A Story to Go, Please

Frank Nack, Abdallah El Ali, Philo van Kemenade, Jan Overgoor, and Bastiaan van der Weij

ISLA, University of Amsterdam

Abstract. In this paper we present an approach for an association-based story environment, in which a priori unrelated experiences represented in images, are stitched together to guide users through interesting city spaces. We describe the associative stories generated by this system, and outline how the notion of 'hypespot' facilitates linking the real world with the structure of the story. We outline the overall architecture of the system and provide a generated story example.

Keywords: location-based storytelling, guided story space exploration.

1 Introduction

Visiting a place often means either exploring territory by comparing what we perceive with what we know, or listening to stories that tell us about people's experiences at this location some time in the past. The first option gives us the current now and here but we miss the rich hidden experience fabric of a place. The second option forces us to step back from the real and dive into representation forms that provide a specific view, such as travel guide books like Baedeker (art) or Lonely Planet (lifestyle). These media forms do not necessarily represent our current mood or conflict in a given context at a location. We could of course enforce face-to-face encounters with people who are able to knit us into the rich tapestry of a location, yet those are usually hard to encounter.

The integration of low-cost pervasive and personal technology such as mobile devices into our everyday life has changed the way we interact with a location [12]. We can now leave traces of our emotional and intellectual experience as digital audio-visual attachments to any location and make them available to others, for example through augmented reality browsers. Creating such traces is widespread, as people commonly upload text, images (drawn or captured), videos, or audio files to social web environments such as Flickr[1], Facebook[2], or Myspace[3]. Additionally, most mobile devices are now equipped with GPS, which means the generated content is geo-tagged with the location of creation.

However, this available material was likely not created with the intention of being presented in a context beyond its intended one. Thus, computationally

[1] `http://www.flickr.com`; last retrieved: 07-08-2010.
[2] `http://www.facebook.com`; last retrieved: 07-08-2010.
[3] `http://www.myspace.com`; last retrieved: 07-08-2010.

R. Aylett et al. (Eds.): ICIDS 2010, LNCS 6432, pp. 74–85, 2010.

creating a chronology of events, characters, and settings from user-generated content poses a real challenge. Specifically, the system would need to be able to organize arbitrary experience representations in terms of 'mimesis' or 'diegesis'[4] [2] and render them structurally meaningful, with the golden aim of automatically presenting a story not about a place, but a story about being at a place.

This paper presents a method that facilitates people, who wish to explore interesting and popular city spaces, to be guided by the experiences of others. The 'Story-To-Go' system leads users along known-to-be interesting locations, where interesting is defined as the amount of digital visual content recently left by others at a place. The system generates and presents a story in an associative, homeopathic manner, leaving the completion of the story to the visitor's own motivational and psychological attributes. As a use case, we will use the well-known classic Odyssey structure, which forms a poetic, non-linear plot of events as part of a journey [6]. The motivation behind choosing the Odyssey was largely driven by a pragmatic need, testing to what extent a well-defined story structure can be meaningfully perceived in a story engine that is essentially powered by people's associative capacities.

2 Mimesis Systems

Relevant work is performed by Fujita and Arikawa [4] and Landry [9], where they support users' experience of locations through story generation with location-based visual media. Fujita and Arikawa developed Spatial Slideshow, a system that maps personal photo collections and represents them as stories providing route guidance, sightseeing guidance, and so on. Each story is conveyed through a sequence of mapped photographs, presented as a synchronized 3D animation of photographs that induce a motion effect. Landry developed Storytellr, a system that utilizes the storage and tagging systems of Flickr to aid creation of retrospective stories based on a three-step process: annotation, search, construction. Both systems are interesting with respect to story generation, however they function as support systems rather than automatic story generation systems.

