A SMART METHODOLOGY TO IMPROVE THE STORY-BUILDING PROCESS

Angelo Gaeta
Matteo Gaeta
Giuseppe Guarino
Sergio Miranda

University of Salerno, Department of Information Engineering,
Electrical Engineering and Applied Mathematics, Italy
gguarino@unisa.it, agaeta@unisa.it, mgaeta@unisa.it,
semiranda@unisa.it

Museum narratives are created from a conceptualization of events that can be structurally organized and referred to as the story. Therefore, the main process of developing a museum narrative is one of story-building. This paper presents a methodology to enrich the creation of stories for digital storytelling. The methodology is at the basis of a smart authoring system that supports authors of digital storytelling in the identification of contents from external repositories and their automatic mapping on different narrative structures, according to syntactic and semantic rules and well-defined narrative structures. Furthermore, the methodology leverages on semantic models and technologies such as ontologies, clustering and text analysis, and supports the development of a smart environment for creation of stories starting from a set of pictures. We have validated our methodology for a specific narrative structure, i.e. the Dramatic Arc, for the creation of an educational storytelling related to the museum exhibition about Second World War.

1 Introduction

Nowadays, museums have to find new channels of communication and create other opportunities to activate the participation of their audiences about cultural exhibition. In fact, a successful museum solicits visitor's participation and involves them in the experience. In this way, museums become places of learning but need to incorporate different ways of teaching their audience. One way to do this is through the use of digital storytelling that revolves around the idea of combining the art of telling stories with a variety of digital multimedia, such as images, audio, video and offers a personalized perspective by participants of events (Robin, 2006; Wyman & Smith, 2011). In general, the main objective of storytelling is the use of narrative as a means created by the mind to frame the events of reality and explain them according to a logic of sense. The storytelling is developed starting from the assumption of two basic principles: the organization of human experiences takes place thanks to the stories and the narrative is a process that equips individuals with a cultural sensitivity that enables them to activate reflective processes and training, especially in groups (Gail & Brewster, 2002).

In the context of using storytelling as a means of narration of museum exhibitions (Villaseñor, 2007) that, for example, could be displayed to end users in an interactive way like *serious game* (Antoniu *et al.*, 2013), the paper presents a methodology designed to enrich the development of stories related to museum objects, with additional information retrieved from external datasets. The methodology will be developed in a smart authoring system designed in the FIBAC project[1] (Gaeta A. *et al.*, 2014) and leverages on an approach that uses internal knowledge of museums and sources of external content to enrich storylines (Wolff *et al.*, 2013).

We have enhanced this approach with the adoption of a computational linguistic technique based on the Rhetorical Structure Theory to identify and organize events from textual information analyzing the text fragments and relating them according to a narrative structure.

The rest of the paper is as follow. Section 2 gives details on the methodology. Section 3 gives an overview of the smart system to support authors of storytelling that we are developing in the FIBAC project. Section 4 presents the semantic models and technologies adopted in the project to enable the methodology. Section 5 describes a case study about the generation of additional stories related to the Second World War exhibition. Section 6 motivates which parts of the methodology are original by a comparison with other works and section 7, lastly, reports conclusions and future works.

---

[1] FIBAC is an Italian co-funded project on the valorization of cultural resources in real and virtual museums http://www.ponrec.it/open-data/progetti/scheda-progetto?ProgettoID=5236

This paper is a revised and expanded version of the paper entitled "RST-based methodology to enrich the design of digital storytelling", presented at the 6-th International Conference on Intelligent Networking and Collaborative Systems, INCoS-2014, Salerno, Italy, 2014 (Gaeta A. *et al.*, 2014).

2 The methodology to enhance stories design

Based on the approaches described by Wolff (Wolff *et al.*, 2013), Farhi (Farhi & Kamel, 2004) and Kim (Kim *et al.*, 2006), our methodology consists of three phases (as numbered in Fig. 1):

1. *Retrieval of content/information*: the main goal of this phase (realized by Retrieve module of the authoring system) is to get information from external data sources about the events related to the cultural artefacts of interest for the museum narrative. Our work is mainly focused on retrieving textual information. Specifically, the required operations are:

   a. Set up a set of "seed" events related to the cultural artefacts. Events can be described using some properties according to a metadata schema like *CIDOC/CRM*[2]. For example, some properties of events could be: *start time, end time, location, agent, activity, art historical period, value, genre, style and movement, object, value, materials, dimensions* and *theme* (Wolff *et al.*, *op. cit.*). The final property, *theme*, is more subjective than the other properties. While the first thirteen properties are related to an individual event and are fairly objective, a theme might describe a group of events in a particular context, i.e. within the story that is being developed (*Ibidem*).

   b. Identify *settings* and *themes* aspects for the current set of "seed" events of the step a. Place and time values that co-occur regularly across a set of story events provide a *setting* for the story. Additional frequently occurring event property values, e.g. for people, genre or activity, provide a story *theme* that gives a context to the story that can help interpretation (*Ibidem*). We find the settings for the set of events by using *k-means clustering* algorithm with *Euclidean distance* (Yaday & Sharma, 2013) (see subsection 4.1) to identify the strongest temporal and location values in the event set, by clustering the events and then finding the centroid of the cluster (Wolff *et al.*, *op. cit.*). The theme is derived from the other properties (see step a.) used to describe the events of the "seed set" (*Ibidem*).

   c. Get a list of *settings* and *themes* to be used to query and filter

---

[2] http://www.cidoc-crm.org/

content from external data sources (*Ibidem*).

d. Query and filter the chosen external repository, like *Linked Data Semantic Repository* (Kiryakov *et al.*, 2009) or Learning Objects Repository (Sampson & Zervas, 2013), using the list of *settings* and *themes* obtained in step c. and predefined queries. For example, the queries to be used could be related to: *birth, death, creation of artworks, acquisition of artworks, publications* (Wolff et al., *op. cit.*), *military conflicts, film subject, cultural events*, etc.

e. For each object resulting from the query, retrieve the text that is associated with object's description (e.g. from "*dbp-ont:abstract*" and "*fb:common.topic.description*" properties).

2. *Content/Information analysis and mapping on a narrative structure*: the main goal of this phase (realized by *Analytics* and *Mapping* modules of the authoring system) is to analyse the retrieved content/information in order to identify specific events of interest that can be mapped on a narrative structure. Our work is focused on Rhetorical Structure Theory[3] (RST) for text analysis[4] and on the identification of a specific narrative structure.

3. *Mapping the narrative structure stages on a storytelling ontology*: the goal of this phase (realized by *Mapping* module of the authoring system) is to model narrative structure stages into a storytelling ontology. The need of this mapping is due to the possibility of making reasoning about the aspects of the narration's conceptual structure and managing the knowledge in a network of interconnected events.

In the following paragraphs we describe the key elements of the methodology: the Storytelling ontology (Curate), the Narrative Structure (Dramatic Arc) and the RST, when applied to specific case study devoted to design an educational enriched storytelling for expert and engaged visitors.

