Integrating Commonsense Knowledge into the Semantic Annotation of Narrative Media Objects

Mario Cataldi¹, Rossana Damiano¹, Vincenzo Lombardo¹, Antonio Pizzo², and Dario Sergi³

¹ Dipartimento di Informatica and CIRMA, Università di Torino, Italy
 ² Dipartimento Dams and CIRMA, Università di Torino, Italy
 ³ Show.it, Torino, Italy. email: {cataldi,rossana,vincenzo}@di.unito.it,

antonio.pizzo@unito.it, sergi@show.it

Abstract. In this paper we present an innovative approach for semantic annotation of narrative media objects (video, text, audio, etc.) that integrates vast commonsense ontological knowledge to a novel ontologybased model of narrative, Drammar (focused on the dramatic concepts of 'character' and 'action'), to permit the annotation of their narrative features.

We also describe the annotation workflow and propose a general architecture that guides the annotation process and permits annotation-based reasoning and search operations. We finally illustrate the proposed annotation model through real examples.

Keywords: media annotation, narrative annotation, ontology

1 Introduction

Nowadays, in a web 2.0 reality, everyone produces stories by creating new materials or by editing and re-interpreting existing media objects (video, text, audio, etc). A large part of these materials are very likely to have a narrative content of some kind. Consider, for example, home made fiction and personal stories in YouTube videos or narratives in blogs and news sites: with such an explosion of contents, search tools for narrative are required. As a consequence, the issue of efficient annotation of the narrative contents is becoming critical.

The narrative format, then, is intrinsically relevant: according to cognitive psychologists [2], it corresponds to the way people conceptualize stories. In crossmedia communication, the narration of stories from the standpoint of characters (so, in a *dramatic* fashion) has become a shared feature of media, as pointed out by [7, 23].

Currently, many retrieval systems exploit the freely associated user generated annotations (tags, descriptions and meta-data) to provide access to the contents [13]; unfortunately, these annotations are often lacking or loosely related to the story incidents. For example, consider a movie segment in which a character escapes from a prison; usually, the annotation of this narrative object relies on its perceivable properties: the actor who plays the character, his physical aspect, the location of the prison, the name of the director. The narrative features of the segment (the actions of the character and the purpose of these actions) are not accounted by current approaches, while they are useful for retrieval and editing, in a search and reuse perspective.

In this paper we present a novel semantic annotation system of narrative media objects (video, text, audio, etc.), centered on the two notions of *character* and *action*. This system is part of the CADMOS project, (Character-centred Annotation of Dramatic Media ObjectS)⁴; the ultimate goal of CADMOS is to test the benefits of narrative annotation for the production and reuse of narrative multimedia. In order to limit the arbitrariness of the annotation, we leverage a large–scale ontology-based semantic knowledge. In this way, the resulting annotation can be used for advanced retrieval operations to provide a more flexible and efficient access to the media objects.

The paper is structured as follows: we first survey, in Section 2, the relevant literature on story in computational systems; in Section 3 we describe the architecture for annotation and retrieval of narrative media objects. Then, in Section 4 we describe the narrative ontology, its theoretical background and the implementation, while in Section 5 we analyze how the narrative annotation of media object is constrained to shared semantics by integrating large commonsense ontologies. Finally, in Section 6 we present a real case study based on the annotation of Alfred Hitchcock's movie "North by Northwest" (1959).

2 Related Work

In this section, we survey some relevant research projects that employ some form of annotation for the analysis and production of media object. These projects, mostly targeted on video, encompass some narrative concepts as part of the annotation they rely on. In Section 4, we review the literature on narrative and drama studies that support the basic tenets of the narrative model underlying our approach.

The Advène project [21] addresses the annotation of digital video, and is not specifically targeted to the narrative content. In Advène, the fragments of the video are annotated with free textual description of the content, cross-segment links, transcribed speech, etc. This information can be exploited to provide advanced visualization and navigation tools for the video. For example, as a result of the annotation, the video becomes available in hypertext format. The annotation is independent from the video data and is contained in a separate package that can be exchanged on the net.

Complementary to this effort, the EU-funded ANSWER project⁵, aims at defining a formal language for script and scene annotation. However, the per-

⁴ Cadmos project is funded by Regione Piemonte, Polo di Innovazione per la Creatività Digitale e la Multimedialità, 2010–2012, POR-FESR 07–13.