For generative story engines, Pérez et al. [15] describe 3 different models of story generation, where they distinguish between 'structure predictability' and 'content predictability'. The structure focuses on narrative transmission and content on events, people, and things. Their notion of 'structural predictability' inspired our approach. A different approach, but also focused on structure, is the work by Multisilta & Mäenpää [11]. The authors apply narrative structure, in their case jazz music, to generate video stories on mobile phones. The narrative is told using the structure of a jazz concert, which begins with the presentation of the theme, followed by solos of various musicians, variations, and finally the band plays the theme in a new, elaborated way. They concluded that instead of generating a coherent, intense and motivated storyline out of occasional video clips, the end result was more like a montage than a story.

[4] Whereas mimesis is about showing a story (comprised of actions), diegesis is about telling a story (which usually includes a narrator).

An analogous approach to the representation of narrative structure focuses on the applicability of story grammars (context-free and generative) to text understanding, where the main influences came from Propp's work on Russian folk-tales [16] and Chomsky's transformational grammar [8]. The foundations of using story grammars to generate narratives is problematic [1], mainly because the resulting systems were either too rigid or they produced very broad sets of story sentences. Nevertheless, we reuse the idea of a context-free story grammar to create an identifiable story structure where the narrative coherence is complemented by the user's own associative capabilities.

3 The 'Story-to-Go' System

The 'Story-To-Go' system is part of ongoing work within the MOCATOUR[5] project, which aims to establish computational methods to facilitate tourists with contextualized and media-based access to information while they freely explore a city. In 'Story-To-Go' we take the concept of location-awareness in our location-based multimedia messaging (LMM) prototype in order to create location-based stories. Below, we describe the LMM prototype briefly followed by the story engine.

3.1 LMM Prototype

The LMM prototype [3] allows augmenting locations using three different media types: text, drawing, and photos. A user can create a free drawing using touch-based input, type text, or snap photographs, where the media expression is fixed at the position and orientation of the user at the time of creation. These media-based expressions are assumed to create a digital memory of the user's experience. Anyone with the application installed on their multimedia-enabled mobile device and is at the same place where the media expression was created can view it. Here, a user is presented with a camera-view, where the anchored expression appears as an Augmented Reality (AR) overlay on top of the camera view. The user is guided towards an expression by means of an on-screen indicator arrow. The augmentations generated by this tool and related metadata form the conceptual building blocks of the material used by the 'Story-To-Go' engine.

3.2 Story Engine

The 'Story-To-Go' story engine is a client-server based system that covers three essential elements. First, it provides the syntactic structures to generate a story in the form of an Odyssey (structure predictability). This part establishes the structures that later allow the system to generate an associative story. Second, it supplies the metadata structures associated with the geo-tagged augmentation

[5] Mobile Cultural Access for Tourists - `http://mocatour.wordpress.com`; last retrieved: 07-08-2010.

data (content predictability), which facilitates the content selection and presentation mechanisms that provide the visual story stimuli. Third it contains the generation rules that facilitate the clustering of expressions into hypespots and the generation of stories. We look at each of these in detail below.

3.2.1 Story Syntax

A story in 'Story-To-Go' is presented by showing a sequence of expressions that follow the semantic coherence associated with the Odyssey. As stated earlier, the motivation behind choosing the Odyssey is due to it's well established structure, where our aim is to revive that structure (albeit associatively) in the user's mind as s/he moves from one location to another. Figure 1 depicts the graphical representation of the Odyssey structure.

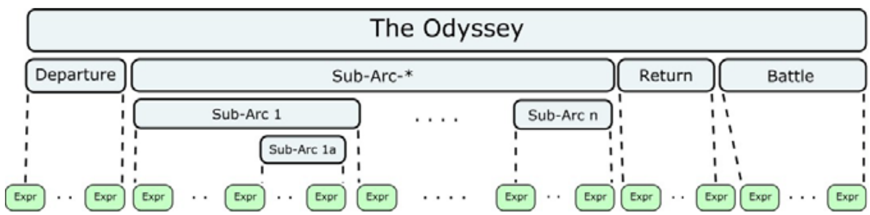

Fig. 1. Graphical representation of story structure of type 'Odyssey'

Syntactically, a complete Odyssey is an arc that consists of sub-arcs, where some arcs are mandatory, such as departure or battle, and others are placeholders for events that can happen in any order. Syntactically, a sub-arc consists of expressions and possibly other sub-arcs, which allows for nested structures. The elementary unit of a sub-arc is an expression. This is in line with our attempt to generate stories that are mimetic by nature.

