

A Thematic Approach to Emerging Narrative Structure

Charlie Hargood
Learning Societies Lab
School of Electronics and
Computer Science
University of Southampton
+44 (0)23 8059 7208
cah07r@ecs.soton.ac.uk

David E. Millard
Learning Societies Lab
School of Electronics and
Computer Science
University of Southampton
+44 (0)23 8059 5567
dem@ecs.soton.ac.uk

Mark J. Weal
Learning Societies Lab
School of Electronics and
Computer Science
University of Southampton
+44 (0)23 8059 9400
mjlw@ecs.soton.ac.uk

ABSTRACT

In this paper we look at the possibility of using a thematic model of narrative to find emergent structure in tagged collections. We propose that a thematic underpinning could provide the narrative direction which can often be a problem with stories from existing narrative generation methods, and present a thematic model of narrative built of narrative atoms and their features, motifs and themes. We explore the feasibility of our approach by examining how collaborative tags in online collections match these properties, and find that while tags match across the model the majority are higher level (matching broader themes and motifs rather than more specific features) which may require further investigation into their utility.

Categories and Subject Descriptors

H.1 [Models and Principles]: General.

General Terms

Standardization, Human Factors, Experimentation.

Keywords

Narrative, Narrative Generation, Thematics, Emergent Structures

1. INTRODUCTION

User generated content on the web (such as blog entries, photos, videos, etc) are often accompanied by explicit virtual structures in the form of tags (overlapping named collections) and other domain-specific collections such as albums (for photos), channels (for videos) and reading lists (for books). These explicit structures give rise to other emergent structures sometimes referred to as folksonomies, such as emergent vocabularies (e.g. tag clouds) and taxonomies (e.g. Wikipedia categories).

Folksonomies have several advantages over generated metadata as they permit richer semantic annotation of collections as opposed to auto-generated metadata [2].

Narratives (or stories) are a commonly prevalent form of

information representation that are well established as an engaging way of representing an experience. Narrative generation is a field that seeks to explore alternative representations of narrative, and investigate the possibility of automatically generating custom stories from information collections. There are a wide variety of different techniques for narrative generation ranging from structured narrative grammars to emergent narratives. However the narratives generated can seem flat, lacking engagement and direction.

In our work we are exploring a thematic approach to solving some of the problems with narrative generation. The thematic approach focuses on themes within a story to give narratives a sense of direction and purpose. For example, rather than simply recounting photographs taken during a holiday in chronological order, it might emphasize photos with themes such as relaxation or celebration to create alternative narratives based on the same resources and the same events. It is our belief that the thematic approach would generate richer stories that benefit from a thematic subtext.

In this paper we present a model representing the thematic element of narratives. We also explore whether these elements map to the tags found on shared online resources in order to explore the feasibility of our approach as a method of creating emergent narrative structure from collaboratively tagged materials.

2. BACKGROUND

2.1 Narratology

2.1.1 Structuralism

Narratology is the study of narrative within literature. It is primarily focused on narrative analysis and on deconstructing existing narratives but it also provides a rich theoretical basis for narrative generation techniques.

Structuralism is an approach to narrative analysis that aims to deconstruct narrative and to learn about the components from which a story is built and how they are connected and contrasted against each other within a narrative work. For narrative generation this approach is particularly attractive as it defines tangible objects within a narrative that can be modeled and used to represent parts of a generated narrative.

Structuralism asserts that a narrative may be deconstructed into a *story* and a *discourse* [4] where the story represents a chronology of all the information to be communicated and the discourse represents what parts of the story are told and how those parts are presented (shown in Figure 1).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WebScience '08, June XXX, 2008, Pittsburg, Pennsylvania, USA.
Copyright 2008 ACM 978-1-59593-XXX-X/08/06...\$5.00.

The story element of this narrative is constructed by the observations and experiences that make up the subject of the narrative. In a virtual collection of resources the story represents the collection itself, containing all observations and experiences. The discourse however represents what parts of the story are told and how it is told; if the collection is the story then the result of narrative generation (telling the story) is the discourse.