---

[3] http://www.sfu.ca/rst/

[4] Indexing techniques for image objects using image relation framework for RST are also proposed (Khurshid & Shoaib, 2013). In order to use RST for multimedia objects it was necessary to modify RST relations to accommodate the multimedia objects (Khaldoon & Shoaib, 2009).
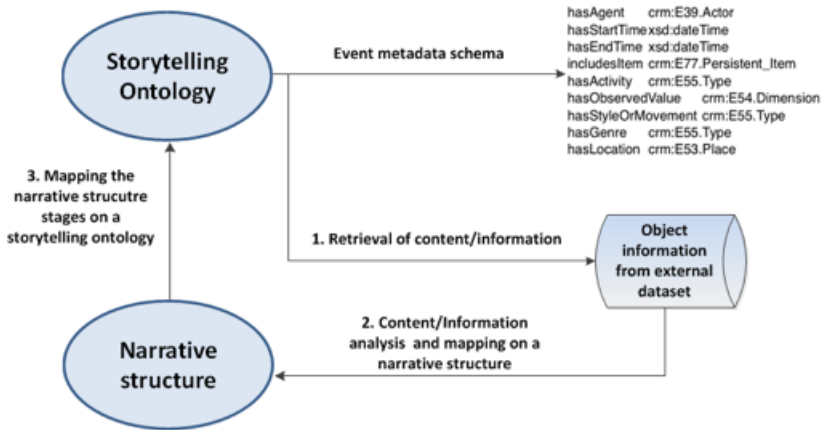
Fig. 1 - Methodology overview

2.1 Curate Ontology

The Curate Ontology[5] is a storytelling ontology designed to describe the main aspects of a museum narrative and their underlying conceptual structure. A museum narrative can describe both the individual cultural artefact and the story that involves a collection of museum objects.

There are five key types of storytelling entity within the ontology. The *narrative* represents a story that is told about one or more museum objects. A narrative is constructed from a *dossier* that is a workspace containing the resources from which a narrative is built. The *story* is the narrative about a particular object. The *object* is the museum cultural artefact of which stories are told. The dossier and the story both contain events. An event is something that has happened. An *event* may be carried out by an actor and it may take place at a particular location and time.

Our work uses Curate Ontology in order to model a storytelling about cultural objects that is organized on stages of a narrative structure.

2.2 Narrative Structures

A narrative structure is the structural model that underlies the order and manner in which a narrative is presented to users (i.e. readers, listeners, viewers). Generally, it can be divided into three main acts: "*setup*", "*confrontation*", "*resolution*". The "*setup*" act is where all of the main characters and their basic situations are introduced and contains the primary level of characterization (exploring the character's backgrounds and own personalities). A problem is

---

5 http://decipher.open.ac.uk/curate/introduction

also introduced and it drives the story forward. The "*confrontation*" act is the pivot of the story and starts when the inciting incident sets things into motion. This is the part of the story where the characters go through major changes in their lives as a result of what is happening. Last act, "*resolution*", is when the problem in the story boils over, forcing the characters to confront it and trying to overcome it (Mangione *et al.*, 2011) (Gaeta M. *et al.*, 2014).

The main best-known narrative structures[6] are: *Dramatic Arc* (Rolfe *et al.*, 2010), *Hero's Journey* (Cao *et al.*, 2011) *Hollywood* or *three act narrative structure* (Thompson, 1999). Our work is focused on the use of "*Dramatic Arc*", the most popular and recognized among the narrative structures and it presents a logical linear narrative with a beginning, a conflict and an end.
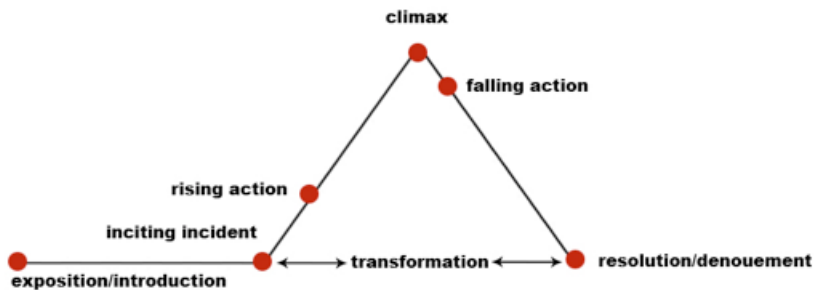
Fig. 2 - Dramatic Arc's stage

In addition, it easily models a set of events and how they are presented to give interest, tension and unique quality to the storyline. As shown in Fig. 2, the stages of the Dramatic Arc are[7]:

1. **Exposition/Introduction**: the main characters and the scene are introduced.
2. **Inciting Incident**: a problem or conflict is introduced, which drives the rest of the story.
3. **Rising Action**: intensity of events increases and the conflict grows.
4. **Climax**: turning point when events and situations change, for better or worse.
5. **Falling Action**: suspense is prolonged as difficulties are confronted and questions are recognized and sometimes answered. The central character typically overcomes conflict in this stage.
6. **Resolution/Denouement**: remaining issues are reconciled. A sense of normalcy is reinstated. Characters, choices, and actions are validated,

---

[6] http://narrativestructures.wisc.edu/home

[7] http://narrativestructures.wisc.edu/home/aristotle/stages

and future possibilities are presented.

2.3 The Rhetorical Structure Theory and its adoption

Rhetorical Structure Theory (RST) is a language-independent theory describing coherence between text fragments. It combines the idea of nuclearity, i.e. the importance of an individual fragment from within the discourse, with the presence of rhetorical relations between these fragments. According to the theory, these relations can be *paratactic* (*Nucleus to Nucleus*, N-N) when they establish relations between fragments that are equally important to the author, or *hypotactic* (*Nucleus to Satellite*, N-S) when they connect a less important unit with a unit the author views to be more important (Iruskieta *et al.*, 2013).

The relations offered by RST and narrative principles applied to these relations (Mann & Thompson, 1987) can be used to organize events extracted from free text in a model of storytelling. The main relations of RTS are shown in Table 1(Nakasone & Ishizuka, 2007).

Table 1
RST RELATIONS

Relation	Description
Background	An event A is referred to as the context in which another event B occurs.
Contrast	For each story that wants to have narrative quality it's important to show some kind of conflict between two or more events. Conflicts, implemented as contracts, give the stories an opportunity to increase understanding of their audience and an interest by creating narrative tension.
Solutionhood	This relation provides a way to define how a "Contrast" relation will be resolved.
Evaluation	This relation provides a definitive conclusion about an event.
Elaboration	With this relation an event is shown to give more details than another one.
Sequence	This relation establishes a linear temporal link between two events. The relation is useful for applying stories in sequence but should not be used as the primary way to connect events.
Cause	With this relation an event is identified as the cause of another event.
Result	This relation indicates that an event is shown as a direct consequence of another event.