The compositionality of the sub-arc is an important aspect of this representation. It has local meaning and provides a way for presenting the story in an episodic fashion. Its recursive character allows for stories of arbitrary length in terms of number of expressions and it facilitates branching between sub-arcs of different stories. A type of sub-arc is defined in a similar way as the story-type, except that it employs extra constraints on the meaning of the expressions that it consists of. Figure 2(a) shows the abstract definition of sub-arc, whereas Figure 2(b) presents an example of a particular type of sub-arc.

The definitions in Figure 2 represent the constraints on tags (see Section 3.2.2), which are denoted by the statements between square brackets. For example, [beginning] refers to the expression being tagged with 'beginning'. Preferred conditions are stated within parenthesis, necessary conditions without. A sub-arc definition describes its own size and complexity. The size is expressed in terms of a minimum and maximum amount of expressions that the arc can cover and the complexity refers to the maximum number of nested sub-arcs it can contain. The

```
a <X> ::= <opening>(<e>|<sub-arc-n>)*<closing>
  <sub-arc-n> ::= <sub-arc-(n-1)><sub-arc> | <sub-arc>
  <opening> ::= <e []>
  <closing> ::= <e []>

b <X> ::= <introduction><body><climax>
  <introduction> ::= <e [beginning]><e [beginning, (conflict)]>*
  <body> ::= <e []>*
  <climax> ::= <e [end]>
```

Fig. 2. (a) BNF of an abstract sub-arc. (b) BNF of a composite sub-arc.

abstract sub-arc definition in Figure 2(a) does not enforce any size or complexity limits, except that it should contain at least two expressions. The example composite sub-arc defined in Figure 2(b) only requires the opening and closing expressions to be tagged correctly. In Figure 3 we provide a basic example of sub-arc definitions for the Odyssey.

```
<departure> ::= <e [beginning, landscape, continuity]><e [disruption]>
<love> ::= <e [2 people, beginning]>
           <e [union, attraction, beginning]>
           <e [passion, (beginning, disruption, tension)]>*
           <e [passion, tension, 2 people]>*
           <e [separation, ending, (loose, sad)]>
<battle> ::= <e [beginning, antagonist, disruption, fear, evil]>
             <e [fear, conflict, 2 entities]>*
             <e [end, win|loose, (relief)]>
<illusion> ::= <e [illusion, vague]><e [illusion, clarity, (climax)]>
<return> ::= <e [ending, union, (relief)]>
```

Fig. 3. BNF of sub-arcs for type 'Odyssey'

3.2.2 Story Semantics

Standard characteristics of a narrative in 'Story-To-Go', like identifiable characters or an explicit plot, are not used. Instead our presentation of expressions attempts to create a narrative by making use of the association created by the user. We thus use syntax and minimal semantic structures to create a coherent Odyssey narrative, but leave the more detailed semantics to be filled in by the user. In the end, the level of semantic detail that a story such as the Odyssey can contain is directly linked to the quality of the available metadata associated with the available media expressions, but our hypothesis is that this relation is logistic as opposed to linear.

The expressions used by 'Story-To-Go' are images, which requires the system to understand the visual semantics on three levels in order to support the story generation process: 1) What the image denotes 2) What connotations the image gives 3) How the image can be used in a story. These three levels are used in 'Story-To-Go' to provide the relevant metadata categories, namely: denotation, connotation and narranotation.

