacking in most collection databases, is the event information related to the the content of the object, such as the event on the righthand side of Figure 1. The content event dimension enriches the object with the event "*The Attack on Yogyakarta*", and its event properties: actor "*The Netherlands*", time stamp "*19 December 1948*" and location "*Yogyakarta*".

However, not all event dimensions lead the user towards the same interpretation of an object. For example, in Figure 1, the "Attack on Yogyakarta" event is more useful for an interpretation of the object from a perspective on wars than the "Mohammed Toha Paints" event. We therefore need to provide the user with a possibly meaningful relationships between events to help guide his interpretation process. We therefore consider event-event relationships.

3.2 Event-Event Relationships

In this subsection, we show the basic event-event relations that are defined by three elements, i.e. two different events and a relationship type. We distinguish three basic relationship types, corresponding to three of the event properties:

A *topological* or location-based relationship: two events are related because they involve the same place (e.g., Yogyakarta)

B *conceptual* or type-based relationship: two events are related because they involve the same type (e.g., War)

C *biographical* or actor-based relationship: two events are related because they involve the same actor (e.g., Toha).

Using these event-event relationships we can generate prototypical historical narratives. We define a narrative by its topic, the event-event relationship type and the sequence of

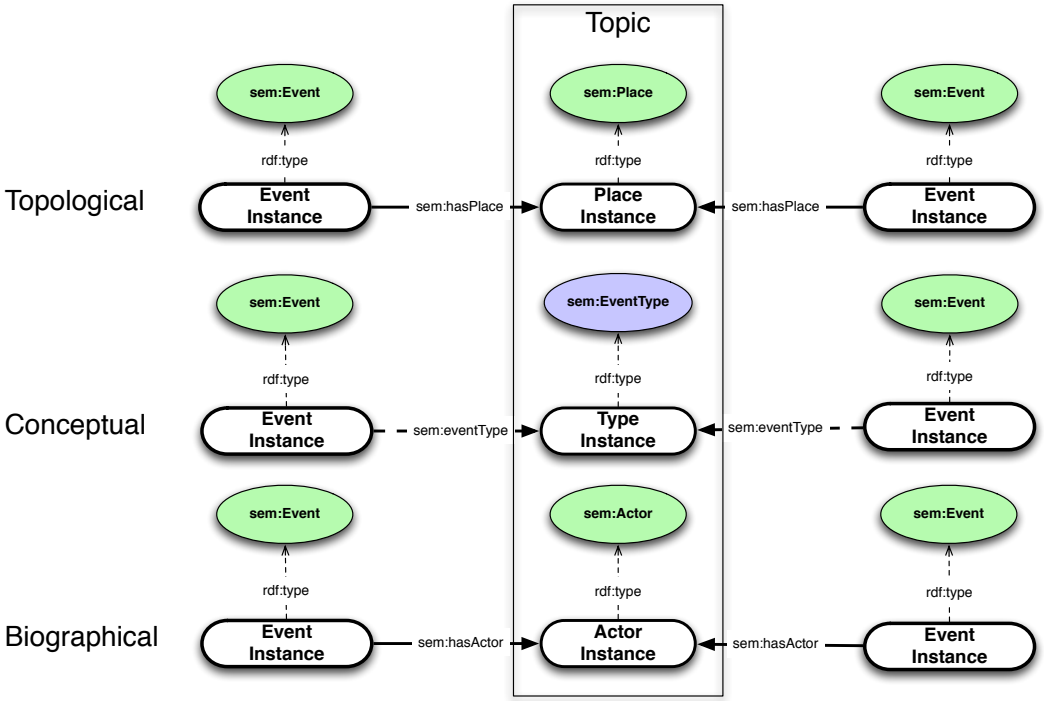

Figure 2: Prototypical types of event-event relations

related events. For example, the narrative $N_i(Actor_j)$ is of a biographical type, where the events $E_1 ... E_n$ are related via the biographical event-event relationship and $Actor_{j1}=...=Actor_{jn}$. Analogously, we define a *conceptual narrative* with the conceptual event-event relationship and a *topological narrative* with the topological event-event relationship. Note, that we do not include the temporal dimension as a possible type of relationship to form a proto-narrative with, as each event-event relation designates a temporal structure [2].

In Subsection 4.2, we give an example of a biographical narrative, where events are related by a biographical relationship.

Figure 2, shows how the components of the prototypical narratives in SEM. A topological narrative could for example be about the location "Yogyakarta", which would be the topic of this narrative. To help the user interpret the history of Yogyakarta, we want to provide him with information about all meaningful events and their properties related to our topic "Yogyakarta".