⁵ http://www.answer-project.org/

spective of ANSWER is not the annotation of existing material for search and reuse, but the pre-visualization of a media object to be produced, with the aims of helping the author to pursue her/his creative intent and of optimizing the production costs. Again, ANSWER does not address the narrative aspects, but rather the filmic language by which the narrative will be conveyed. This choice is explicit in the project design, since it relies on the semantic layer provided by a Film Production ontology. This ontology constitutes the reference model for the Director notation, the input language for the pre-visualization services.

A media independent project is provided by the OntoMedia ontology, exploited across different projects (such as the Contextus Project [10]) to annotate the narrative content of different media objects, ranging from written literature to comics and tv fiction. The OntoMedia ontology [11] mainly focuses on the representation of events and the order in which they are exposed according to a timeline. In this sense, it lends itself to the comparison of cross-media versions of the same story, for example, a novel and its filmic adaptation, while it does not cover in a detailed way the role of the individual characters and the description of their behavior.

The CADMOS project shares with these approaches the basic assumption that a media object can be segmented into meaningful units and, given some kind of formal or semi-formal description, the units can be accessed and navigated. Moreover, it does not restrict its interest to video, although it recognizes that video constitutes a most suitable test bed for a character-based, narrative annotation: being the most representative example of what Esslin terms 'dramatic media', i.e., media displaying live action [7], the video medium assigns the characters a primary role in the exposition of stories.

Finally, the objectives of the I-Search project (a unIfied framework for multimodal content SEARCH) partially overlap with those of CADMOS. I-SEARCH defines a unified framework, called Rich Unified Content Description (RUCoD) which describes a Content Objects (be it a textual item, a 3D model, or else) in terms of its intrinsic features, social features and user-related, emotional properties [5]. Based on this description, content objects are delivered to users through a innovative graphical interface. Differently from CADMOS, I-SEARCH includes a low level analysis component in its architecture for the automatic acquisition of content information; however, it does not include an explicit representation of this information, partly because it assigns more relevance to emotional and social features

3 The Architecture of CADMOS

The architecture of CADMOS, illustrated in Figure 1, includes six main modules: the User Interface, the Annotation Manager, the Ontology Framework, the Ontology Mashup, NL-to-Onto module and the DMO Repository.

The dramatic media objects (DMO) are stored and indexed within the DMO Repository. A User Interface assists the user in the annotation process. The entire annotation workflow is led by the Annotation Manager which communi-

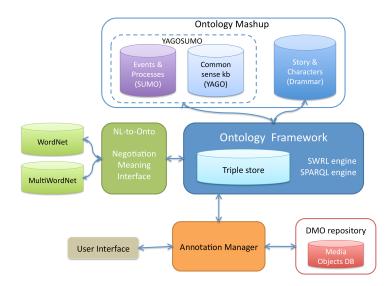


Fig. 1. The architecture of the Cadmos project.

cates with the DMO Repository (permitting to analyze and visualize the original media objects) and with the Ontology Framework, which acts as the pivot component of the architecture. The Annotation Manager and the user interface are an extension of the Cinematic system, described in [15].

The Ontology Framework carries out the reasoning services requested by the Annotation Manager; for example bridging the gap between the natural language input from the user and the ontological knowledge (Ontology Mashup). The mediation among these two modules is conducted through the NL-to-Onto module, which exploits the integration of WordNet with Multi-WordNet⁶ to help the user disambiguate the intended meaning of the input descriptions, and translate them to the correspondent English terms. In fact, as explained in detail in Section 5, given a user input expressed in her/his native language (currently, the only available languages are English, Italian, Spanish, Portuguese, Hebrew and Romanian), this module first tries to disambiguate the sense of the inserted term (in the user native language) by proposing to the user different possible meanings of the term; then, when the user has selected the most suitable meaning, it translates the correspondent term in English by leveraging the WordNet linguistic knowledge.

Finally, the Ontology Mashup maps each English term to an ontology expression, using the vast knowledge expressed by two well-known ontologies: the Suggested Upper Merged Ontology (SUMO [19]⁷) and Yet Another Great Ontol-

⁶ http://multiwordnet.fbk.eu/english/home.php

⁷ http://www.ontologyportal.org/

ogy (YAGO [25]⁸), merged into YAGOSUMO [16]⁹. YAGOSUMO incorporates almost 80 millions of entities from YAGO (which is based on Wikipedia and WordNet) into SUMO, a highly axiomatized formal upper ontology. This combined ontology provides a very detailed information about millions of entities, such as people, cities, organizations, and companies and can be positively used not only for annotation purposes, but also for automated knowledge processing and reasoning. This univocal mapping is possible thanks to the integration of WordNet in YAGOSUMO. While YAGOSUMO represents the commonsense knowledge for describing characters, objects and actions, the narrative ontology expresses the knowledge about the dramatic structures of the narrative domain. The role of these two knowledge sources in the annotation process is described in detail in Section4.

