

Historical Event-based Access to Museum Collections

Chiel van den Akker¹, Lora Aroyo², Agata Cybulska¹, Marieke van Erp², Peter Gorgels³, Laura Hollink⁴, Cathy Jager³, Susan Legène¹, Lourens van der Meij², Johan Oomen⁵, Jacco van Ossenbruggen^{2,6}, Guus Schreiber², Roxane Segers², Piek Vossen¹ and Bob Wielinga^{2,7}

¹ Faculty of Arts/VU University Amsterdam

² Department of Computer Science/VU University Amsterdam

³ Rijksmuseum Amsterdam

⁴ Web Information Systems Group/Delft University of Technology

⁵ Netherlands Institute for Sound and Vision

⁶ Centrum Wiskunde & Informatica (CWI), Amsterdam

⁷ Human-Computer Studies Laboratory/University of Amsterdam

{C.vandenAkker,AK.Cybulska,S.Legene,P.Vossen}@let.vu.nl

{L.M.Aroyo,Marieke,Lourens,Schreiber,RH.Segers}@cs.vu.nl

{P.Gorgels,C.Jager}@rijksmuseum.nl

L.Hollink@tudelft.nl

joomen@beeldengeluid.nl

Jacco.van.Ossenbruggen@cw.nl

B.J.Wielinga@uva.nl.

Abstract. This paper presents research in the context of two multidisciplinary projects aimed at facilitating the history domain with an automatic approach for event extraction and modelling. To realise this, the Semantics of History project is providing a historical ontology and a lexicon to support the detection of historical events in textual data whilst the Agora project focusses on exploring the modelling aspects of historical events and employing the combined results in an event-driven browse and search approach. Furthermore, the historical events are used as a flexible model to identify semantically relevant relationships between objects in highly diverse museum collections, creating meaningful ‘cause’ and ‘effect’ links along the key event dimensions ‘who’, ‘what’, ‘where’ and ‘when’. This should finally support the (re)interpretation process of history research, by allowing end-users to create their own personal narratives, leading to theoretical reflection on the meaning of digitally mediated public history in contemporary society. In this paper, we give a high-level overview of the research challenges in the realisation of a desired search and browse scenario. Finally, we outline the open issues and future research.

1 Introduction

There is a vast amount of historical knowledge locked in museum collections. This knowledge is often explicitly present in textual descriptions accompanying museum objects or implicitly present through the fact that an object belongs to a particular collection and was collected for a particular purpose. In this sense, objects from one museum collection only tell part of the story, as they present a view on the past from one only perspective, limited to their collection. Through combining objects from different collections, a more comprehensive view of a certain historical period can be given. When unlocked, this knowledge can help casual users understand the significance of museum objects

and historical events better and aid experts (curators, art historians and historians) in their search for objects relevant to a topic.

The Agora project started in October 2009 with the aim of facilitating context-driven browsing and search in heterogeneous museum collections. The context that unites these collections is provided by historical events that can be linked to the collection objects, as historical event-descriptions are comprised of causal language, locations, the actors involved and the time of the event. Agora is a four year project with a team made up of experts from the computer science and the history departments at the VU University in Amsterdam^{8,9}, as well as from the Netherlands Institute for Sound and Vision (henceforth: S&V)¹⁰ and Rijksmuseum Amsterdam (henceforth: RMA)¹¹. The goal of Agora is threefold: (1) a historical event thesaurus linked to museum artefacts, (2) a semi-automatic event modelling approach that satisfies both the needs of experts and the general public, and (3) an online social platform in which both the general public and expert historians can explore various perspectives on events, build their own narratives, and contribute to the evolution of the event thesaurus.

The Semantics of History project started in August 2009 with the aim to model changes of historical reality over time and through different writer perspectives which are revealed in historical text archives. For this purpose, Semantics of History aims at developing a historical ontology and a lexicon which will facilitate detection of historical events in textual data. The resulting models and the event extraction from text will be implemented and tested in the context of the Agora project. Semantics of History is funded by the VUA Interfaculty Research Institute CAMeRA and is carried out in cooperation between two departments at the VU University in Amsterdam: the linguistics department and the computer science department.

The results of both Agora and Semantics of History will be deployed in a social cultural heritage platform that will allow different users (e.g., experts, interested laypersons and secondary school students) to have event-based access to the S&V and RMA collections. As Agora and Semantics of History are tightly interwoven, we will from this point on report on the combined results of these projects.

