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 to explore unknown territory either through a comparison between what we perceive with what we know or through the encounter of stories that tell us about people's experiences at this location some time in the past. Until recently we could access a location's memory mainly through media surrogates, such as travel literature recording the experiences of an author. Examples of this include travel guide books, of which Baedeker (art) or Lonely Planet (lifestyle) are the best known. Moreover, one can view photos, drawings, films or audio files to gain an idea about a location. These type of presentations usually help for the preparation of a visit, but not the direct experience of a location. The other way to explore a location is through face-to-face encounters with people who are able to knit us into the rich but hidden experience fabric of a place, 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 (see Fig. 1). 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. Location-based content anchored at popular locations already reveals

[1] `http://www.flickr.com`; last retrieved: 11-06-2010
[2] `http://www.facebook.com`; last retrieved: 11-06-2010
[3] `http://www.myspace.com`; last retrieved: 11-06-2010

to visitors myriads of digital experiences that clutter the visitor's perception of the location. An example of this semantic overbearing is the ∼20.000 hits returned by a Flickr search for "de Dam", the central square in Amsterdam.

Fig. 1. A mock-up illustrating a digital graffiti overlay.

To help people better understand a location, we can make use of this large and continuously growing digital experience space. However, this available material was likely not created with the intention of being presented in a context beyond its intended one. Thus, computationally 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 a city space guided by experiences of others, to do so without restricting their experiential ability. The described system emphasizes location exploration by leading users along interesting locations, where interesting is defined as the amount of digital 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. To achieve this, we use the well-known classic Odyssey structure, which forms a poetic, non-linear plot of events as part of a journey [6].

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

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

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.

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 a mobile research platform developed on the Android operating system. It 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 adapt MOCATOUR's creation/presentation interface to create stories from the experience representations. Below, we introduce the creation/presentation interface followed by the story engine.

3.1 Creation and Presentation Interface

The MOCATOUR annotation interface [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

[5] Mobile Cultural Access for Tourists - `http://mocatour.wordpress.com`; last retrieved: 11-06-2010

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 (see Fig. 1). 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 basic material used by the story 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 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. Figure 2 depicts the graphical representation of the Odyssey structure.

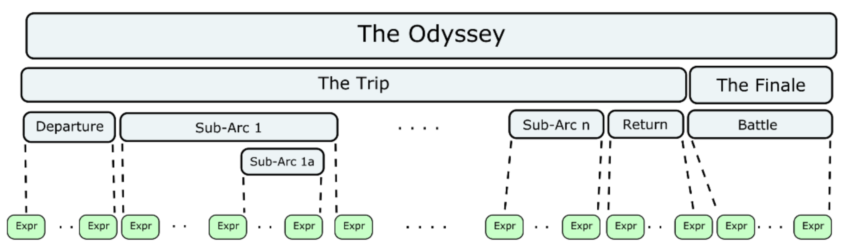

Fig. 2. Graphical representation of story structure of type 'Odyssey'.

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. We express this story form using a context-free grammar, where Fig. 2 provides a formal definition in Backus-Naur Form.

```
<story> ::= <arc>
<arc> ::= <sub-arc>*
<sub-arc> ::= [<sub-arc>|<e []>|<e [R]>]*
<e []> ::= expression
<e [R]> ::= expression with requirement R
```

Fig. 3. Backus-Naur Form of the general story structure.

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. On a global syntactic scale, a particular type of story is defined by the types of sub-arcs that it consists of as well as their quantity. Fig. 3 presents the definition of the particular story type we employ, where the adaptation is mainly provided on the sub-arc level.

```
<odyssey> ::= <departure><sub-arc-*><return>
<battle> <sub-arc-*> ::= <sub-arc-(*-1)><sub-arc>
<sub-arc> ::= <departure>|<love>|<battle>|<illusion>|<return>
```

Fig. 4. Backus-Naur Form of the overall ‘Odyssey’ structure.