However, providing the user with all meaningful events and their properties related to a topic may overwhelm the user, as for some topics hundreds of events may be relevant. If we think of the narrative as one dimension along which we can order events, we can use the event properties related to the event to provide a second dimension ordering. One can, for example, group all events by "type" in the Yogyakarta narrative; then it becomes clearer to the user whether a place is for example more often associated with war than with peace. Topological narratives can be ordered by time, actor, and type. Conceptual narratives can be ordered by time, actor, and place. Biographical narratives can be ordered by time, type, and place.

3.3 Additional Relationships

Besides the prototypical event-event relationships, there are other ways to create a historically meaningful narrative. A valid narrative about Mohammed Toha can for example also include events in which his brother was the actor, because Mohammed Toha and his brother share a biographical relationship. For this type of narrative we consider events in which one of the properties shares a part of relationship in the case of locations and concepts (e.g., Yogyakarta being a part of Java and an attack being a part of a war) or a biographical relationship in the case of actors (e.g., sibling relationships). For locations and persons, these types of relationships can be modeled in SEM through external vocabularies such as TGN[3] and FOAF[4] respectively. For event types in the historical domain, there currently does not exist an ontology that models relationships between event types.

We do not exclude the possibility that there may be other relationships that can constitute a valid narrative. For a particular perspective on a painter, one can imagine a narrative that includes events that have been painted by that painter as well as by other painters. However, such examples are context-dependent, and not easily generalizable, we therefore (for now) we limit our focus to the prototypical narrative structures described in Subsection 3.2.

[3] `http://www.getty.edu/research/tools/vocabularies/tgn/`
[4] `http://xmlns.com/foaf/spec/`

Event Description	#
Complete	11
Event + Time stamp	14
Event + Location	10
Event + Actor (agens)	11
Event + Actor (patiens)	12
Event + Actor (witness)	4

Table 1: Event descriptions present in "A Child in War"

4. DH FOR "A CHILD IN WAR"

The Rijksmuseum Amsterdam collection, like many cultural heritage collections, does not contain structured historical event information that ground the collection objects in their historical context. As mentioned in Section 2, some historical descriptions may be present as running text in an object description field in the database, but in most cases, this information will have to come from external resources, such as the "A Child in War" book. In this section, we describe to what extent we can model historical events and relationships from this book to enrich cultural heritage collections with so they can be used to facilitate the interpretation of cultural heritage.

4.1 Event Dimensions

The "A Child in War" book refers to 36 different historical event descriptions in its text. Some of the event descriptions in the text refer to the same event ("Operatie Kraai" and "Tweede Politionele Actie"[5]), but for now we treat every event description as a unique event, as the different verbalisations may denote a different perspective on the event. This may sound as a small amount of events, but the book only deals with events witnessed by a boy living in Indonesia in 1948 and 1949. Furthermore, even this small number of events provides us already with a large variety of narratives to generate that present a multitude of perspectives on this period of Dutch-Indonesian history.

We can compile 12 complete event description from the book (event name, time, location and actors), for the remaining 24 cases, only partial event information (e.g., only the actors or only the actors and locations) can be found in the book. In Table 1, we show the properties we find in the text for the events. The text does not contain any event type information, therefore this is not included in the event dimensions we can compile for "A Child in War".

4.2 Narratives

In this section, we discuss an example of a prototypical narrative in the "A Child of War" use case. Suppose the user selects the artwork *Soldier with a chicken* from the Rijksmuseum Amsterdam collection. This aquarelle depicts a scene of a soldier in the KNIL division[6] stealing a chicken. If the user is interested to explore further the historical space around KNIL, we can generate a biographical narrative of all the events, where KNIL is involved as an *actor*. As shown in Figure 3, there are three events that fulfill this requirement, namely *Bersiap* and *Operatie Kraai* (both of event

[5]"Operation Crow" and "Second Police Action"
[6]Royal Netherlands East Indies Army