The annotations are supplied manually through tagging. This simple yet effortful mechanism was chosen because automatically acquired annotations such as name of creator, creation time and date, or location coordinates do not tell us much about the visual content. Additionally, extraction of low-level features, such as color, texture, or shapes, are insufficient for inferring the high-level intentions behind a created image expression.

Each metadata category has its own specific tags that can be added to an expression. Users can annotate expressions every time they encounter an expression, while observing an image in isolation or within a story context. Each of the metadata categories will be discussed below.

Denotations

For story generation, it is important to recognize what is in an image and especially relevant for detecting character agency within the image so that sub-arcs can be constructed. However, since logical continuity might cause problems, due to a lack of related material, denotative annotations in 'Story-To-Go' ask for high-level categories instead of detailed descriptions. In this way, expressions can be loosely matched. Users can add any number of these tags to an image. Over time the number of denotative annotations for an image will increase depending on the image's popularity as well as its reuse within different story settings. We used the following general categories: *people, animals, text, landscape, abstract image, entity count in image, gender count in image (male, female, neutral).*

Connotations

Connotations describe how a user feels about an image. In 'Story-To-Go', we use a semantic differential scale [14] to calculate the connotation value of an image. This scale is designed to measure people's attitudes toward objects, events, or concepts. People express their feeling by choosing a position on an interval scale between multiple pairs of bipolar adjectives or nouns. The configuration of positions on the scale forms the representation of a user's connotative interpretation of an image. We used the following connotation pairs, based on their usefulness to propagate a story forward [2,10]:

Passion	Void
Continuity	Disruption
Attraction	Fear
Win	Lose
Clarity	Vagueness
Tension	Relief
Good	Evil
Happy	Sad
Union	Separation

Narranotations
Narranotations provide information about how an expression can be used as an element within a story. They are needed to identify the type of sub-arc an expression can be used in, as well as the order expressions appear in within a sub-arc. In line with the Odyssey structure, we use the following fixed set of annotations: *beginning, ending, antagonist, conflict, illusion.*

While denotations and connotations can be thought of as generic metadata structures that can usefully enrich any expression, narranotations add more specifically to the development of a story. However, narranotations aim to capture the use of the isolated expression, so users do not require knowledge about the global structure of the story. We are aware that this type of tagging requests the use of a system like 'Story-To-Go' on a more frequent basis and hence see this as an attempt towards strongly contextualized tagging.

3.2.3 Story Generation

The 'Story-To-Go' story engine essentially performs two tasks: organizes the available images and related metadata according to hypespots, and then generates the story which is presented to the user in real time while exploring the city.

Hypespots: The aim of 'Story-To-Go' is to guide visitors through currently interesting places in a city. To achieve this, we cluster available material (images and metadata) by grouping expressions that are together at a location and that were created recently (a hypespot). A hypespot, covering 150 meter, forms a convenient real world unit to use in the creation of a story. This means that a hypespot represents a place that people frequently visit and create content there; furthermore, given the recency of expressions, it provides grounds to establish 'interesting' locations. The hypespot creation process is shown in Figure 4. This process happens in a dynamic environment, as every added image or a change in related metadata changes the expressiveness of a location, which is addressed by abstracting over all available metadata periodically every 12 hours (half a day), in order to provide the blocks of data the engine requires for generating a story.

The internal structure of a hypespot cluster makes use of the 3 metadata categories described earlier. Denotative and narranotative metadata for every image in a proximity range of 150m are handled similarly, as both provide tags that either inform about the availability of expressions or the applicability for structural placement. This allows for a simple selection process to be performed in which the relevance of a tag is determined by the frequency of its appearance. For connotative metadata, we filter for ratings that are not spread evenly [7] to ensure a bipolar representation that makes it easy to select dominant tags.