It is important to note that the current architecture also permits annotationbased user queries through the User Interface; in this case, the Ontology Framework translates the user request into a SPARQL and performs the requested operation on the triple store (which contains the annotated information). The result is returned to the Annotation Manager that also retrieves the relevant associated media objects and presents them to the user through the User Interface.

4 The Drammar Ontology

According to literary studies [20], from a dramatic point of view, each story develops along two orthogonal axes: *characters* and *plot*; in fact, each story contains a series of incidents, made of characters' actions and, sometimes, unintentional, or naturally occurring, events. The plot can be recursively segmented into units; in cinema, for example, they usually form three layers, respectively called 'scenes', 'sequences' and 'acts' [14].

In drama, the character plays a central role, as it is the medium through which the story is conveyed to the audience. As acknowledged by contemporary aesthetics [4], character is a powerful instrument of identification [3], contributing to the emotional engagement of the audience "in sympathy with the narrative character" [8]. Thus, the typical workflow of linear storytelling relies upon the notions of "character bible" [24], which reports the characters' typical attitudes, personality and behaviors, and the "story", organized into a set of hierarchical *units*, which form the plot tree.

The formal model of agency known as BDI (Belief, Desire, Intentions) has proven effective to model story characters in computational storytelling. According to this model, characters, given their *beliefs* about the world, devise plans to achieve their goals (or *desires*), and form the *intentions* to execute the actions contained in their plans. While beliefs and intentions constitute the mentalistic component of the model, actions constitute the external, perceivable component of the model. In a story, actions have different levels of granularity in the plot

⁸ http://www.mpi-inf.mpg.de/yago-naga/yago/

⁹ http://www.mpi-inf.mpg.de/ gdemelo/yagosumo.html

tree. At the high level, characters' actions can be described as complex actions, that incapsulate sequences of simpler actions at the lower levels – in the same way as the action "dating" includes "inviting somebody out", "reserving a table", etc. In parallel, at the highest levels, characters' goals tend to persist, while low–level, immediate goals tend to be continuously modified in reaction to the plot incidents.

Being inspired by the notion of bounded rationality [1], the BDI model, by itself, is not sufficient to capture the essentials of characters. In fact, it is important to notice that emotional and moral aspects must be accommodated in a rational background to model actual characters. According to the cognitive framework proposed by Ortony, Clore and Collins [18], emotions stem from the appraisal of one's and others' actions based on a combination of self-interest and moral evaluation. Cognitive studies, then, have pointed out the relation between intentions, i.e., action plans, and emotions and formalized their integration in a computational model [9].

Finally, the sequence of the incidents in a story is represented by the changes in the world state. This component accounts for the narratologists' claim that plot incidents must be causally connected to each other as a necessary condition for story construction [22]. For the story to be consistent, the state of the world that holds after a certain unit must be consistent with the logical preconditions of the unit that follows it in the narration.

The Drammar ontology has been designed with the twofold goal of providing, at the same time, an instrument for the conceptual modeling of drama facts, and a formal tool for the practical task of annotating the narrative properties of dramatic media objects. The ontology is organized to reflect the tripartite structure of plot, characters and units mentioned above. The essential part of the ontology is shown in Figure 2.

The top level of the ontology consists of four disjoint classes. The DramaDynamics, the DramaEntity, the Structure classes (which contain the definition of proper drama), and the Relation class (which encodes some relevant properties of the expressive means by which drama is exposed to the audience in a specific narrative, such as the visual properties that relates characters and objects on scene).

The DramaDynamics class models the evolution of drama though the sequence of incidents (actions or events) that bring the story world from a state to another, affecting the characters' mental states, i.e., their belief, goals, values and emotions. Incidents occur in drama units; states are established as a consequence of the incidents, and constitute the units' preconditions and effects. States can hold in the story world (FactualState class) or in the characters' mind (RepresentedState class), i.e., constitute the content of their beliefs or goals (for example, a character has the goal that a certain state is true or believes it to be true).