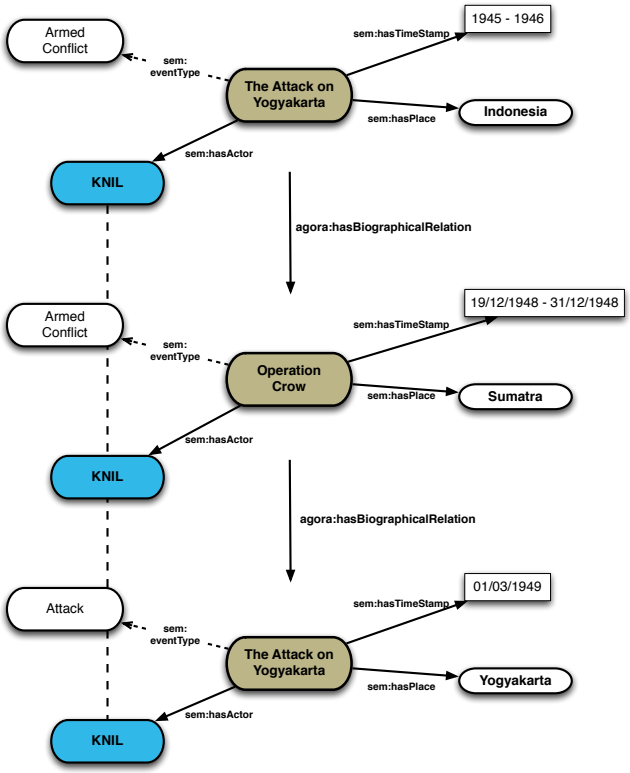

Figure 3: Example of a biographical narrative about the KNIL

type *armed conflict*), as well as *Grote Aanval* (of event type *attack*). The biographical narrative allows the user to explore these three events, all involving KNIL as an actor, from different perspectives through their event properties. For this purpose, we can generate the corresponding views for this narrative by:

1. presenting all events in groups according to their types
2. presenting all events in a chronological order on a timeline
3. presenting all events in a ranked order based on the frequency of their properties, e.g., number of actors involved
4. presenting all events on a map based on their location

In this way, the artwork *Soldier with a chicken* is placed in an extensive historical context, which can be explored by the user from multiple perspectives.

4.3 Current Technical Limitations and Issues

By far, the large majority of knowledge about events is described in diverse text collections such as the "A Child in War" book. Structured event descriptions that facilitate the development of digital hermeneutics are sparse or biased towards a specific collection or topic. The development of digital hermeneutics in the historical domain is therefore challenged by the current state-of-the-art in language technology

Figure 4: View of Toha's Paintings in Agora Demo

to extract the event information from texts. In this section, we identify the main challenges we face in the automatic creation of historical narratives for digital hermeneutics.

Granularity The historical event descriptions we currently recognize and model constitute the so called named events that have a proper name such as *First Police Action*. These events allow for reasonably light-weight and scalable event extraction. However, for extracting phrasal events the text needs to be fully preprocessed including word-sense disambiguation[13]. Even then, not all implications of an event can currently be inferred and extracted automatically. A (translated) excerpt from "A Child in War" illustrates the difficulties of extracting knowledge from natural language: *"Dutch soldiers search Mohammed Toha's house looking for his brother, who joined the guerillas. Mohammed is forced to sit still."* From this sentence it can be inferred that there was a 'raid' event with a certain purpose, namely to find Toha's brother who is considered an enemy of the Netherlands. Another difficulty is that the verb *search* is highly polysemous. To fully model and make this information useful in a digital hermeneutics application, the found event properties have to be grounded in the real world, for example, it also has to be made explicit that Toha's house is in Yogyakarta in order to find related events along the topological dimension. The different information layers that text provides us with makes it very difficult to capture these events and implications.

Data sparsity Another issue that pertains to extracting events from textual documents are the many event elements that are not mentioned in the catalogue. Some named events only appear as a subscript of Toha paintings and do not contain any information about the event itself. Additionally, written text is known to have many anaphora and coreferences that are not easy to resolve. In "A Child in War", an event called "de Grote Aanval[7]" is mentioned 11 times, but this is not a canonical event in other Dutch resources, and one has to know that this refers to the takeover of Yogyakarta by the Indonesian guerillas on 1 March 1949. To further complicate matters, this event is sometimes also referred to as "verrassingsaanval[8]".Also, many event elements are implicitly present in the text or simply never mentioned. Although deep linguistic processing of Dutch text has not reached an optimal level of usability, there are ways around this, for example by combining evidence from different resources about the same events.

Typing of events The conceptual relations between events demand for typing the historical events with some semantic meta-layer. Unfortunately, few ontologies exist that are specifically designed to type historical events. Existing ontologies such as IPTC[9] contain some classification of event types, but are biased towards news events and not optimal for the historical domain. Top ontologies [7, 5]) impose too many constraints and do not contain a domain specific modeling of classes. However, first we need to find out what level of abstraction and expressivity is needed to reason over the event instances, allowing to make meaningful narratives based on event types. In order to do so, an event-based collection access system will have to be tested. A demonstration of a system with such functionality is presented in the next section.

5. "A CHILD IN WAR" DEMO

We show how the historical event dimension for objects can be used in an online collection access system. In Figure 4, we show a screen shot of the Agora demo[10]. The screen shot shows three of Toha's paintings in the center of the screen. For each object the main view is an image of the object, and if possible, the 'object record' can be flipped to reveal a textual description. When a user selects one of the objects,

[7]"the Big Attack"

[8]"surprise attack"

[9]`http://www.iptc.org/site/News_Exchange_Formats/EventsML-G2/`

[10]`http://agora.cs.vu.nl/demo`

in the lefthand sidebar the event dimension of the object is shown, the event properties can be used as facets to filter on. On the righthand sidebar next to the center screen, the objects are shown that share one or more event properties with the selected object. In the sidebar next to that, the related events are shown, which can be used to progress in the narrative. In the top bar, the navigation path the user has followed so far is shown.