Based on the clustered material, the engine identifies and ranks the new set of hypespots in the data set by clustering the expressions according to spatial distance and activity recency. Here a variant of the QT clustering algorithm [5] was used with a temporal parameter integrated in the distance measure. This way only the maximum cluster diameter has to be defined. Once hypespots are identified, the engine searches in each hypespot for available story arcs, utilizing the sub-arc representation structures as defined in Section 3.2.1. When multiple

Fig. 4. Preprocessing scenario. Left: example data set. Middle: after clustering. Right: after sub-arc finding.

ways of constructing a sub-arc are possible, the system selects the best one based on the extent a given subarc complies to the preferred attributes. The found arcs are stored as frames. When a story gets generated these frames are used as the building blocks for the larger story structure.

Story Generation Mechanism: When a user starts a story, the system calculates basic context parameters, such as current position, end position of story and available time, to establish spatial bounds of the search space. An example heuristic looks like this: max number of hypespots = max distance / max number of hypespots per 300 m^2. Within this space the hypespots and their available sub-arcs are queried. The results are analyzed by making use of a partial-order planning algorithm [13], which is constrained by both story definitions described in Section 3.2.1 and the location of hypespots. The system then presents the material across different hypespot locations. The schemas shown in Figure 5 illustrate possible story paths across hypespots.

Fig. 5. Three possible stories

3.3 'Story-to-Go' Proof of Concept

We implemented a proof of concept of both the preprocessing and the story engine in SWI-Prolog and performed two evaluations. Given our limited set of material and participants, the evaluations were meant to be indications about the usefulness of the approach and not solid, empirical proof of its applicability.

In the first evaluation, we wanted to investigate the applicability of the preprocessing algorithm on a large data set. Due to the novelty of the environment we could not test our algorithm on a large enough set of real-world data. We decided to develop a simulation environment in which we applied both preprocessing as well as the story engine on a randomly generated set of 1000 expressions with variable locations spread over an area of 1 km^2 using a randomly chosen set of metadata for the 3 categories. We then processed the sets of hypothetical content, resulting in an average of 132 hotspot clusters, in which an average of 396 sub-arcs could be extracted.

In the second evaluation we wanted to gain two insights: first, could users identify the overall odyssey structure as well as the individual sub-stories if they are confronted only with the visual representation? Second, would our participants be able to tell a story using the material? 50 people, aged between 20 and 45, from different faculties of the University of Amsterdam participated in this experiment. Each participant was informed about the context of the research and the two tasks requested to be performed by them. Each participant was also provided with all metadata structures and associated keywords.

For the first task we asked the participants to generate multimedia messages (i.e., a digital graffiti). We wanted to make sure that the participants focused on the expressiveness of the multimedia message and not initially on the building of a story. Each participant was therefore provided with two locations, each represented as a camera snapshot taken with the mobile device's camera, and a set of related annotation keywords covering all 3 annotation classes. The provided locations all represented places well-known by the participants. This was to ensure that they could envision the complete environment even though they were only represented with a 2D fraction of it. The participants could use the LMM prototype or any other tool to generate the multimedia messages. They had a week to finalize this task. The outcome was 84 multimedia messages in the form of image overlays.

After the provision of all the multimedia messages, participants were presented with one of the odyssey stories generated by the system (see Figure 6). The participants received the Odyssey story (as a series of concatenated images) without the associated metadata. Each participant was asked to identify the story blocks and the related category of the block and state if the overall story structure passes as an odyssey. Afterwards they were asked to provide the story they made out of the images. Eleven participants provided the requested material, where the majority could identify the story segments (e.g., this segment is about love; here something violent is happening). However, none could get the section borders correct. All participants stated that the story structure could pass as an Odyssey but that it could also represent any type of drama structure.