This paper is structured as follows. In Section 2, the motivation for this project is given by discussing the shortcomings of current collection access and modelling and by describing the needs from different user groups in two use-cases. In Section 3, the technical challenges of Agora and Semantics of History are presented, along with the approaches that we are investigating. To conclude, we will present open issues in Section 4. In Figure 1, the different parties, goals and domains that play a role in the two projects are summarised.

2 Motivation

In the humanities domain, there are different user groups with a great need for advanced cross-collection access to resources. In our case, we are dealing with the question of how museums can present the specific information that belongs to objects in their collection in a way that strengthens users' historical understanding and involvement in relevant historical debates. We are also asking ourselves how we can prevent users from ending up perusing the collection in a zapping-like,

⁸ <http://www.cs.vu.nl/>

⁹ <http://www.let.vu.nl/>

¹⁰ <http://www.beeldengeluid.nl>

¹¹ <http://www.rijksmuseum.nl/>

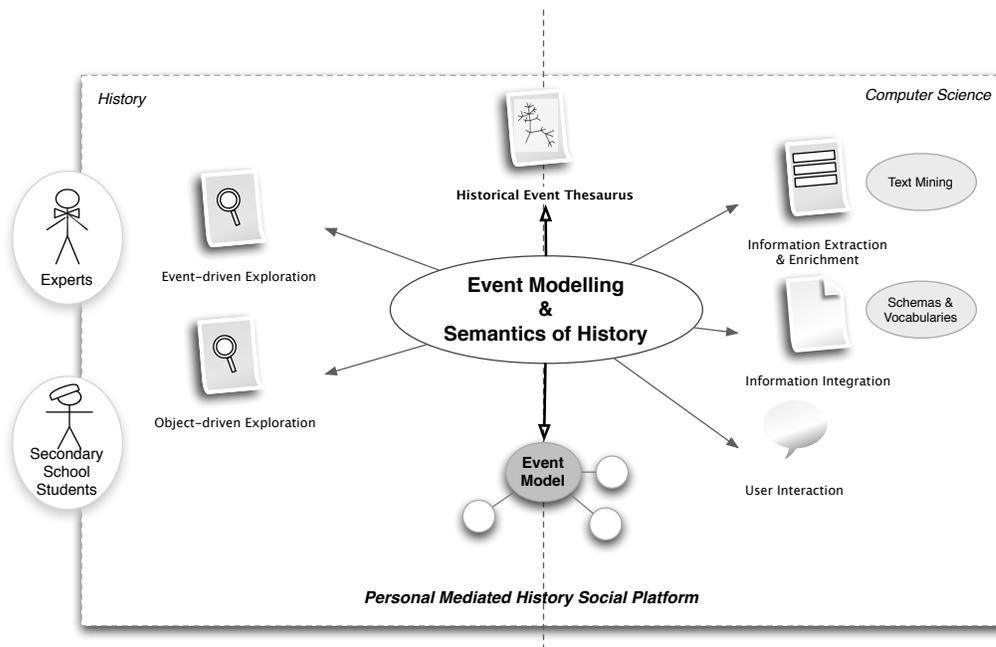


Fig. 1. Overview of domains and goals involved in Agora

incidence-based viewing that will only lead to a confirmation of his or her preconceived views and insights as no relations between objects are presented that are novel or surprising to the user. The answer of museums to this challenge so far has been to prepare thematic ‘Web-specials’: portals and other Web-presentations that present historical narratives as an extension of the regular exhibition and education practice. With an approach that is centred around historical events for collection access we aim to strengthen the role of museum collections in the public discourse about the past.

Although museum collections and their accompanying information are becoming available in digital forms, search is often limited to keyword search and browsing through predefined facets[1]. These access methods are not optimal; in keyword search, for example, it is not clear how the retrieved results are related to each other as they are simply presented in a list. Specifying a search query through facets resolves this problem, as it enables users to specify which relations they find relevant. However, facet browsing is often limited in that there is usually one set of facets that may not provide sufficiently finegrained access to all artefacts. Most museum collections, for example, are searchable through facets that describe meta-data that is available for every object such as *title*, *year*, *artist*, *technique*, *dimensions* and *object id*, but users may also want to search via locations that play a role in the object (for example because a location is depicted in the artwork, or because the birthplace of the artist might be relevant). Combined access to heterogeneous collections from different institutions only augments the shortcomings of keyword and facet access strategies as keyword search will often return more unorganised results when more collections are searched, and facets from different collections are often incompatible.