The DramaEntity class contains characters (and objects) and their properties. Of paramount importance for drama annotation are the characters' properties that represent their propositional attitudes, i.e., beliefs and goals. Drama

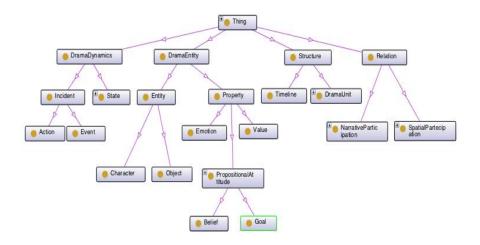


Fig. 2. The class hierarchy of the Drammar ontology.

entities and drama dynamics are connected through the characters' properties of being agent of actions in order to achieve certain goals. Action class represents a well know design pattern, as, through its properties, it links the execution of actions by the characters with a specific segments of the drama.

The Structure class encompasses the notion of plot tree (a hierarchically organized set of units) and the relation of units with the media objects. Incidents happen inside drama units, with characters doing actions purposely to bring about their goals, under the influence of their values and emotions. The property of units of having preconditions and effects connects the incidents in a unit with the state holding before and after it occurs.

The Relation class models the relations among the drama objects at the expressive level, such as the spatial relations. The latter, together with the scene layout, rely on the filmic codes through which the characters' actions are exposed to the audience.

5 Ontology-based Meaning Negotiation

Annotation ambiguities are one of the most critical issue that every annotation method should take into account: in fact, ambiguities create a mismatch between what the annotator intended and the community requirements. They are usually caused by unclear (e.g. incomplete, inconsistent, overlapping) definitions of the terminology employed within the annotation. Ontologies are instrumental in facilitating this negotiation process in large scale online communities. In fact, ontologies currently play a central role in the development and deployment of many data applications, especially in media environment. They can be defined as structured information that describes, explains, locates, and makes easier the retrieval, use, and management of information resources. Within the CADMOS approach, we explored ways to develop ontology-guided meaning negotiation [17], with the goal of avoiding annotation ambiguities. However, since the use of ontologies can be difficult in the annotation process, we explored ways to develop ontology-guided meaning negotiation from natural language input.

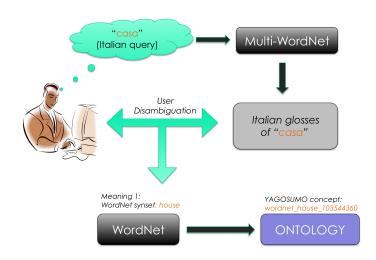


Fig. 3. Ontology-based disambiguation method: the annotator starts the process by searching for some term in her own language. The system retrieves the glosses related to the different meanings of the inserted term and proposes them to the annotator. The annotator selects the most suitable gloss and the system retrieves the related ontological concept by using the associated WordNet synset.

In CADMOS, the YAGOSUMO ontology allows the use of a shared vocabulary for describing resource content and capabilities, whose semantics is described in a (reasonably) unambiguous and machine-processable form. Ontologies also enable effective cooperation between multiple annotators or for establishing consensus where unsupervised (or semi-supervised) systems cooperate with human annotators. Therefore, this domain knowledge helps exclude terminological and conceptual ambiguities, due to ambiguous interpretations. When using YAGO-SUMO for annotation purposes, the annotator can be constrained to use its terminology, through a negotiation step. Moreover, in order to facilitate the work of the annotator, we include in our system, a multi-language tool that permits to initially express the annotation by using the language of the annotator. For example (see Figure 3), the Italian word "casa" can be mapped to different synsets in Multi-Wordnet, ranging from the 'house' and 'home' to 'firm', but

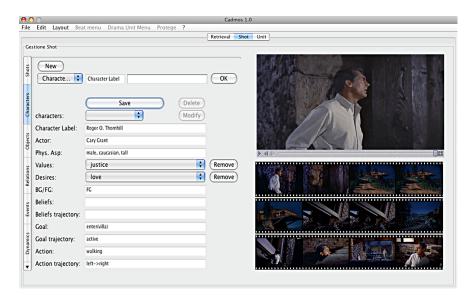


Fig. 4. The prototype of the CADMOS Annotation Interface

only one of them maps to the YAGOSUMO concept that corresponds to the annotator's meaning (here, 'house'). In particular, for each element in the annotation interface (presented in Figure 4), which is associated to a class/predicate in the Drammar Ontology, the system implements the following steps:

- the annotator initially expresses the content as a short natural language term in her/his own language: the term will be forwarded to Multi-Wordnet¹⁰ and the possible meanings of the inserted term are shown by reporting them to the annotator with the different related glosses;
- then, the annotator is able to disambiguate the meaning of the term by selecting the gloss that best match her/his initial thought;
- each gloss is the mapped univocally to the representative English WordNet synset;
- as a final step, YAGOSUMO takes as input the synset and returns the internal related concept. Moreover, it is then possible to analyze the assertions related to the concept and the predicates in which it is involved.