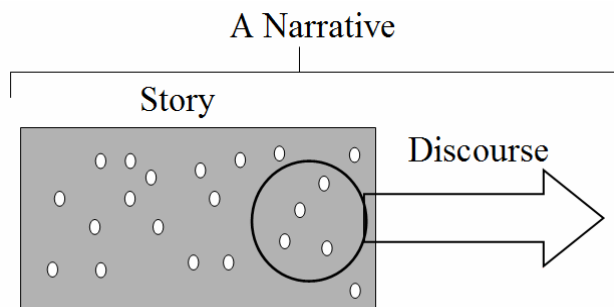


Figure 1. A Narrative can be deconstructed into Story and Discourse

The discourse is the result of a multitude of different mechanics including how the story is presented, what medium is used, the style, the genre, and the themes of the narrative. Thematics is the discipline of approaching themes within narrative in a structuralist way, deconstructing and analyzing the relations between the components that communicate a theme within a narrative.

Tomashevsky deconstructed thematic elements into themes (broad ideas such as ‘politics’ or ‘drama’) and motifs (more atomic elements directly related to the narrative such as ‘the helpful beast’ or ‘the thespian’) [11]. He describes how themes are constructed from other themes (sub themes) and motifs. A motif is the smallest atomic thematic element and refers to an individual element with the narrative which connotes in some way the theme. Themes may always be deconstructed into other themes or motifs whereas a motif may not be deconstructed.

2.1.2 Semiotics

Semiotics or semiology is the study of signs and how we extract meaning from them. Saussure wrote that all signs are built of two parts [9], a signifier (the physical signal from the sign such as the appearance of an apple) and a signified (the denotation of that sign such as the concept of ‘apple-ness’ or ‘fruit’).

Barthes made a distinction between denotative signs (signifiers that lead directly to their signified, such as a word having a literal meaning) and connotative signs (signifiers that lead indirectly to some contextual or culturally important signified, such as the red light implying Stop to a driver) [5]. Barthes goes on to point out that should a sign *connote* something then the signifier of such a sign would itself be built out of a denotative sign (a picture of a red light *denotes* a red light, red light *connotes* Stop). In such a way we can draw contextual cultural concepts from static basic objects that in a particular context have a greater meaning.

2.2 Narrative Systems

Narrative generation has been the focus of a wide range of systems, varying in both their methods and application. Some

systems use narrative generation techniques as a way of adding more meaning to information, for example Topia [3] where search results are presented as a discourse. Using narrative as a way of representing information in this way is similar to various hypertext projects such as AHA! [7] where the omission, emphasis, and spatial presentation of information creates a discourse and gives the information presented additional meaning. In other systems, such as the virtual storyteller [10], the aim is to completely generate an entertaining story rather than represent existing content.

There are many different methods used to generate narrative but they largely fall into two types; grammar narratives, and emergent narratives. Grammar narratives work by extensively modeling the rules of a given genre and using structuralism to create a grammar of narrative elements. A story is then generated by fitting prewritten narrative segments together by identifying which narrative element they match and following the rules of the grammar. An example of such a system is Artequakt [1] and to an extent Card Shark [6]. Emergent narratives generate a story by simulating the story setting and simply presenting what occurs. Often this is achieved by using agents to play the parts of characters within a story that follow the rules of the environment and using a director agent to influence the actor agents into a creative narrative. Examples of such emergent narratives are Façade [8] and the Virtual Storyteller [10].

While existing techniques often succeed in generating narratives they have several drawbacks. Narratives generated from story grammars are heavily bound to the rules of a given genre and become very formulaic, and emergent narratives can seem like a bland account of a set of actions as the generation is based on a simple report of what happened in sequence, and as such lacks emphasis and flavor. Both techniques generate narratives that can tend to lack any authorial voice, leading to narratives without any emphasis, creating stories without an objective that can seem directionless. A human author imbeds meaning, subtle themes, and his/her own goals into a piece – these are lacking in any computer generated narratives. If direction, emphasis, or the authorial voice could be incorporated into generated narratives then it would lead to less bland or formulaic stories.

3. A THEMATIC APPROACH

3.1 The Model

Authors use themes to communicate a subtext within a narrative. This subtext may be an agenda or simply an emphasis of a particular part of the narrative or even simply an emphasis of the author's own style. This subtext gives a narrative direction beyond merely communicating a chronology leading to deeper narratives and giving an authorial voice to stories. We propose a thematic underpinning to narrative generation techniques so that richer narratives with direction may be generated.

To do this we go back to Tomashevsky's structuralist work on thematics. *Features* within the narrative denote *Motifs* and from these *Themes* can be identified.