In addition to the general technical shortcomings of current access methods, there are shortcomings that stem from the cultural heritage domain. Users in this domain have a strong desire to explore collections through a personal narrative or from a personal perspective. Currently, cultural heritage collections lack event-based annotations that would provide the context to facilitate such explorations, but cultural heritage institutions have expressed the wish to have a formal definition of events to include in their annotation of collections.

We argue that collection access through events can remedy these bottlenecks as events provide the context that can link a variety of objects together, providing a more comprehensive overview than facets. To facilitate cross-collection access, we aim to develop an event thesaurus with which different collections can align their internal thesauri. Furthermore, our approach will combine both searching and browsing as this ensures maximum flexibility for the user to explore the collection whilst keeping track of relations between objects. We speculate that by providing a social platform for history, laypersons and experts will complement each other in the process of creating a digitally mediated public history. We believe in an open and social environment in which lay and expert users can together explore and contribute to the evolving collections of objects, events and thesaurus terms. In this way, we will research and develop ways to support a dynamic (community-based and event-centred) creation of narratives of digitally presented material objects as well as multimedia objects. Events are central to narrative and perspective. For a narrative is a sequence of events with a beginning, middle and end, and different sequences of events provide different perspectives on those events.

We envision two types of exploration: object-driven and event-driven exploration. Object-driven exploration involves a search or browsing activity where the user starts by selecting an object from the collections and subsequently finds new objects and events through the relations with the first object. In the event-driven exploration, the user starts by selecting an event and builds a sequence of related events and objects. As a user may hop between events and objects on his or her search through the collections the object-driven and event-driven explorations alternate. An example of a cross-collection exploration scenario in the Agora platform is presented in Figure 2. In this scenario, the object-driven and event-driven exploration is presented as an alternative to the typically currently used small, fixed set of relations imposed by the owner of the collection. It enables the user to wander through the RMA and S&V collections via event-based relations between collection objects that are most relevant to the user. In the example illustrated in Figure 2, the user starts by selecting the RMA print “Arrival of Van Spilbergen in Kandy”. This object is related to an event that has *VOC* as actor and *Batavia* as place. In this way, the user can explore these facets and discover in the results another RMA painting “The Castle of Batavia, seen from West Kali Besar” depicting the Tradeport of Batavia. Via this object, the user can find another object from the RMA collection that depicts an event that takes place such as “A Tea Visit in Batavia”. This object is related to a set of events, such as *acts of colonialism*, which can also have sub-events, e.g. *Police Actions*. The user can choose any of these events or sub-events to explore the collections further and for example arrive at one of the S&V videos that reports on the *Police Actions*. The user then continues the sequence along the facet of another sub-event *Indonesian War for Independence*, which offers the S&V video “Suriname and the Netherlands Antilles” annotated with the sub-event *Suriname’s Independence*. Finally, the user is recommended (from an external resource) the sculpture “Slavery” located in the Oosterpark in Amsterdam, annotated with the same *Suriname*.