In Figure 3 the entire process is shown (with a concise example): the annotator starts the process searching for the meaning of the italian term "casa" (the Italian translation of "house"). Using the NL-to-Onto module (Section 3) the system retrieves five different meanings and reports them to the user by showing the related italian glosses. Using these natural language explanations, the annotator is now able to identify the intended sense of the inserted term.

¹⁰ http://multiwordnet.fbk.eu/english/home.php

Therefore, the system automatically retrieves the related WordNet synset (by using the WordNet sub-module) and identifies the ontology concept associated to the synset (by exploiting the integration between YAGOSUMO and WordNet store in the Ontology Mashup module).

6 Annotation Example

In order to describe the output of the annotation process we resort to an example, taken from Hitchcock's "North by Northwest". In particular, we describe the annotation of an excerpt of the film, where the protagonist, Roger (Cary Grant), is approaching, unseen, the villain's luxurious house to find out where Eve (Eva Marie Saint) is segregated. The related video segment (a media object) shows Roger in the act of approaching the gate of the house in order to gain a view of the inside.

Therefore, considering the introduced dramatic annotation model, *roger* is an instance of the **Character** class, so it also belongs to the **DramaEntity** class. The existence of this individual however, is not tied to any particular segment of the story: the relation between *roger* and the segment we are considering is given by the fact that he is the agent of an action occurring within the segment. The ontology contains a commitment toward the intentional stance [6], as any occurrence of the action class must be tied to a character's goal that constitutes the purpose for which it is executed. The following triples, represented in Turtle format and stored in the triple store of the Ontology Module (see Section 3), describe Roger's physical aspect, his propositional attitudes and the actions he performs.

In detail, the properties from line 3 to 5 (actor, aspect, gender) state that the character's role is played by the actor Cary Grant and that he looks like an elegant man. Note that the annotation has been constrained through the use of the YAGO knowledge base (from YAGOSUMO, stored in the Ontology Mashup module, see Section 5). Lines 6 to 8 record its relation to other individuals: the **isAgentOf** property specifies the action it performs, the **hasGoal** property connects it to his goals and the **appearsIn** property links the character instance to unit instances. The relation between a certain action and the goal instance that constitutes its purpose is represented as a property of the action itself, as in the following fragment:

:hasPosition1 :roger ;
:occursIn :unit1 .

The action instance represented in these triples is connected to the unit it occurs in (line 6, occursIn), to its action type (type, line 3), i.e., the ontological definition of the action in the domain knowledge (selected by the annotator through the mechanism described in Section 5), to the goal that motivates the character (line 4, hasPurpose) and to the entities that fill the action's thematic roles in its natural language counterpart (expressed according to a positional notation inspired by VerbNet [12], with *roger* as the first argument of the "walk" verb).

This structured annotation permits to properly define not only the basic features of the selected media object (the elements that are shown and their fundamental visual properties) but also the dramatic features and their correlations implied by the narrated story. Thus, the annotation can be also positively used for retrieval purposes by leveraging the drama knowledge imposed by the model and the vast common sense entities definitions reported within the ontologies (stored in the Ontology Mashup module). Moreover, the RDF triple storage also permits, through the Ontology Framework module, reasoning operations that can be used for advanced query operations (for example, the user can now retrieve media objects where some character is performing an action for a specific purpose, i.e. a goal).

7 Conclusions

5

6

In this paper we presented a novel semantic annotation system of media objects that relies on an explicit model of narrative, encoded as an ontology, focused on the dramatic concepts of character and action. We described the theoretical motivations for this work as well as a concrete ontology-based proposal for the annotation process. Moreover, we highlighted the general architecture for a narrative-based annotation system and we presented a novel ontology-based terminology disambiguation method that permits to negotiate the meaning of each term inserted by the annotator. This annotation model enables the interoperability among the processes that create new narrative content and the content aggregators, with benefits for the reuse of existing dramatic media objects and for the creation of new (and already annotated) ones.