Each story's narration phase can be associated with RST's relation in order to organize and relate specific events. A possible approach of association that has been used in more recent works (*Ibidem*; Nakasone *et al.*, 2009) is shown as follow (the symbol "→" represents the meaning of the relation "*associate with*"):
- Story *introduction → Background* relation
- Story *conflict → Contrast* relation

- Story *resolution → Solutionhood* relation
- Story *conclusion → Evaluation* relation

According to this approach, we have determined an association between the individual stages of the "Dramatic Arc" narrative structure and RST relations in order to establish which events should belong to the different phases of a story, as shown in Table 2.

Table 2
ASSOCIATION BETWEEN "DRAMATIC ARC" STAGE AND RST RELATION

"Dramatic Arc" Stage	Description	Associated "RST" Relation
Exposition/ Introduction	This stage introduces the main characters and sets the scene.	Background/Circumstance/Sequence
Inciting Incident	Stage in which a problem or a conflict that drives the rest of the story is introduced.	Contrast/Cause/Result
Rising Action	In this stage event's intensity increases and the conflict grows.	Elaboration from Contrast/Cause/Result relation
Climax	Stage where there is a turning point in which the events and situations change.	Sequence from Contrast/Elaboration relation
Falling Action	The suspense is prolonged because the difficulties need to be addressed. Usually, the main character overcomes the difficulty in this phase.	Solutionhood from Contrast/Elaboration relation
Resolution/ Denouement	In this stage pending issues are reconciled. A sense of normality is restored. The characters, the choices and the actions are validated and future possibilities are presented.	Evaluation or Sequence from Solutionhood/ Circumstance relation

3 Smart Authoring System: a logical view

The Fig. 3 shows the logical view of the authoring systems we are developing in the FIBAC project. At the core of the system, there are three main modules mirroring the three phases of our methodology (described in the previous section 2).

As mentioned in the introduction, the starting point of the system is a set of digital cultural artefacts (e.g. pictures), part of the *internal repository*, that are correlated or are part of the story to be authored.

These artefacts are the input of the *Retrieval* module that analyses these contents in order to gather a set of information to query external repositories and retrieve additional contents to create a story. In our case, as we will better describe in the next section, this module leverages on clustering techniques and

query builder in order to retrieve additional information from external sources.

This additional information is then analysed by the *Analytics* module. This module implements a set of text analytics based on syntactic and semantic rules. In the context of the FIBAC project, this module provides an implementation of the Rhetorical Structure Theory allowing an author to analyse and decompose the corpus according to a set of rhetorical relationships (see section 4.2 for more information).

The *Mapping* module, lastly, provides the mapping between the textual information, analysed and decomposed in atomic units by the Analytics module, and narrative structures such as, for instance, the Dramatic Arc. This is done leveraging on well-defined narrative structure templates and storytelling ontologies, such as the Curate ontology. The modules ***R**etrieval, Analytics* and *Mapping* compose the RAM layer of Fig. 3.

The *Ontologies* and *Rules&Templates* layers support all the phases and provide the necessary flexibility to adopt the systems in different domains and for different narrative structures.
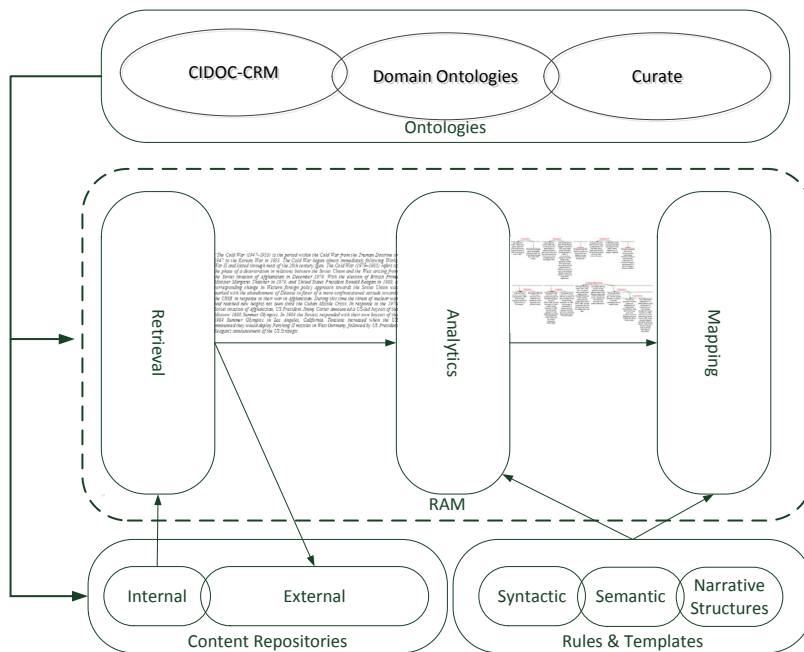
Fig. 3 - Smart Authoring System: a logical view

4 Semantic models and technologies to support the methodology

The following subsections give an overview of the models and technologies

that are used in each phase of the methodology.

4.1 Metadata Schema, Clustering, Semantic Repository and Query Language

CIDOC/CRM Metadata schema is used to model the properties that describe the events (see step a. of the Phase 1 in the description of methodology) (Wolff *et al.*, *op. cit.*).

K-means clustering algorithm, with Euclidean distance (Yadav & Sharma, 2013) is used to identify the strongest temporal and location values in an event set, by clustering the events and then finding the centroid of the cluster (see step b. of the Phase 1 in the description of methodology).

The clustering algorithm is as follow (Wolff *et al.*, *op. cit.*):

1. Cluster events on time, using k=3.
2. For each k cluster, use the time centroid of the cluster to select a set of story events that happened at this time, then find the location and theme centroid of this event set (using the other properties that describe the events). Add the setting (time plus location) for each to the list of story-settings and the theme to the list of theme settings. This maximizes temporal information.
3. Cluster on location, using k=3.
4. For each k cluster, use the location centroid of the cluster to select the set of story events that happened in this location, then find the time and theme centroid of this cluster (using the other properties that describe the events). Add the setting (time plus location) for each to the list of story-settings and the theme to the list of theme settings. This maximizes location information.

Linked Data Semantic Repository (LDSR) (Kiryakov *et al.*, 2009) with a large amount of RDF data[8], like FactForge[9], is used as external dataset to be queried (Wolff *et al., op. cit.*) in order to retrieval additional information about events. SPARQL[10] is used to query the LDSR dataset (*Ibidem*).

4.2 Text Analysis

The main goal of the Phase 2 of the methodology is how to identify important sentences that are essential to understand the context of text related to object's description in order to identify specific events within it. Our idea is to use annotations based on computational linguistic theory, i.e. Rhetorical Structure Theory (RST) analysis, which defines a set of rhetorical relations that are

---

[8] http://en.wikipedia.org/wiki/Resource_Description_Framework

[9] http://www.ontotext.com/factforge

[10] http://www.w3.org/TR/rdf-sparql-query/

used to describe how the sentences are combined to form a coherent text. As such, RST analysis discovers relations within a sentence or among sentences (see Table 1). In RST, the two text spans are further differentiated as *nucleus (N)* and *satellite (S)*. The nucleus texts are more essential to the overall purpose of the document and are comprehensible independently of the satellite (Kim *et al.*, 2006). A contextual analysis of text description, based on a combined approach between the identification of "*Cue phrases*" (Farhi & Kamel, 2004; Kim *et al.*, *op. cit.*) and *recognition criteria*[11] (implemented as rules), helps to discover and classify rhetorical relations (Corston-Oliver, 1998). Table 3 shows the recognition criteria of RST relations set that can be identified from the text related to object's description.

