

a thematic model for narrative generation

charlie hargood, lsl, ecs, university of southampton

problem and motivation



Within dynamic hypermedia systems it is often the goal to create representations of information tailored to the users need with generated content that is both relevant and high quality. Narrative generation systems aim to generate an entertaining and engaging experience for a user by generating content within a story framework.

While narrative generation systems experience some success they often suffer from limitations in quality producing bland and unengaging stories that can lack any authorial direction. My research is in this field and aims to use themes to enrich generated narratives. I have developed a thematic model for representing themes in narrative generation that is based on work in structuralism. I have also built a prototype that utilises this model to build themed photo montages (a very basic, plotless form of narrative) in order to evaluate the models ability to represent themes that can be successfully conned.

Here I explain some of the background research behind both narrative generation and using narratology to build a thematic model, explain the unique thematic model that I have developed, and offer my largest contribution to date in the form of an evaluation of the effectiveness of the thematic approach compared to simple keyword search.

background and related work

narratology

Although narratology, as a study of literature, is mostly focused on the analysis of narrative it provides a detailed insight into how narratives are built.

One approach to narratology, structuralism, deconstructs narrative and aims to learn about the components from which a story is built and how they are connected and contrasted against each other. As this defines tangible objects within a narrative that can be modeled there is much narrative generation can use from structuralism as it can seek to generate the structures that structuralists have defined. Most structuralist theories asserts that a narrative is composed of any series of human experiences [9], and may be deconstructed into a story and a discourse [4] where the story (or fabula) represents a chronology of all the information to be communicated and the discourse (or sjuzhet) represents what parts of the story are told and how those parts are presented (shown in Figure 1).

The story element is constructed by the experiences that make up the subject of the narrative. In a virtual collection of resources the story represents the collection of experiences represented as resources. The discourse however represents what parts of the story are told (the story selection) and how it is told (the story presentation); if the collection is the story then the result of narrative generation (telling the story) is the 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. The study of thematics approaches themes with a structuralist method of deconstruction and attempts to identify the narrative elements that communicate themes.

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 a structure of themes being built out of sub-themes and motifs. A motif is the smallest atomic thematic element and refers to an individual element within 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.

narrative generation

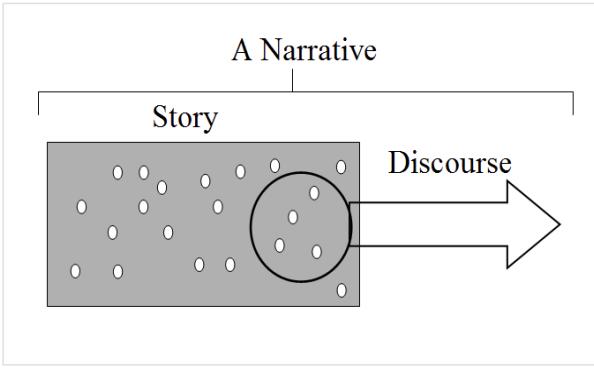


Figure 1: Story and Discourse

Narrative generation has a variety of applications in systems that deal with different information, as a narrative can be any collection of human experience it is not limited to written prose but to any representation of human experience. Some systems use narrative as a lens through which to view a larger collection, for example PhotoCopia [12] which presents narrative photo montages. Some systems generate narratives to add more meaning to information, for example Topia [3] where search results are presented as a discourse. Using narrative as a representation of information in this way is similar to various hypertext projects such as AHA! [6] where the omission, emphasis, and spatial presentation of information creates a discourse that makes the presented information more meaningful. In other systems with entertainment as an objective, such as the Virtual Storyteller [10], the aim is to completely generate an entertaining story rather than represent existing content.

Different methods of narrative generation often fall into two types; grammar narratives, and emergent narratives. Grammar narratives work by modeling the rules of a given genre and using structuralism to create a grammar of narrative elements. A discourse is then generated by fitting prewritten narrative segments together using the rules of the grammar. An example of such a system is Artequakt [2] and, to an extent Card Shark [5]. In contrast emergent narratives generate a story by presenting a simulation of the story setting, often 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 emergent narratives are Façade [8] and the Virtual Storyteller [10].