We assume a situation where a story is compiled with many small segments of narrative that are structured together, in this case the selection of these small atomic segments and their content are key to communicating a theme. We use the term Narrative-Atoms or *Natoms* to describe these segments; small atomic pieces of

narrative that cannot be further broken down, for example a single photo or paragraph. The content of these natoms is rich with information, however only some of it visible to a machine (such as generated metadata and authored tags), we call these visible computable elements *Features*. Natoms contain any number of features which may or may not work towards connoting a theme in a story. Features can each denote a motif, a basic thematic object that has connotations within the story, for example the feature *cake* denotes the motif of *food*. These motifs in turn connote broader themes in the context in which they are presented, for example *food* in the context of a gathering may connote *feasting*. These themes, when combined with other themes or motifs could in turn be used to further connote other themes, for example *feasting* might connote *celebration*.

This forms the foundation of our thematic model of a narrative:

- Natoms contain tagged features
- Features denote motifs
- Themes are connoted by other themes and motifs

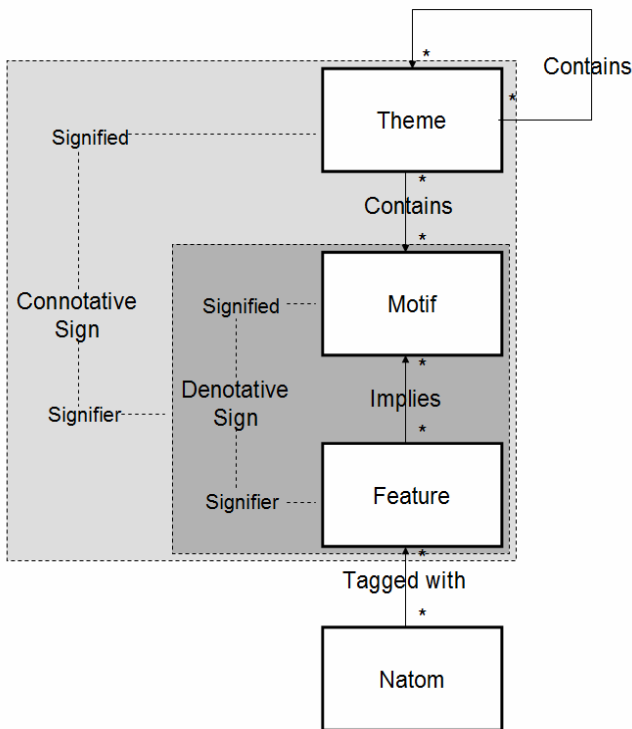


Figure 2. The Thematic Model

The model is shown in Figure 2, which also shows how the parts of the model map to Barthes' idea of denotative signs as the signifiers for connotative signs. Features denote Motifs because motifs are directly associated with the feature (normally as a generalized version of it). Themes are broader concepts communicated over the entirety of the narrative, typically by numerous motifs. By their nature they cannot be denoted as they rely on some cultural context which cannot be contained within

a natom, as such a theme is a connotation of the motifs, and by extension the features, within the narrative.

This model is but one part of a narrative generation system, it contains no rules for the presentation of elements or the narrative structure. However it can be used to select natoms to be used within a discourse. As such we could use themes constructed from this model to influence the story selection in grammar or emergent narratives to give them a thematic subtext.

When a narrative is formed a part of the story is selected and then presented as a discourse [4]. We can consider virtual collections of resources as our story, and should we want to create a discourse to tell a story of Tuesday it would select all the natoms (photos, blog entries, etc.) of that day. Using an appropriately populated thematic model we could examine the features of those natoms in order to identify motifs and thus potential themes. Natoms that connote these popular themes could then be selected or emphasized to create a final discourse that felt more purposeful. If the virtual collections were very large we could set out to look for natoms that supported particular themes, for example, by using public photo collections to create a discourse (a photo montage) with the themes of *family*, *winter* and *Christmas*.

Because features could be tagged in any way for such a system to work every motif object would need a list of features that could denote the motif. In turn theme object will also require some way of knowing what motifs are suitable for them, however in this case it is less simple as themes are contextual things not simply denoted. It seems likely that a theme should be described as having core thematic elements that are required for a theme to be communicated, such as a *wedding* theme requiring a *bride* motif, as well as optional thematic elements that exaggerate or promote the theme but are not essential (such as a *religious* theme). Themes would need to keep a set of required and optional thematic elements (both motifs and sub themes). The power of the thematic approach will be proportional to the quantity and richness of these feature-motif and motif-theme connections.