Table 3
CORRESPONDENCE BETWEEN RST RELATIONS AND "CUE PHRASES"

RST Relation	Recognition criterion	Cue Phrases
Contrast	No more than two nuclei; the situations in these two nuclei are (a) comprehended as the same in many respects (b) comprehended as differing in a few respects and (c) compared with respect to one or more of these differences	Whereas, but, however, although, by contrast, even though, though, while, when, after.
Elaboration	S presents additional detail about the situation or some element of subject matter which is presented in N or inferentially accessible in N in one or more of the ways listed below. In the list, if N presents the first member of any pair, then S includes the second: *set:: member; abstraction:: instance; whole:: part process:: step; object:: attribute; generalization:: specific*.	Also, sometimes, usually, for example, in addition, in particular, in general, instead.
Circumstance	S sets a framework in the subject matter within which the reader is intended to interpret N.	After, before, while, post, following.
Condition	Realization of N depends on realization of S.	As long as, if...then, if, so long as, unless, until.
Cause-Effect	S could have caused the agent of the volitional action in N to perform that action; without the presentation of S, the reader might not regard the action as motivated or know the particular motivation; N is more central to writer's purposes in putting forth the N-S combination than S is.	Because, since, as, as a consequence, as a result, thus, therefore, due to, lead to, consequently, regardless.
Sequence	There is a succession relationship between the situations in the nuclei.	Until, before, and, later, then, during, succeeding.
Purpose	S is to be realized through the activity in N.	In order to, so, that, for the purpose of.

[11] http://www.sfu.ca/rst/01intro/definitions.html

RST Relation	Recognition criterion	Cue Phrases
Background	S increases the ability of the reader to comprehend an element in N.	With, probably.
Evaluation	The reader recognizes that S assesses N and recognizes the value it assigns.	With, so, but, which, even so.
Means	S presents a method or instrument which tends to make realization of N more likely.	By, with, using.
Volitional-result	N could have caused S; presentation of N is more central to Writer's purposes than is presentation of S;	Respond, as, regardless.
Solutionhood	N is a solution to the problem presented in S.	Proposed solution, options.
Summary	S presents a restatement of the content of N, that is shorter in bulk.	That is, in other words, in short, to summarize, summarizing.

4.3 Avoiding Knowledge Overlapping

An important issue when a story is created is to check if the information (i.e. text) retrieved from external sources is already described in the internal knowledge base, in order to avoid to enrich museum narratives with no new adding knowledge. This task is constituted of two main phases:

1. Analysis of the text data related to the events of the story for deriving implicit relationships between concepts described through a set of attributes on the one hand and the attributes themselves on the other, and design of a formal context of these concepts and attributes, structured according to a specific model (e.g. lattice).

2. Verification of the degree of similarity between the concepts of the formal context model and the structured model (e.g. ontology) of the internal knowledge base. The constraint to be satisfied is that both schemas are modeled according to the same structure (e.g. graph, vector).

In relation to the first phase, it is possible to use an approach known in literature, based on Formal Concept Analysis (FCA) (Ganter *et al.*, 2005), that is schematized in the Fig. 4 (Cimiano *et al.*, 2005).
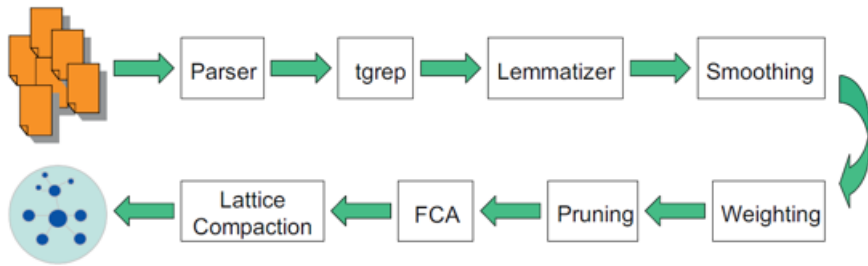
Fig. 4 - Avoiding Knowledge Overlapping - Phase 1

As shown in Fig. 4, the process involves eight steps (Cimiano *et al., op. cit.*):

1. *Step 1* (*Parser*): the text corpus is part-of-speech (POS) tagged[12] and parsed in order to obtain a parse tree for each sentence.
2. *Step 2* (*tgrep*): some phrase dependencies like verb/subject, verb/object and verb/prepositional are extracted from these parse trees. In particular, pairs are extracted consisting of the verb and the head of the subject, object or prepositional phrase.
3. *Step 3* (*Lemmatizer*): the extracted pairs are assigned to their base form.
4. *Step 4* (*Smoothing*): to avoid data sparseness, the collection of pairs is smoothed, where the frequency of pairs which do not appear in the corpus is estimated on the basis of the frequency of other pairs.
5. *Step 5* (*Weighting*): the pairs are weighted according to some statistical measure.
6. *Step 6* (*Pruning*): only the pairs over a certain threshold are transformed into a formal context.
7. *Step 7* (*FCA*): FCA is applied on the formal context. The resulting lattice is transformed into a partial order which is closer to a concept hierarchy.
8. *Step 8* (*Lattice Compaction*): as FCA typically leads to a proliferation of concepts, the partial order is compacted, removing abstract concepts and leading to a compacted partial order, which is the resulting concept hierarchy.

In relation to the second phase, the formal model of concepts obtained in the previous phase is used in conjunction to the structured schema of the internal knowledge base in an ontology matching methodology to verify the similarity of the concepts of both models.

One of the ontology matching approach, well known in literature, that can be applied is described by Melnik (Melnik *et al.*, 2002), where two data structures

---

[12] Part-of-speech tagging consists in assigning each word its syntactic category, i.e. noun, verb, adjective, etc.

(modelled as graphs) are taken by an algorithm as input to produce, as output, a mapping between corresponding nodes of the graphs.