References

- 1. Bratman, M.E., Israel, D.J., Pollack, M.E.: Plans and resource-bounded practical reasoning. Computational Intelligence 4, 349–355 (1988)
- 2. Bruner, J.: The narrative construction of reality. Critical Inquiry 18(1), 1–21 (1991)
- 3. Carroll, N.: Beyond Esthetics: Philosophical Essays. Cambridge University Press, New York: Cambridge (2001)
- Carroll, N. (ed.): Special Issue on Narrative, Journal of Aesthetics and Art Criticism, vol. 67. Wiley Blackwell (2009)

11

- Daras, P., Semertzidis, T., Makris, L., Strintzis, M.: Similarity content search in content centric networks. In: Proceedings of the international conference on Multimedia. pp. 775–778. ACM (2010)
- 6. Dennett, D.C.: The Intentional Stance. The MIT Press, Cambridge, MA (1987)
- 7. Esslin, M.: The Field of Drama. Methuen, London (1988 (1987))
- 8. Giovannelli, A.: In Sympathy with Narrative Character, Journal of Aesthetics and Art Criticism, vol. 67, pp. 83–95. Wiley Blackwell (2009)
- Gratch, J., Marsella, S.: The architectural role of emotion in cognitive systems. Integrated models of cognition systems p. 230 (2007)
- Jewell, M.O., Lawrence, K.F., Tuffield, M.M., Prugel-Bennett, A., Millard, D.E., Nixon, M.S., m. c. schraefel, Shadbolt, N.R.: Ontomedia: An ontology for the representation of heterogeneous media. In: Multimedia Information Retrieval Workshop (MMIR 2005) SIGIR. ACM SIGIR, ACM SIGIR (2005)
- Jewell, M., Lawrence, K., Tuffield, M., Prugel-Bennett, A., Millard, D., Nixon, M., Shadbolt, N.: OntoMedia: An ontology for the representation of heterogeneous media. In: In Proceeding of SIGIR workshop on Mutlimedia Information Retrieval. ACM SIGIR (2005)
- 12. Kipper, K.: VerbNet: A broad-coverage, comprehensive verb lexicon. Ph.D. thesis, University of Pennsylvania (2005)
- Kiryakov, A., Popov, B., Terziev, I., Manov, D., Ognyanoff, D.: Semantic annotation, indexing, and retrieval. Web Semantics: Science, Services and Agents on the World Wide Web 2(1), 49 – 79 (2004)
- 14. Lavandier, Y.: La dramaturgie. Le clown et l'enfant, Cergy (1994)
- Lombardo, V., Damiano, R.: Semantic annotation of narrative media objects. Multimedia Tools and Applications pp. 1–33 (2011), http://dx.doi.org/10.1007/s11042-011-0813-2, 10.1007/s11042-011-0813-2
- Melo, G.d., Suchanek, F., Pease, A.: Integrating yago into the suggested upper merged ontology. In: Proceedings of the 2008 20th IEEE International Conference on Tools with Artificial Intelligence - Volume 01. pp. 190–193. IEEE Computer Society, Washington, DC, USA (2008)
- Moor, A.D.: Ontology-guided meaning negotiation in communities of practice. In: Proc. of the Workshop on the Design for Large-Scale Digital Communities at the 2nd International Conference on Communities and Technologies. pp. 21–28 (2005)
- Ortony, A., Clore, G., Collins, A.: The Cognitive Structure of Emotions. Cambridde University Press (1988)
- Pease, A., Niles, I., Li, J.: The suggested upper merged ontology: A large ontology for the semantic web and its applications. In: In Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web (2002)
- 20. Prince, G.: A Dictionary of Narratology. University of Nebraska Press (2003)
- Richard, B., Prié, Y., Calabretto, S.: Towards a unified model for audiovisual active reading. In: Tenth IEEE International Symposium on Multimedia. pp. 673–678 (Dec 2008), http://liris.cnrs.fr/publis/?id=3719
- 22. Rimmon-Kenan, S.: Narrative Fiction: Contemporary Poetics. Routledge (1983)
- Ryan, M.L.: Narrative Across Media: The Language of Storytelling. University of Nebraska Press (2004)
- 24. Seger, L.: Creating Unforgettable Characters. Henry Holt and Company, New York (1990)
- Suchanek, F.M., Kasneci, G., Weikum, G.: Yago: A Core of Semantic Knowledge. In: 16th international World Wide Web conference (WWW 2007). ACM Press, New York, NY, USA (2007)