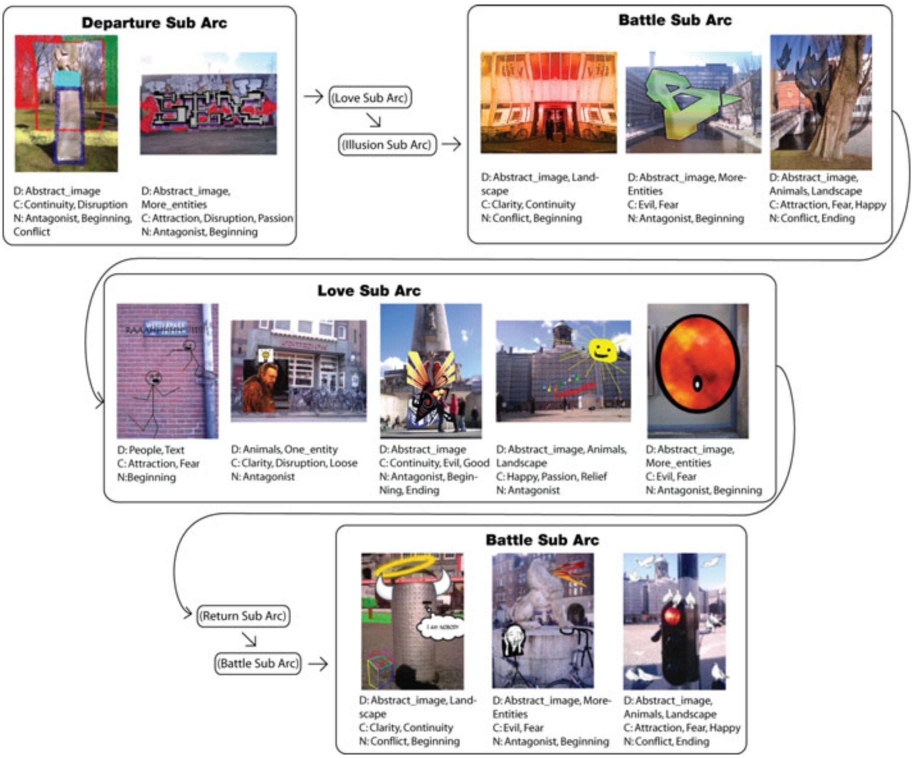

Fig. 6. Example story resulting from our second evaluation

The provided stories showed that all participants only at the beginning looked for indicators of the stories personage and then later connected the appearing material according to their mental representations. Thus, it was not relevant for the participants that the characters were actually present in each image, as long as they could relate some content to some character.

We are aware that presenting the story images together as a coherent whole is different than presenting a single image at a location, where the story is completed through the traversal of different locations in a city. However our evaluation supported our initial hypothesis: namely, that associative stories can be generated from user-generated content and understood by other consumers.

4 Conclusions and Future Work

Using the well-defined Odyssey structure, we have developed a method for generating and presenting an associative location-based story using user-generated media content, where the stories are based on the amount and recency of user activity. The proposed method uses a grammar-inspired compositional story representation based on sub-arcs. The semantics of stories are rooted in the content metadata. While we used the metadata to compose our Odyssey stories, mental associativity from the user is required for the Odyssey to be perceived as such.

The evaluations we carried out using the system-generated stories were not based on real-world user-generated content. This was because geo-tagged media messages required for the system to generate stories do not exist yet. Furthermore, motivating users to supply the necessary metadata poses an ongoing research challenge. It is important to further investigate how the established tagging categories of this work can be automatically populated with tags already available at sites such as Flickr. However, if incentive mechanisms that raise awareness in users about what stories the system can generate if they contribute their content are designed into the system, it might gradually motivate them to tag more often. Despite these limitations, our proposed method demonstrated a proof of concept system that can take arbitrary content and generate location-based stories on the fly.

To enhance the user experience of consuming location-based stories, future work should address the issue of making the user interaction with the story more interactive. One method of doing this is to allow the user to veto the system-provided routes about which location to go next and instead allow the user to pick her own route, where the generated storyline is constructed dynamically to accommodate the users decision. This kind of non-linearity would require the system to make quick, real-time updates as the user moves, which is currently not supported. However, the generic and compositional representation of stories used in our proposed method makes this possible.