The Fig. 5 shows a simple scheme illustrating the method (*Ibidem*).
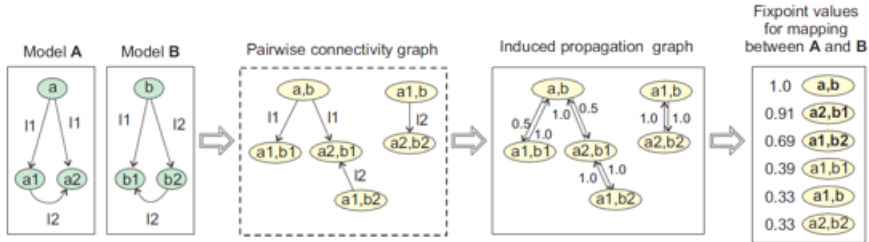
Fig. 5 - Avoiding Knowledge Overlapping - Phase 2

As shown in Fig. 5, the process involves four steps (*Ibidem*):

1. *Step 1*: the internal data models are structured as *directed labeled graphs*[13]. The top left part of Fig. 5 shows two models A and B to be matched.
2. *Step 2*: an auxiliary data structure, named *similarity propagation graph*, is derived from models A and B and used in the fixpoint computation of the method. The propagation graph is computed from A and B by means of a *pairwise connectivity graph (PCG)*, defined as follows:

   (x, y) $p,(x', y') \in PCG(A, B) \Leftrightarrow (x, p, x') \in A$, and $(y, p, y') \in B$
   Each node in the connectivity graph is an element from $A \times B$. Such nodes are called *map pairs*. The connectivity graph, for the proposed example, is shown as a dashed rectangle in Fig. 5.
3. *Step 3*: an *induced propagation graph* for A and B is created from the *pairwise connectivity graph* resulting in the previous step. For each edge in the connectivity graph, in the propagation graph there is an additional edge going in the opposite direction. The weights placed on the edges of the propagation graph indicate how well the similarity of a given map pair propagates to its neighbours and back. These so-called *propagation coefficients* range from 0 (no contribute) to 1 (full contribute) inclusively and can be computed in many different ways. The approach is based on the intuition that each edge type makes an equal contribution of 1.0 to spreading of similarities from a given map pair. If there is exactly one edge out of map pair in the connectivity graph, its weight is 1.0 while, if there are two edges out of map pair, the weight

---

[13] Each edge in a graph is represented as a triple (s, p, o), where s and o are the source and target nodes of the edge, and the middle element p is the label of the edge.

is equally distributed among them (0.5 and 0.5).

4. *Step 4*: the fixpoint values for mapping between the models A and B is calculated by a specific iterative computation[14]. It is also defined a threshold value above which the concepts can be considered similar.

In our methodology, the *Avoiding knowledge overlapping* task can be performed in two occasions:

1. After the *Phase 1* of our methodology (see section 2), when the text corpus related to cultural objects is retrieved from external datasets it is possible to analyse the full text data in order to verify if all concepts associated with this information are already described in the museum knowledge base. If a high (or total) number of concepts is present in the internal dataset, the process of storytelling might be stopped because this information would not add anything new to the museum narrative. In this case, the process is optimized and no redundant information would enrich the knowledge of the museum.

2. Or, alternatively, after the *Phase 2* of our methodology (see section 2), when a set of stories is created based on the correlation of events in the stages of a narrative structure it is possible to verify if the single events of a story are already described in the museum knowledge base. In this case, the eventual combination of redundant events and no mapping between events and narrative stages could determine to ignore some stories for the enrichment of museum narrative.

4.4 Dramatic Arc-Curate mapping

The Fig. 6 shows the mapping of a narrative structure (i.e. Dramatic Arc) with a storytelling ontology (i.e. Curate).

Each stage of the narrative structure represents, in ascending order, a section of the story within the storyline that contains a particular event to be told.

For each "*Section*" of the story's narration, the Curate Ontology manages a "*Constituent Story*" class which is used to assign a position to the section within the storyline (by using the "*hasPosition*" property). In this way, it's possible to model and order the different stages of a narrative structure within a storytelling ontology.

---

[14] For more details about fixpoint computation see (Melnik *et al.*, 2002).
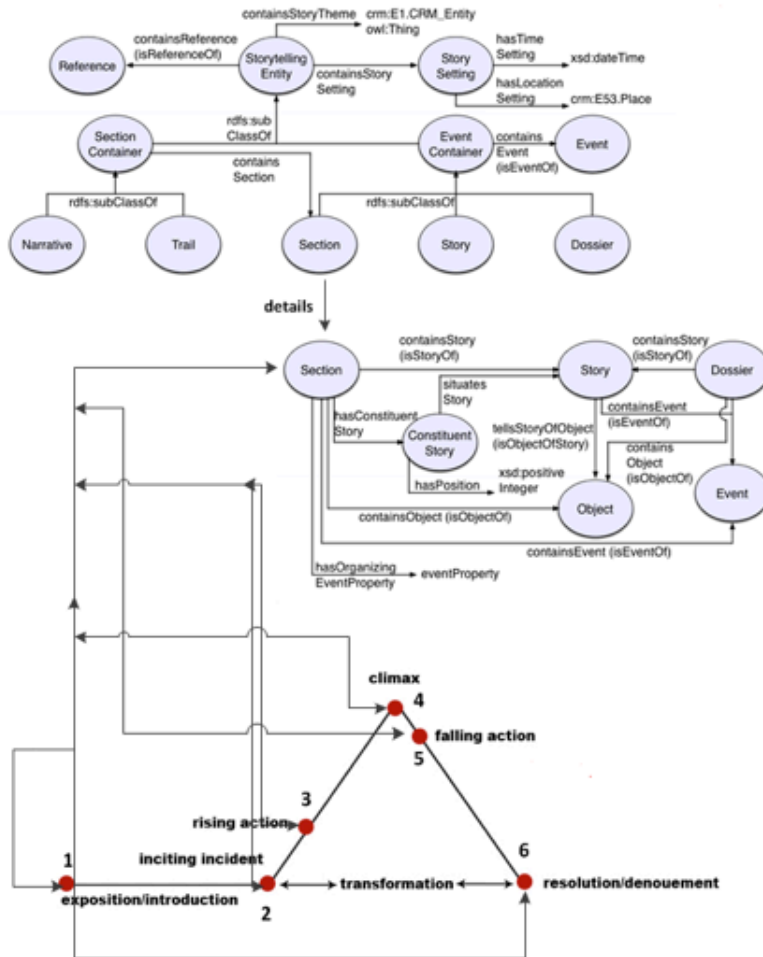
Fig. 6 - Dramatic Arc-Curate mapping

5 Case study

In order to validate the proposed methodology, we present a case study of its application based on a scenario about a museum curator who wants to enrich the creation of a narrative about *Second World War* with new stories associated to external content related with the cultural objects of the exhibition.

Phase 1: Retrieval of additional content/information

Step a) The Table 4 shows the "seed" events set (cultural artefacts consisting of picture and description, that is supposed to be stored in the museum reposito-

ry and have been related manually by the curator) that has been considered for our experimentation. The properties of the events have been modeled according to CIDOC/CRM metadata schema.