6. FUTURE WORK

With the current explosion of online collection portals and increasing role of technology in the humanities, it is to be expected that in the near future digital hermeneutics will be at the center of attention of both humanity scholars and computer scientists. It will do so as a theory of interpretation of information and communication in a digital environment, as well as a context for the development of Web applications with the goal of supporting the interpretation process for its users.

Agora intends to realize the interpretation of historical objects in online cultural heritage collections in the context of a social platform where different groups of users can experience the added value of using historical events and narratives in the exploration of integrated collections. The object-event and event-event relations discussed in this paper not only support object- and event-driven browsing and searching in integrated collections; it also provides meaningful relationships to help users interpret objects and their associated historical events. We are currently working on structuring events in a historical event thesaurus and using this to enrich the object metadata with [10, 11].

We are working towards a user-curated environment where users will be able to *(i)* contribute additional event-related metadata to objects, *(ii)* identify links between events, and *(iii)* create, save and share historical narratives. This is a step towards making the Web a place of dialogue between cultural heritage institutions and their audience. Such a dialogue could lead to a new understanding of cultural heritage collections, both for the general public, and expert historians.

7. ACKNOWLEDGMENTS

The Agora project is funded by NWO, the Netherlands Organisation for Scientific Research, under the CATCH programme, grant 640.004.801.

8. REFERENCES

[1] R. Capurro. Digital hermeneutics: An outline. *AI & Society*, 35(1):35–42, 2010.

[2] A. Danto. *Narration and Knowledge*. Columbia University Press, 1985.

[3] D. Davidson. *Essays on Actions and Events*. Oxford University Press, 1980.

[4] P. Eckhardt and P. Sigmond. *Kind in de Oorlog; Mohammed Toha schildert Yogyakarta 1948-1949*. Rijksmuseum, 2009.

[5] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, and A. Oltramari. Wonderweb deliverable d18. Technical report, Laboraty for Applied Ontology - ISTC-CNR, 2003.

[6] K. Müller. Museums and virtuality. In R. Parry, editor, *Museums in a Digital Age*, chapter 30, pages 295–305. Routledge, 2007.

[7] I. Niles and A. Pease. Towards a standard upper ontology. In *Proceedings of the international conference on Formal Ontology in Information Systems (FOIS'01)*, 2001.

[8] R. Parry. *Recording the Museum: Digital Heritage and the Technologies of Change*. Taylor & Francis, 2007.

[9] Y. Raimond and S. Abdallah. The event ontology. `http://purl.org/NET/c4dm/event.owl`, 2007.

[10] R. Segers. Extracting and modeling historical events to enhance searching and browsing of digital cultural heritage collections. In *Doctoral Consortium Extended Semantic Web Conference (ESWC2011)*, 2011.

[11] R. Segers, M. Van Erp, L. van der Meij, L. Aroyo, G. Schreiber, B. Wielinga, J. van Ossenbruggen, J. Oomen, and G. Jacobs. Hacking history: Automatic historical event extraction for enriching cultural heritage multimedia collections. In *To appear in: Proceedings of the 6th International Conference on Knowledge Capture (K-CAP'11)*, 2011.

[12] R. Shaw, R. Troncy, and L. Hardman. Lode: Linking open descriptions of events. In *4th Annual Asian Semantic Web Conference (ASWC'09)*, Shanghai, China, 2009.

[13] M. Stevenson and Y. Wilks. Word-sense disambiguation. In R. Mitkov, editor, *The Oxford handbook of computational linguistics*. Oxford University Press, 2003.

[14] C. van den Akker, L. Aroyo, A. Cybulska, M. van Erp, P. Gorgels, L. Hollink, C. Jager, S. Legene, L. van der Meij, J. Oomen, J. van Ossenbruggen, G. Schreiber, R. Segers, P. Vossen, and B. Wielinga. Historical event-based access to museum collections. In *Proceedings of the First International Workshop on Recognising and Tracking Events on the Web and in Real Life (EVENTS2010)*, Athens, Greece, May 2010.

[15] M. van Erp, J. Oomen, R. Segers, C. van den Akker, L. Aroyo, G. Jacobs, S. Legêne, L. van der Meij, J. van Ossenbruggen, and G. Schreiber. Automatic heritage metadata enrichment with historic events. In *Museum and the Web 2011*, 2011.

[16] W. R. van Hage, V. Malaisé, R. Segers, L. Hollink, and G. Schreiber. Design and use of the simple event model (sem). *Journal of Web Semantics*, 2011.

[17] P. Walsh. The web and the unassailable voice. *Archives and Museum Informatics*, 11(2):77–85, 1997.