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. Fig. 4(a) shows the abstract definition of sub-arc, whereas 4(b) presents an example of a particular type of sub-arc.

```
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. 5. (a) BNF of an abstract sub-arc. (b) BNF of a composite sub-arc.

The general form of metadata used in our system are tags applied to media expressions (see Sect. 3.2.2). The definitions in Fig. 4 represent the constraints on tags, 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 abstract sub-arc definition in Fig. 4(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 Fig. 4(b) only requires the opening and closing expressions to be tagged correctly. In Fig. 5 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. 6. 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 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 can contain is directly linked to the quality of the available metadata associated with the available media expressions.

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 narranotations.

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. For the denotation and connotation categories, at least one instantiation should be provided during the image creation process so that the expression can be used as part of a story. However, in ‘story-to-go’, this requirement is not enforced on the user. Rather, 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 the image, gender count in the 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. 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: Within ’story-to-go’ users can generate images that represent an experience at a particular location. They can also add metadata from the 3 annotation categories at any time. The annotation process is highly dynamic, as every added image or a change in related metadata changes the expressiveness of a location. In ‘story-to-go’ this dynamism 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.

As the aim of ’story-to-go’ is to guide visitors through currently interesting places in a city, 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 Fig. 7.

Fig. 7. Preprocessing scenario. Left: example data set. Middle: after clustering. Right: after sub-arc finding.

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 Sec. 3.2.1. This involves searching for defined configurations of annotations (metadata tags) in order to identify expressions that can be placed in a sub-arc. When multiple 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 Sec. 3.2.1. and the location of hypespots. The system then presents the material as an Odyssey narrative across different hypespot locations. The schemas shown in Fig. 8 illustrate possible story paths across hypespots.

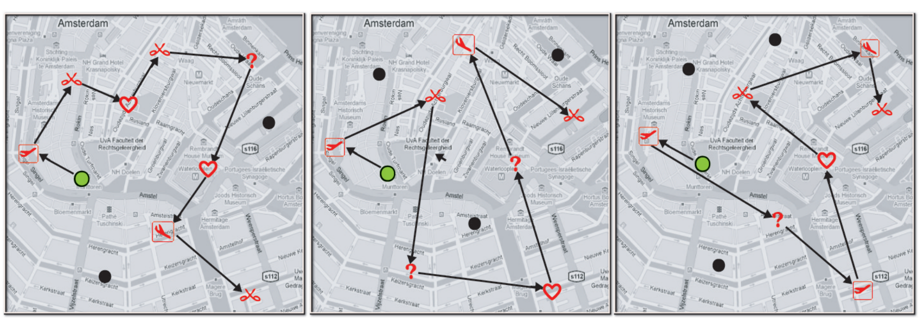

Fig. 8. Three possible stories.

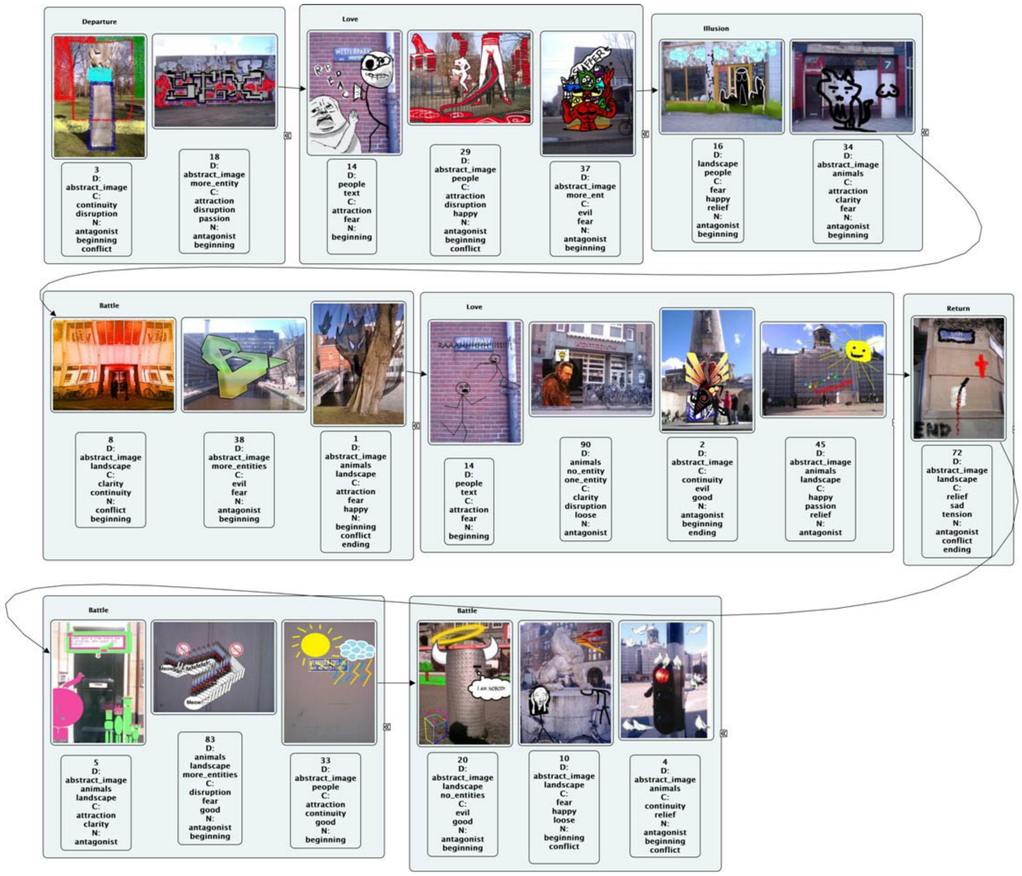

Fig. 9. Example story resulting from our second evaluation.

3.3 Evaluation

We implemented a proof of concept of both the preprocessing and the story engine in SWI-Prolog [6] and performed two evaluations. 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.