Table 4
FIBAC CASE STUDY - "SEED" EVENTS

Cultural artefact (Image objects)	Event description	Event properties
01.09.1939 - German soldiers remove the barrier of a border between Germany and Poland.	The invasion of Poland (September 1, 1939).	Start time: 1939; End time: 1939; Location: Poland; Agent: German soldiers; Activity: barrier removal between Germany and Poland; Genre: invasion.
The Germans increases in Balkans.	The invasion of the Balkans and the preliminary attack on West (March-April 1943).	Start time: 03/1943; End time: 04/1943; Location: Balkans; Agent: German soldiers; Activity: attack; Genre: invasion.
The Soviet armored columns advance in the snow during Little Saturn Operation.	Battle of Stalingrad (November 1942-February 1943).	Start time: 11/1942; End time: 02/1943 Location: Russia; Agent: Soviet armored columns; Activity: advancing in the snow and battle in Little Saturn Operation; Genre: military conflict.
Marshal Zhukov signs the document of surrender of Germany.	Battle of Berlin and the end of the Third Reich (May 8, 1945).	Start time: 1945; End time: 1945; Location: Berlin; Agent: Marshal Zhukov, Germany; Activity: battle of Berlin and end of the Third Reich; Genre: surrender.
Churchill, Roosevelt and Stalin at the Yalta Conference, 1945.	Consequences of the War - The Yalta Conference of 1945.	Start time: 1945; End time: 1945; Location: Russia; Agent: Churchill, Roosevelt, Stalin; Activity: Yalta conference; Genre: conference.

Step b). We have applied the *k-means clustering algorithm* in order to obtain a list of *settings* and *themes* that has been used to query an external content dataset. Resulting clusters are shown in the Table 5.

Table 5

FIBAC CASE STUDY - K-MEANS CLUSTERING ALGORITHM

Clustering typology	Cluster description
Clustering time first and then location (+ theme)	Cluster 1: start_time = 1939 and end_time = 1939 location: Poland theme: German soldiers, invasion
	Cluster 2: start_time = 1942 and end_time = 1943 location: Russia theme: battle, snow, Soviet armored columns, operation, Little Saturn, military conflict, Stalingrad
	Cluster 3: start_time = 1945 and end_time = 1945 location: Germany theme: battle, Berlin, Zhukov, surrender
Clustering location first and then time (+ theme)	Cluster 1: location = Balkans start_time = 1939 and end_time = 1943 theme = invasion, attack, German soldiers, barrier removal
	Cluster 2: location = Berlin start_time = 1945 and end_time = 1945 theme = battle, Third Reich
	Cluster 3: location = Russia start_time = 1945 and end_time = 1945 theme = Churchill, Roosevelt, Stalin, Yalta, conference

Step c). The distinct *settings* derived from algorithm are: *Poland 1939, Russia 1942-1943, Germany 1945, Balkans 1939-1943, Berlin 1945, Russia 1945, while the themes are: German soldiers, invasion, battle, snow, Soviet armored columns, operation, Little Saturn, military conflict, Stalingrad, Berlin, Zhukov, surrender, attack, barrier removal, Third Reich, Churchill, Roosevelt, Stalin, Yalta, conference*.

Step d). We have queried external dataset in order to retrieve additional information by using the list of settings and themes and a predefined query typology. The selected dataset has been *FactForge[15]*. For example, a SPARQL query that use "*Russia, 1945*" as setting and "*military conflict*" as theme is:

```
PREFIX fb:<http://rdf.freebase.com/ns/>
PREFIX dbpedia:<http://dbpedia.org/resource/>
PREFIX dbp-ont:<http://dbpedia.org/ontology/>
SELECT DISTINCT ?event WHERE
{?event fb:time.event.start_date ?sd
.?event fb:type.object.type fb:military.military_conflict
.FILTER (regex(?sd, "^1945"))
.?event fb:time.event.locations ?el
.{ {FILTER (?el=dbpedia:Russia)} UNION
{?event dbp-ont:wikiPageWikiLink dbpedia:Russia}}}
```

---

[15] http://factforge.net/

The query returns all military conflict events which began since 1945 and occurred in Russia or can be connected to Russia through the entity's Wikipedia enter.

The resulting events are:

- *dbpedia:Cold_War_(1947-1953);*
- *dbpedia:Battle_of_Berlin; dbpedia:Prague_uprising;*
- *dbpedia:Evacuation_of_East_Prussia;*
- *dbpedia:Heiligenbeil_Pocket;*
- *dbpedia:War_in_Vietnam_(1945–1946).*

Step e). We are interested to present additional story about "*dbpedia:Cold_War_(1947-1953)*" object. So we have extracted the content related with *dbp-ont:abstract* property[16]. A small extract of the text is shown below.

"*The Cold War (1947–1953) is the period within the Cold War from the Truman Doctrine in 1947 to the Korean War in 1953. The Cold War began almost immediately following World War II and lasted through most of the 20th century@en.*"

Phase 2: Content/Information analysis and mapping on a narrative structure

The Fig. 7 and Fig. 8 show the RST analysis of "Cold War" fact (realized with the *RSTTool[17]* by following the rules presented in Table 3).

---

[16] http://factforge.net/resource/dbpedia/Cold_War_%281947%25E2%2580%25931953%29

[17] http://www.wagsoft.com/RSTTool/

Fig. 7 - FIBAC Case Study - RST Analysis for "Cold War" subject - Part 1
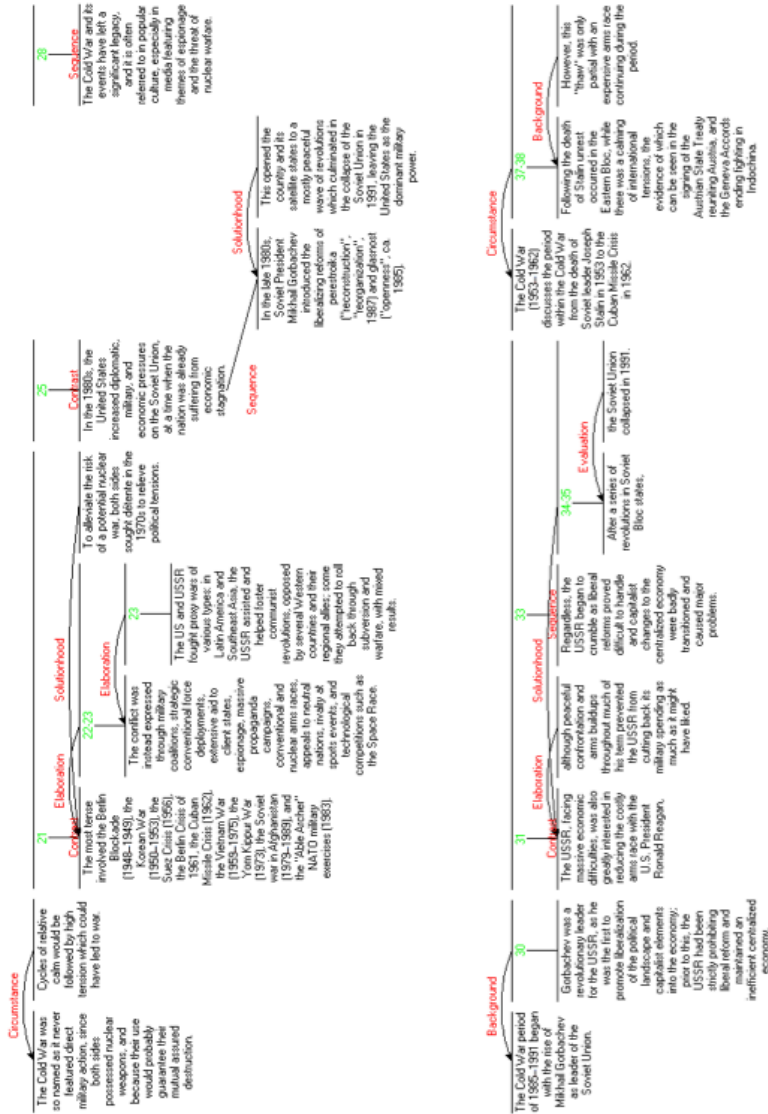
Fig. 8 - FIBAC Case Study - RST Analysis for "Cold War" subject - Part 2