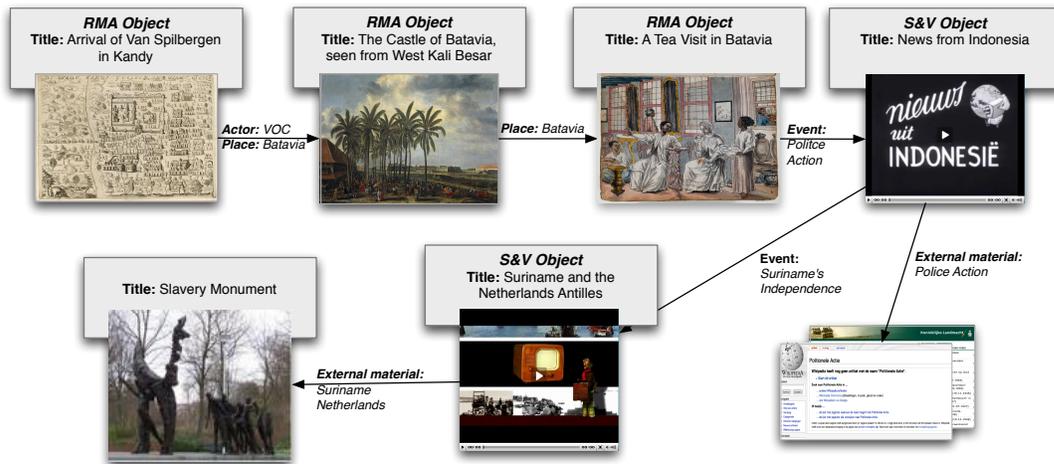


Fig. 2. Agora Exploration Scenario

2.1 Use Cases

We have defined two use cases that illustrate the need to present the same collections in different ways to different user groups. The exploration scenario presented above fits in our first use-case: assisting secondary school students for their Culture and Society end assignment. Since 1998, every student in the two higher tiers of secondary education in the Netherlands is required to write a piece on a particular topic. To facilitate the search for relevant objects and references for (art)historic topics, we will build a use-case that is focussing on event-driven browsing for different museum collections. By combining different objects, students are enabled to present their own view on the events they write about. Different relations between objects and events can result in different views on events.

Our second use case is aimed at facilitating experts in (art)historic research. For historical research, these experts want access to all objects that are related to certain events. To them, Agora platform may present an overview of objects and events a certain actor may be related to, and thus aiding the researcher in his or her information gathering task. We may also help curators to document new objects and add data to old ones, providing structured data by means of the event-model.

3 Challenges

Three domain challenges lie at the basis of our work: (1) sharing terminology, (2) understanding the meaning of events, and (3) building a historical event thesaurus. Each of these challenges is detailed below.

Sharing terminology

A clear definition of the shared terminology is a prerequisite for successful multidisciplinary collaboration. This need is particularly pressing within collaborative projects in the field of cultural

heritage and computer science such as ours as each field has different definitions and theories about shared concepts. Even our central concept, *event*, is treated differently by the parties involved in our project; in the history domain, the notion of an event has been defined as “what agents make happen or undergo”[2]. This has the implication that actions are a species of events. Furthermore, events are concrete particulars. They are unrepeatable entities with a location in space and time[3]. Within computational linguistics, the notion of event is often not defined, and if defined, the definition is mostly pragmatic and broad to ensure reusability across different domains. Another difference is that in computer science ‘event extraction’ often does not stretch beyond the literal task of identifying event labels, participants, locations and time stamps, whereas historians are interested in the interpretation of events. Computer science also considers events mostly as separate entities, whereas historians consider events in their connection with other events. The significance of an event depends on this connection and is usually expressed in the form of a narrative[4]. Through continual dialogue we are acknowledging the differences in each other’s dealing with shared concepts and ensure that we have a stable middle-ground to work from.

Understanding the meaning of events

We are in the process of finding out what the notion of an event means for computer science and for history, and how to incorporate both views in an event model. Although there is no consensus on the definition of event in computer science, most event modelling approaches share the characteristic that they want to model: *Who does what, where and when?*. We take this as at least the minimal requirement an event modelling approach should be able to express. Once a minimal event definition has been developed, we can start to think about modelling additional aspects that play a role in our domain and are closely related to events, namely: granularity, interpretation, perspective, and causality. We are currently investigating the use of the simple event model (SEM) to model historical events[5]. SEM aims to provide the minimal set of classes to describe events, minimising possible clashes between different domain-specific event definitions. It is designed to use external type definitions and has mappings to other models such as DOLCE¹² and LODÉ¹³. We may also borrow from other models such as F[6] for modelling causation and interpretation to extend SEM. However, we do not make any commitments to a particular event model as we aim for a flexible approach in which we can switch to more specific or general models when the need arises.

Building a historical event thesaurus

In order to build a historical event thesaurus, it is important to investigate how historical events are referred to and how they relate to each other in the museum collections. This provides insights into how museum collections are annotated. Next, we will try to model these events in such a way that they can be used to provide better cross-collection access for diverse groups of users. We therefore want the model to be rich enough to capture the intricacies of historical events, but flexible enough to ensure its understandability. The first two steps should result in a thesaurus that can be used to better support users in searching and browsing rich and heterogeneous museum collections.