Future work should also investigate using a stronger user evaluation the extent that users familiar and unfamiliar with the Odyssey structure can understand the narrative development in the Odyssey as they traverse hypespots within and across cities. This would provide powerful evidence to support the use of context-free grammar methods in generating generic story types (e.g., Shakespearean Tragedy) using locations as anchor points for the perceptual consumption and narrative development of story arcs.

Another issue important for enhancing the user experience is to set up user-driven tension bows in sub-arcs and stories. In other words, users should be empowered to give an immediate rating upon experiencing an expression (either in isolation or part of a story) that reflects their interest in what they are viewing. This kind of information acts not only as a quality-control incentive mechanism for users to exert effort in making high quality expressions with metadata, but also provides a story-quality ranking mechanism that permits the system to draw higher quality stories for presentation. To what extent system interactivity and user-driven tension bows can provide a better user experience remains an open question, one that depends on the rate at which tools that allow creating high-quality location-based media expressions are developed and deployed.

Acknowledgments

This work is part of the Amsterdam Living Lab project (PID07071), and funded by the Dutch Ministry of Economic affairs and Amsterdam Topstad. The authors thank Amsterdam Innovation Motor (AIM) for their support.

References

1. Black, J.B., Wilensky, R.: An evaluation of story grammars. Cognitive Science 3(3), 213–229 (1979)
2. Chatman, S.B.: Coming to Terms - The Rhetoric of Narrative in Fiction and Film, pp. 109–123. Cornell University Press, Ithica (1980)
3. El Ali, A., Nack, F., Hardman, L.: A location is worth a thousand experiences: Implications for location-based multimedia messaging. In: Proceedings of the Twelfth international Conference on Multimodal Interfaces, Beijing, China, November 8-12 (2010)
4. Fujita, H., Arikawa, M.: Animation of mapped photo collections for storytelling. IEICE Transactions 91-D(6), 1681–1692 (2008)
5. Heyer, L.J., Kruglyak, S., Yooseph, S.: Exploring expression data: Identification and analysis of coexpressed genes. Genome Research 9(11), 1106–1115 (1999)
6. Johnston, J.: Lecture on the odyssey (2004), `http://records.viu.ca/~johnstoi/introser/homer.htm`
7. Kullback, S., Leibler, R.A.: On information and sufficiency. The Annals of Mathematical Statistics 22(1), 79–86 (1951)
8. Lakoff, G.P.: Structural complexity in fairy tales (The study of man). UC Irvine Press (1972)
9. Landry, B.M.: Storytelling with digital photographs: supporting the practice, understanding the benefit. In: CHI Extended Abstracts, pp. 2657–2660 (2008)
10. Lucas, D.W.: Aristotle Poetics Introduction, Commentary and Appendixes. Oxford University Press, Oxford (1968)
11. Multisilta, J., Mäenpää, M.: Mobile video stories. In: Tsekeridou, S., Cheok, A.D., Giannakis, K., Karigiannis, J. (eds.) DIMEA. ACM International Conference Proceeding Series, vol. 349, pp. 401–406. ACM, New York (2008)
12. Nack, F., Jones, S.: Nomadic information in social environments. In: Proceedings of the 4th Taiwanese-French Conference on Information Technology (TFIT 2008), pp. 256–266. National Taiwan University (2008)
13. Nguyen, X., Kambhampati, S.: Reviving partial order planning. In: IJCAI, pp. 459–466 (2001)
14. Osgood, C.E., Suci, G., Tannenbaum, P.: The measurement of meaning. University of Illinois Press, Urbana (1957)
15. Pérez y Pérez, R., Sharples, M.: Three computer-based models of storytelling: Brutus, minstrel and mexica. Knowledge-Based Systems 17(1), 15–29 (2004)
16. Propp, V.A.: Morphology of the Folktale. Publications of the American Folklore Society, University of Texas Press (1928)