Subsequently, with the mapping proposed in Table 2 we have associated the textual span (nucleus and satellite), identified by RST relation as events (numbered from 1 to 38) in Fig. 7 and Fig. 8, to the "Dramatic Arc" stages in order to obtain linear narrations (stories) about "*Cold War*" subject.

As result, Table 6 proposes the stories that a museum curator can automatically derive. Each numbered event of the story is associated to a picture summarizing the event's description shown in in Fig. 7 and Fig. 8 (the notion "Event X Event Y" represents the timeline of the events).

Table 6
STORIES ABOUT "COLD WAR" SUBJECT

	Exposition/	Inciting Incident	Rising Action	Climax	Falling Action	Resolution/ Denouement
Story 5	Event 36 ↓ Event 37 ↓ Event 38					

Finally, we have applied the *Knowledge Overlapping Avoid* process in order to verify if the events narrated in the stories of Table 6 and Table 7 were already described in the internal knowledge base.

For each story, the analysis of the combination of redundant events and no events-narrative stages mapping allowed us to ignore stories 1 and 5, so that museum narrative about "Second World War" subject can be enriched with the stories 2, 3 and 4.

Phase 3: Mapping the narrative structure stages on a storytelling ontology
With the mapping proposed in subsection 4.4 (see Fig. 6) it's possible to associate each stage of "Dramatic Arc" (with events within it) to a storytelling ontology (in details "Curate" ontology) in order to manage the knowledge about the story.

Case study discussion
The results derived by applying the proposed methodology to the case study point out three main aspects:
1. Ease with which authors can get the skeleton of a story associated to external content (in our case a cultural object). In fact, starting from visual information (e.g. pictures with descriptive text) it's possible to automatically derive additional content whose information can be syntactically analyzed and mapped on a narrative structure, without any extra effort. Authors have only to design and organize the "seed events".
2. Management of overlapped knowledge. The methodology helps the authors to automatically detect any redundant information and to avoid proposing stories that do not add anything new with respect to what is already available in the internal knowledge base.
3. Flexibility with respect to the domains of application. Through the use of rules and templates (i.e. syntactic and semantic approaches and narrative structures) the methodology is independent of the domain and

can be easily applied to different types of stories (e.g. the visual novel).

6 Comparison with other works

Museum narratives, like other forms of narrative, are developed from an underlying conceptualization of events that can be referred to as the story.

Storyspace (Wolff *et al.*, 2012) is a system designed for describing curatorial narratives. The narrative component of the system can produce multiple narratives from the underlying stories and plots of curated exhibitions. Based on the curator's choice, the narrative module suggests a coherent ordering for the events of a story and its associated object. The events come from two sources: from the stories relating to the heritage objects that are chosen for the exhibition and from events that relate not to a single object but to a set, or subset, of the exhibited objects.

Storyspace supports a curator to create stories, plots and narratives about a specific exhibition. With timeline visualisations, the curator can explore and plot the story from multiple different angles, using facets[18] that they have defined as being central to the current curatorial story. This in turn leads to richer narrative construction (*Ibidem*).

Storyscope (Wolff *et al.,* 2013) enhances the approach defined in Storyspace by providing an intelligent support for the selection of events within the story and their interconnection as a coherent structure to be told within the narrative. A plot-reasoning component is adopted to achieve this purpose. It uses both internal knowledge and external information sources to propose content (and, implicitly, events) that can be used to incrementally generate storylines for museum narratives.

Museum narratives in Storyscope are developed through a dossier, which can contains heritage objects (with their own stories), a set of events, either derived from the heritage object stories or else added for an additional context (*Ibidem*).

By comparison, our methodology proposes an approach that enriches the story-building process of Storyscope from two perspectives:
1. The possibility of creating stories associated to specific contents (e.g. cultural or educational objects).
2. Automatic management of overlapped knowledge.

In relation to the first point, an important task in the development of "event based" stories is the extraction of events from data content, generally textual information.

---

[18] Facets can be defined which describe an important property of a curatorial story. Examples include time, location and theme (Wolff *et al.*, 2012).

Byrne (Byrne K. & Klein E., 2009) and Segers (Segers R. *et al.*, 2011) propose to use "*Named Entity Recognition*" techniques to recognize the base concepts related to events such as actors, places, etc. and pattern-based methods for recognizing event names. These approaches have the disadvantage of producing redundancy and co-occurrence of data and the resulting concepts must to be managed and filtered.

Otherwise, our methodology uses the linguistic approach based on *RST* to identify and organize events from textual information only analyzing the text fragments and relating them according to a narrative structure. So, by examining the linguistic form of a text, we are able to make plausible inferences about rhetorical structure and derive events of interest for a storytelling model without worry of managing an undue number of concepts. In this way, we are also able to directly mapping the events into a narrative structure without save the information in the internal knowledge base.

In relation to the second point, it is important to ensure that the additional content retrieved from external sources is not already present in the internal knowledge base in order to avoid getting redundant knowledge that does not offer anything new to the museum narrative compared to what is already known.

Our methodology ensures this result by applying approaches based on the conceptualization of the events of story and on the verification of the similarity degree between concepts.

Conclusions and future work

In this work we have proposed and validated a methodological approach to enrich the design of digital storytelling considering individual needs and cultural context. The aim was to support authors of digital storytelling with the identification of contents from external data sets and their automatic mapping on different narrative structures using semantic technologies and the Rhetorical Structure Theory.

In the context of the FIBAC project, the methodology will be implemented into an authoring system for storytelling that will be able to recommend content for specific visitor profile and experience starting from digital artefacts. Furthermore, we will focus on the simplification of the definition of "seed events" thanks to the use of inference rules (implemented according to a semantic web rule language) that can deduct knowledge without manually analyze the entire knowledge base of the museum. In this way we make leverage on semantic aspects where it's needed; in fact, we adopt a semantic elaboration of these deducted events to gather additional information on cultural artifacts.

Acknowledgements

REFERENCES

Antoniu, A., Lepouras, G., Bampatzia, S., Almpanoudi, H. (2013), *An approach for serious game development for cultural heritage: Case study for an archaeological site and museum*. Journal on Computing and Cultural Heritage 12/2013; 6(4). DOI: 10.1145/2532630.2532633.