In addition, we identify three different types of technical challenges that come into play when one wants to disclose museum collections: (1) information extraction and enrichment, (2) information integration, and (3) user interface design. Fortunately, the information extraction, enrichment and integration will not have to start from scratch, as the Rijksmuseum Amsterdam and Netherlands Institute for Sound and Vision have linked their collections to existing domain-specific vocabularies

¹² <http://www.loa-cnr.it/DOLCE.html>

¹³ <http://linkedevents.org/ontology/>

and thesauri such as AAT¹⁴ and Iconclass¹⁵ but also to general vocabularies, such as WordNet[7]. For the user interface, we can build on previous work from the MultimediaN E-Culture project[8]. In the following subsections, we will first discuss the two technical challenges and then our approach to dealing with them. The issues that the event modelling and object- and event-driven collection exploration requirements pose on the interface design depends on the outcomes of resolving the first two technical challenges and will be addressed in a later stage of the project.

3.1 Information Extraction and Enrichment

One of the biggest challenges is information extraction on historical events from different textual data. We will start by identifying the ‘bigger’ events that have been deemed important enough to receive a proper name (e.g., *French Revolution*, *Second World War*). As the behaviour of references to this type of events is similar to other named entities, we are recasting the identification of these event labels as an named entity recognition task[9]. In order to detect accompanying actors, locations and temporal information and the relations between these, as well as smaller events that do not have a proper name we will first employ state of the art named entity and term recognition techniques[10], followed by relation finding[11].

Once we are able to detect references to events, we will need to identify which references belong to distinct events, and which are a variation on the description of the same event. From manual extraction from a small number of newspaper texts, we observed that by identifying time, locations and participants of events and by defining their relations with those of other events, are able to relate historical events with each other. We are developing different matching functions to determine what descriptions refer to the same historical events and test these in larger and more heterogeneous corpora (consisting of e.g., collection catalogues, historical resources, and secondary literature). For example, every description that involves the same type of event, the same participants, the same location and the same time period is likely to refer to the same event. Another heuristic we are investigating involves descriptions that abstract from any of those elements but that we can identify as being semantically compatible which may indicate that these descriptions also refer to the same event. Typically, we see that different descriptions add or leave out details of what happened or group events into large happenings with bigger impact[12].

By means of the event model, relations between specific event descriptions and the more general ones can be represented; also event presentations influenced by writers’ perspective towards historical events can be recognised and captured. Eventually, we want to be able to maximise the recall for finding events in text regardless of the description and secondly be able to infer added subjectivity and interpretation layers to the events.

3.2 Information Integration

The RMA and S&V collections represent different components of the Dutch cultural heritage. The Rijksmuseum Amsterdam focusses on art, crafts and history. A large part of its one million object collection consists of 17th century Dutch paintings. The Netherlands Institute for Sound and Vision aims at preserving the Dutch audiovisual heritage. Its collection contains about 700,000 hours of radio, television, movie and music material, of which most is less than fifty years old. Although the two collections are different in age and focus, there is a fair overlap in the topics they deal with

¹⁴ http://www.getty.edu/research/conducting_research/vocabularies/aat

¹⁵ <http://www.iconclass.nl/>

as the S&V collection contains, for example, documentaries about events that are also depicted or play a role in the RMA collection.

In order to access these two collections simultaneously we first need to align the collection metadata schemas with each other (e.g., *artist* in the RMA collection database may correspond to *creator* in the S&V collection database). From previous experience in the MultimediaN E-Culture project we have learnt that a good way to do this is to map both to an accepted metadata schema such as Visual Resource Association core categories (VRA)¹⁶[8]. To consolidate the collection integration, we also aim to map the values of the fields in the collection databases to a shared vocabulary[13] or to other relevant external resources such as the Dutch Biography Portal¹⁷.