3.2 An Example

Figure 3 shows an example of an instantiated model. There is a hierarchy of themes (white boxes), motifs (grey boxes) and features (dark grey boxes) under the overall themes of Celebration and Spring. As you can see each of these themes is made up of 2 sub-themes and a motif, the sub-themes in turn are made up of a few motifs and each motif is implied by a feature that could easily be tagged in a natom.

Figure 3 also shows that motifs can be part of a theme but are in no way bound to it. In the example the inclusion of a party motif could be used to connote either the Birthday or Easter themes. Furthermore motifs may be denoted by any number of features, the example shows how a party motif could be denoted from either a Champagne feature or a Balloon feature.

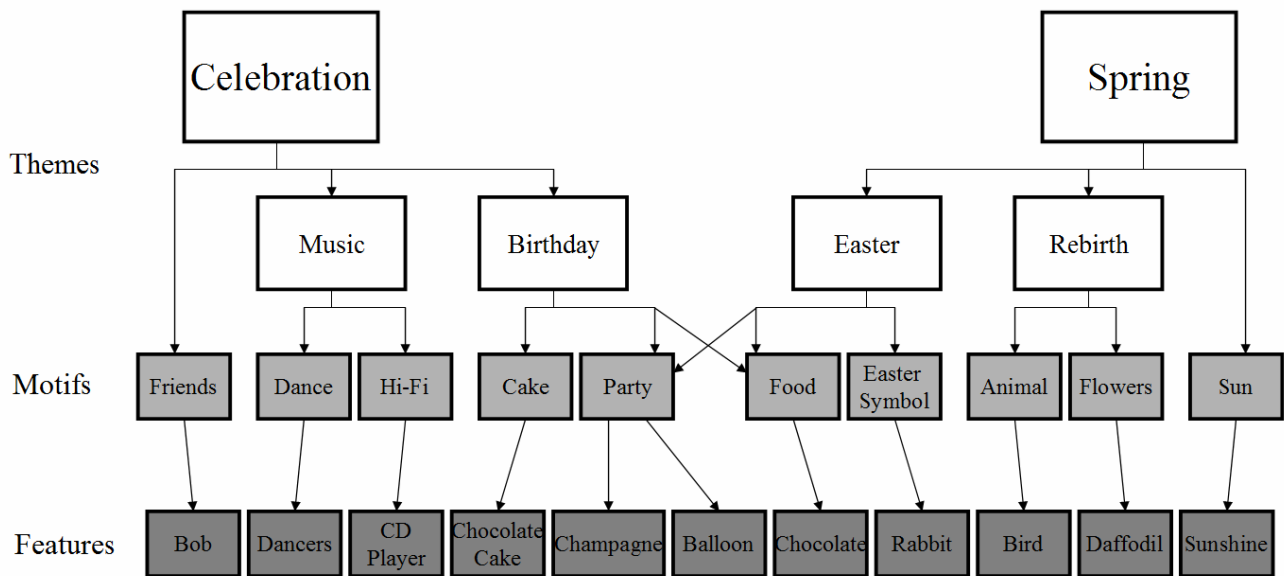


Figure 3 Worked Example

4. NARRATIVES AS EMERGENT STRUCTURES

The thematic model described in Section 3 could be used to influence narrative generation given a set of natoms with appropriate metadata, but how well might current tagging behavior support this?

Any user generated virtual collection is an account of some human experience and as such should contain a potential narrative; in a sense every blog, photo album, and video has a story to tell. But the generation of our own custom narratives from these collections depends on the quality, quantity and nature of the metadata available. For example if natoms were to be tagged mainly with themes then a narrative generation system could find itself starved of features to connote other themes. While the tagging is still relevant to the thematic approach a theme has very few connections, whereas a natom tagged with a feature could be used to denote many different motifs and as such connote many themes. Referring to our example in Figure 3 if a photo of a bottle of champagne was tagged as “Celebration!” it would be accurate, however in this case the photo could only be used to connote a theme of celebration whereas if the same photo were tagged as “Champagne” it could also denote the motif of “Party” and subsequently connote the themes of both “Celebration” and “Spring”. As such it is important to measure where tags fall within the model so that their utility at constructing themes may be assessed.

To evaluate the feasibility of applying this model and generating narratives from virtual collections we decided to survey some existing collections to see how natoms were tagged, with features, motifs, or themes. Using the example above we searched for images on Flickr¹ with each feature, motif, and

theme, searched for tags of items only, and comprised a table of the average number of results for each main theme (Celebration and Spring) from the example. We also modeled a few other super themes (Winter, Hedonism, and Childhood) and surveyed the results for them also. The results are in Table 1, Averages rounded to nearest value.