As the generated stories in the simulation lacked real visual representations, we set up a second evaluation. We asked 50 people to provide us for particular

[6] http://www.swi-prolog.org/

locations their media expressions and related metadata. The provided locations, presented as camera snapshots, were geo-tagged using the GPS functionality on an Android-based mobile device. The participants however were not restricted to generate the expressions (to be overlaid on the provided locations) using the MOCATOUR creation tool. The outcome was 84 images with metadata sets, of which the story engine could generate the structures shown in Fig. 8. Confronted with resulting story, the participants could identify the story segments (e.g., this segment is about love; here something violent is happening) but not always identify the story as an Odyssey. However, after being told that the story reflects an Odyssey it was easier for them to relate the images with each other.

4 Conclusions and Future Work

Using the structure of an Odyssey, we have developed a method for presenting an associative location-based story using user-generated media content, where the extent of the generated stories is based on the amount and recency of user activity. The proposed method makes use of a grammar-inspired compositional story representation based on sub-arcs. This resembles the syntax of regular stories and provides a way to reason about global meaning on a local level. The semantics of stories are rooted in the metadata of the content that is included. While this is what is used to compose stories, mental associativity from the user is required for the story to be perceived as fully coherent.

The evaluations we carried using the system generated stories were not totally based on real-world user-generated content. This was because geo-tagged media expressions required for the system to generate stories as yet do not exist. Furthermore, motivating users to supply the necessary metadata poses an ongoing research challenge. However, if incentive mechanisms that raise awareness in users about what kinds of stories the system can generate if they do so, it might gradually motivate them to tag more often. Despite these limitations, our proposed method demonstrated a promising 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.

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.

5 Acknowledgments

The MOCATOUR project is funded by the Dutch Ministry of Economic Affairs, Amsterdam Innovation Motors, and Amsterdam Topstad.

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. Cornell University Press (1980)
3. El Ali, A., Nack, F., Hardman, L.: A location is worth a thousand experiences: Design implications for location-based experience capture systems. In: Proceedings of the eighteenth ACM international conference on Multimedia (2010; submitted)
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 (November 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 (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 (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, IL (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)