The backbone of our event-driven and object-driven exploration method shall be based on ClioPatria which provides a basis for collection exploration that combines searching and browsing[14]

4 Open Issues

One of the most central open question in this work focusses on what are the recurring historical events that can be traced, in the RMA and S&V collections, based on documentations and attributions of meaning and provenance; and how can we interpret the historical time-lines and narratives that emerge from such a search, using semantics to derive and explain various views, biases, contradictions, opinions and emotional reactions? Currently, in various types of historic documents and collections, events are captured with a single interpretation or perspective. However, we aim at *allowing for multiple local, national, international, and personal perspectives* on historical events and their sequences. In this context we are in search for answers to the following questions:

- how can events be placed in historical sequences, in the context of various collections that address different past-relationships;
- how to include different perspectives in individual events and in event narratives;
- how to extract and model causal relations between events;
- how to involve and motivate the end users in the process of collaborative editing of historical event narratives

Critical for the success of this research is to step upon previous experiences and analyse the implicit historical event model that cultural institutions created by constructing their collections and collection description (i.e., what do events mean to them and how are they represented?). It is interesting to explore the past selection and interpretation processes in order to facilitate new access to (enriched) cultural heritage data and to ultimately investigate how this digitally mediated public history is related to current history writing.

Finally, to allow for effective deployment of the research results we need to specify the envisioned role of a social cultural heritage platform for both the cultural heritage professionals and for the well-informed or interested lay people; and to what extent this platform will be the main drive to maintain the dynamics both in the shared historical thesaurus and the historical events descriptions and their relationships.

¹⁶ <http://www.vraweb.org/projects/vracore4/>

¹⁷ <http://www.biografischportaal.nl>

Acknowledgements

Agora is funded by NWO in the CATCH programme and Semantics of History is funded by VU University of Amsterdam's Interfaculty research institute CAMeRA.

References

1. Cohen, D.J.: History and the second decade of the web. *Rethinking History* **8**(2) (2004) 293–301
2. Ricoeur, P.: *Time and Narrative*. Volume 1. Chicago and London (1984)
3. Davidson, D.: *Essays on Action and Events*. Oxford (1980)
4. Danto, A.: *Analytical Philosophy of History*. Cambridge (1968)
5. van Hage, W.R., Malaisé, V., de Vries, G., Schreiber, A.T., van Someren, M.: Combining ship trajectories and semantics with the simple event model (sem). In: *Proceedings of 1st ACM International Workshop on Events in Multimedia (EiMM09)*, Beijing, China, ACM (October 23 2009)
6. Scherp, A., Franz, T., Saathoff, C., Staab, S.: F—a model of events based on the foundational ontology DOLCE+DnS ultralight. In: *Proceedings The Fifth International Conference on Knowledge Capture (K-CAP 2009)*, Redondo Beach, CA, USA, ACM (2009)
7. Fellbaum, C., ed.: *WordNet: An Electronic Lexical Database*. The MIT Press (1998)
8. Schreiber, G., Amin, A., Aroyo, L., van Assem, M., de Boer, V., Hardman, L., Hildebrand, M., Omelayenko, B., van Ossenbruggen, J., Tordai, A., Wielemaker, J., Wielinga, B.: Semantic annotation and search of cultural-heritage collections: The MultimediaN E-Culture demonstrator. *Journal of Web Semantics* **6**(4) (2008) 243–249
9. Sundheim, B.M.: Overview of results of the muc-6 evaluation. In: *Proceedings of the 6th conference on Message understanding*. (1993) 13–31
10. Ratinov, L., Roth, D.: Design challenges and misconceptions in named entity recognition. In: *Proceedings of Thirteenth Conference on Computational Natural Language Learning (CoNLL 2009)*, Boulder, CO, USA (2009) 147–155
11. Suchanek, F.M., Ifrim, G., Weikum, G.: Combining linguistic and statistical analysis to extract relations from web documents. In: *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, Philadelphia, PA, USA (2006) 712–717
12. Cybulska, A., Vossen, P.: Event models for historical perspectives: determining relations between high and low level events in text, based on the classification of time, location and participants. In: *To Appear in: Proceedings of LREC 2010*, Valletta, Malta (2010)
13. Tordai, A., Omelayenko, B., Schreiber, G.: Thesaurus and metadata alignment for a semantic e-culture application. In: *Proceedings of the 4th international conference on Knowledge capture (K-CAP'07)*, Redondo Beach, CA, USA, ACM (2007) 199–200
14. Wielemaker, J., Hildebrand, M., van Ossenbruggen, J., Schreiber, G.: Thesaurus-based search in large heterogeneous collections. In: *The Semantic Web - ISWC'08*. Volume 5318 of LNCS., Tenerife, Spain, Springer-Verlag (May 2008) 695–708