Table 1. Average Tag Types

Main Theme	Themes	Motifs	Features
Celebration	1,915,532	1,929,864	44,557
Spring	503,078	1,830,234	214,397
Winter	1,601,127	1,365,610	39,866
Hedonism	8,940	1,800,366	73,384
Childhood	615,775	346,701	204,390

The results shown are varied, all thematic elements appear as tags, however themes and motifs seem more popular than features (with the exception of Hedonism). It is possible that this is because users tag collections more frequently with broader concepts (themes and motifs) rather than identifying specific elements (features). If true this could make using our thematic model difficult as the relatively few number of features tagged might prove insufficient to build suitable themes.

However this could also be due to our search being driven from the top of the model (by themes). Any motif can be denoted by a wide range of features, so for any given motif there may be more feature tags in total than tags for the motif itself. If we were to search for a full range of features that could denote these motifs then we would almost certainly find a more even spread of feature and motif/theme tags.

¹ <http://www.flickr.com>

5. CONCLUSION AND FUTURE WORK

In this paper we have presented a thematic model of narrative based on the work of structuralism in narratology. Our model consists of narrative atoms (natoms), features, motifs and themes. We believe that the model could be used to create emergent narrative structures from virtual collections that may have more focus and direction than narratives created just through existing approaches.

The thematic model can be used to select natoms that together promote given themes, but this will only result in collections or montages without some additional structure. We believe that the thematic underpinnings when coupled with discourse generation will create greater coherence and causality within generated narratives; other natom meta-data (such as format and creator) could also be considered in order to create a stronger sense of authorial voice and style.

Our initial survey of online collections has indicated that a thematic approach should be possible, although it is not yet clear whether there would be enough tags relating to features (which are more flexible than tags that relate to motifs/themes) to create narratives with an arbitrary theme. More work needs to be done to measure whether it would be viable to do this from existing collaborative tagging collections or if such an approach would be limited to collections that had been formally tagged in a specific fashion.

Our intention is to build a thematic narrative generation system based on online collections in order to investigate the effect empirically, and to evaluate the effectiveness of thematic selection in creating narratives with a perceived focus and direction

6. REFERENCES

- [1] Alani, H. Kim, S. Millard, D. Weal, M. Hall, W. Lewis, P. and Shadbolt, N. 2003. Automatic Ontology-based Knowledge Extraction and Tailored Biography Generation from the Web. *IEEE Intelligent Systems*, 18, 14-21.
- [2] Al-Khalifa, H and Davis, H. 2006. Folksonomies versus Automatic Keyword Extraction: An Empirical Study. *IADIS International Journal on Computer Science and Information Systems (IJCSIS)*, Col 1, 132-143
- [3] Alberink, M. Rutledge, L. and Veenstra, M. 2003. Sequence and Emphasis in Automated Domain-Independent Discourse Generation. *Information Systems*, 1-10.
- [4] Barthes, R. 1966. Introduction to the Structural Analysis of Narrative. A Roland Barthes Reader. Heath, S, Trans. Sontag, S, Ed. 1994. London: Vintage.
- [5] Barthes, R. 1996. *Mythologies*. Lavers, A, Trans. London: Vintage.
- [6] Bernstein, M. 2001. Card shark and thespis: exotic tool for hypertext narrative. *Proceedings of the twelfth ACM conference on Hypertext and Hypermedia*, 41-50.
- [7] DeBra, P. Aerts, A. Berden, B. de Lange, B. Rousseau, B. Santic, T. Smits, D. and Stash, N. 2003. AHA! The Adaptive Hypermedia Architecture. *Proceedings of the fourteenth ACM conference on Hypertext and Hypermedia*, 81-84.
- [8] Mateas, M. and Stern, A. 2003. *Façade: An Experiment in Building a Fully-Realized Interactive Drama*. Game Developers Conference 2003.
- [9] Saussure, F. 1974. *Course in General Linguistics*. Baskin, W, Trans. Glasgow: Fontana.
- [10] Theune, M. Fass, S. Nijholt, A. and Heylen, D. 2003. *The Virtual Storyteller: Story Creation by Intelligent Agents*. TIDSE 2003: Technologies for Interactive Digital Storytelling and Entertainment.
- [11] Tomashevsky, B. 1965. *Russian Formalist Criticism: Four Essays*. *Thematics*. Lemon, L, T. and Rees, R, J, Ed. University of Nebraska Press, 66-68