Byrne, K., Klein, E. (2009), *Automatic Extraction of Archaeological Events from Text*. Proc. of the 37th Int. Con., Williamsburg, Virginia, USA, March 22 - 26, 2009.

Cao, Y., Klamma, R., & Jarke, M. (2011), *The Hero's Journey-Template-Based Storytelling for Ubiquitous Multimedia Management*. Journal of Multimedia, 6(2), 156-169.

Cimiano, P., Hotho, A., Staab, S. (2005), *Learning Concept Hierarchies from Text Corpora using Formal Concept Analysis*. Journal of Artificial Intelligence Research 24 (2005) 305–339.

Corston-Oliver, S. (1998), *Computing Representations of the Structure of Written Discourse*. Technical Report MSR-TR-98-15 Microsoft Research.

Farhi, M., Kame1, H. (2004), *Rhetorical Structure Theory for content-based indexing and retrieval of Web documents*. Proc. of the 2nd Int. Con. on Information Technology: Research and Education.

Gaeta, A., Gaeta, M., Capuano, N., Mangione, G.R., Pierri, A. (2014), *Enhancing Educational Experiences with Remediation Theory: the Case of the FIBAC Project*. F. Xhafa, L. Barolli, F. Palmieri, M. Koeppen, V. Loia (Eds.), Proc. of the 6th Int. Conf. on Intelligent Networking and Collaborative Systems (INCOS 2014), September 10-12, 2014, Salerno, Italy, pp. 492-497, IEEE Computer Society, 2014.

Gaeta, A., Gaeta, M., Guarino, G. (2014), *RST-based methodology to enrich the design of digital storytelling*. Proc. of the 6th Int. Conf. on Intelligent Networking and Collaborative Systems (INCOS 2014), September 10-12, 2014, Salerno, Italy, pp. 720-725, IEEE Computer Society, 2014,

Gaeta, M., Loia, V., Mangione, G.R., Orciuoli, F., Ritrovato, P., Salerno, S. (2014), *A methodology and an authoring tool for creating Complex Learning Objects to support interactive storytelling*. Computers in Human Behavior, Volume 31, Issue 1, February 2014, Pages 620-637.

Gail, E., Brewster, J. (2002), *Tell it Again!. The New Storytelling Handbook for Primary Teachers*. Harlow: Penguin English, 2002. Print.

Ganter, B., Stumme, G., Wille, R. eds. (2005), *Formal Concept Analysis: Foundations and Applications*. Lecture Notes in Artificial Intelligence, no. 3626, Springer-Verlag,

ISBN 3-540-27891-5.

Iruskieta, M., Aranzabe, M.J., Diaz dellaraza, A., Gonzalez-Dios, I., Lersundi, M., Lopez de Lacalle, O. (2013), *The RST Basque TreeBank: an online search interface to check rhetorical relations*. Anais do IV Workshop "A RST e os Estudos do Texto", pp. 40-49, Fortaleza, CE, Brasil, 2013.

Khaldoon, K. and Shoaib M. (2009), *A Paradigm for Rhetorical Structure Theory (RST) Based Relations to Accommodate Multimedia Objects*. IADIS International Conference eBsociety, Spain, pp. 297-299, 2009.

Khurshid, S.K., Shoaib, M. (2013), *RIO: Rhetorical Structure Theory Based Indexing Technique for Image Objects*, The International Arab Journal of Information Technology, Vol. 10, No. 5, September 2013.

Kim, S., Bracewell, R.H., Ahmed, S. and Wallace, K.M. (2006), *Semantic Annotation To Support Automatic Taxonomy Classification*, 9th Int. Design Con. - Design 2006.

Kiryakov, A., Ognyanoff, D., Velkov, R., Tashev, Z., Peikov, I. (2009), *LDSR: Materialized Reason-able View to the Web of Linked Data*. Proc. of CEUR Workshop, 2009.

Mangione, G.R, Orciuoli, F., Pierri, A., Ritrovato, P., Rosciano, M. (2011), *A new model for storytelling complex learning objects*. Proc. of the 3rd IEEE Int. Con. on Intelligent Networking and Collaborative Systems, INCoS 2011.

Mann, W.C., Thompson, S.A. (1987), *Rhetorical Structure Theory: A Theory of Text Organization*. ISI/RS-87-190, 1987.

Melnik, S., Garcia-Molina, H., Rahm, E. (2002), *Similarity Flooding: A Versatile Graph Matching Algorithm and its Application to Schema Matching*. Proc. of the Int. Con. on Data Engineering, pp. 117-128.

Nakasone, A., Prendinger, H., Ishizuka, M. (2009), *ISRST: Generating interesting multimedia stories on the web*. International Journal of Applied Artificial Intelligence, Vol. 23, No. 7, Sept 2009, pp 633-679.

Nakasone, A., Ishizuka, M. (2007), *ISRST: An Interest based Storytelling Model using Rhetorical Relations*. Edutainment 2007, Springer LNCS 4469, June 11-13, 2007, Hong Kong, China, pp. 324-335.

Robin, B.,R. (2006), *The Educational Uses of Digital Storytelling*. In C. Crawford et al. (Eds.), Proc. of Society for Information Technology & Teacher Education Int. Con. 2006 (pp. 709-716). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE).

Rolfe, B., Jones, C.M., Wallace H. (2010), *Designing Dramatic Play: Story and Game Structure*. Proc. of the 24th BCS Int. Con. on Human-Computer Interaction (HCI2010), Sept. 6–10, 2010 (pp. 448–452). Dundee, UK:British Computer Society.

Sampson, D.G,. Zervas, P. (2013), *Learning object repositories as knowledge management systems*. Knowledge Management & E-Learning, Vol.5, No.2. Jun 2013.

Segers, R., van Erp, M., van der Meij L. (2011), *Hacking History via Event Extraction*. Proc. of the 6th Int. Con. on Knowledge capture, Banff, Canada, June 25-29, 2011.

Thompson, K. (1999), *Storytelling in the new Hollywood: understanding classical*

narrative technique. Harvard University Press.

Villaseñor, M. C. (2007), *Building Echoes. The Role of Storytelling in Museums and Galleries*.

Wyman, B., S. Smith, *et al.* (2011), *Digital storytelling in museums: observations and best practices*. Curator 54(4): 461-468.

Wolff, A., Mulholland, P., Collins, T. (2013), *StoryScope: using theme and setting to guide story enrichment form external data sources*. Proc. of the 24th ACM Con. on Hypertext and Social Media, pp. 79-88.

Wolff, A., Mulholland, P., Collins, T. (2012), *Storyspace: a Story-driven Approach for Creating Museum Narratives*. Proc. of the ACM Con. on Hypertext and Social Media, Milwaukee, USA, June 25-28, 2012.

Yadav, J., Sharma, M. (2013), *A Review of K-mean Algorithm*, International Journal of Engineering Trends and Technology (IJETT) – Volume 4 Issue 7- July 2013.