Existing techniques often succeed in generating narratives but 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 their 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.

uniqueness of the approach

The uniqueness of this approach lies in the use of thematics. Traditionally narrative generation is more concerned with the narratological and plot objectives of a story without as much focus to other messages an author seeks to imbed or anything as subtle as themes. However these existing methods can lead to bland and unengaging results, although ones with sound plots. we believe that by adding thematic objectives to the process of narrative generation we could produce richer narratives with more direction, that as a result may be perceived as higher quality.

thematic model

In previous work [7] we proposed a thematic underpinning to narrative generation in the form of a thematic model that described how themes are constructed within a narrative. The thematic model is largely based on Tomashevsky's work on thematics. The foundation of the model(as shown in Figure 2). It describes narratives as being built of \textit{natoms} (narrative atoms) which contain \textit{features} that denote \textit{motifs} which in turn connote \textit{themes}.

For example, we might view a digital photo as a natom, and the tags on that photo as the features that denote a particular motif. Thus a photo tagged with 'daffodil' could denote the motif of 'flower', which connotes the theme of 'spring'. Themes can themselves build up into new themes, for example the theme of 'christmas' can be used to connote the theme of 'winter'.

thematic builder prototype

In order to evaluate the effectiveness of the model a prototype system was built that utilised an instance of the model. The prototype uses the model to select images from Flickr that have strong relevance to particular themes. The prototype went under the working name of the Thematic Model Builder (TMB).

This instance of the model was built in xml and four themes were modeled and expanded (all sub themes and motifs were modeled as well): winter, spring, celebration, and family. The process of defining an instance of the model for particular themes is a complex and subjective one [7]. We explored a systematic method for building themes based on semiotics. Initially we identify what \emph{connotes} that theme, these connotative signs will make up the themes sub themes and motifs. However, these signs become sub-themes only if all of the aspects of their concept in turn connote the theme being built, otherwise the sign should become a separate theme in

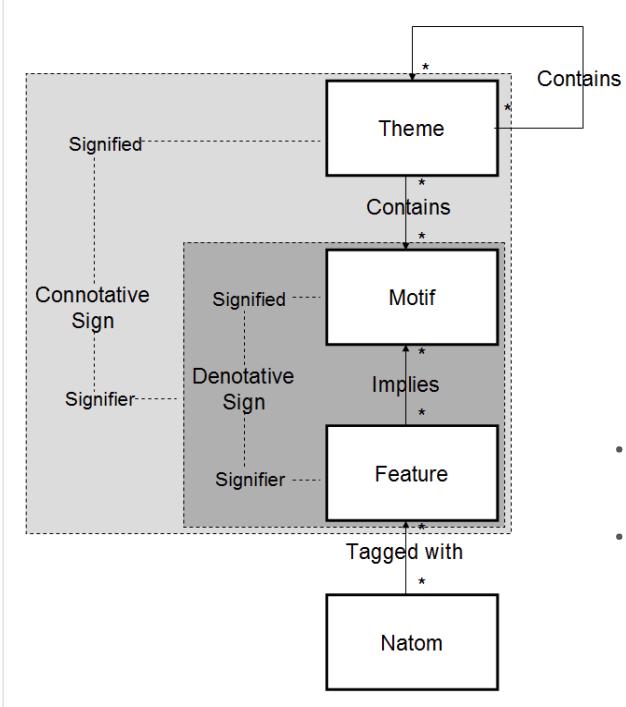


Figure 2: Thematic Model

its own right. Thematic objects anchored to a particular device within the narrative become motifs which have their features defined by likely tags that denote the object.

The prototype itself was written in java with a simple JSP front end. For the purposes of this prototype and evaluating the model, Flickr was chosen as a source of atoms. As a folksonomy its items have rich semantic annotations in metadata [1] that make the features in each image apparent and it has a large freely available body of resources. The library of images (the fabula) was generated by making a keyword search of Flickr on the desired subject and storing the top n images (where n is the desired size).

The system then followed an algorithm of measuring the thematic quality of each natom in the fabula. It returns the natoms with the highest scores according to two metrics:

- **Component coverage:** the proportion of high-level sub-themes or motifs that a natom has features for - this is useful for measuring how strongly a natom matches the desired theme. (for example, winter expands several high-level sub-theme and motifs including christmas, snow and cold. A natom matching just one of these has less coverage than one that matches many)
 - **Thematic coverage:** the proportion of desired themes that a natom has features for - this is useful for searches with multiple themes

The TMB Prototype allows us to compare the effectiveness of selecting photos according to their theme with the process of selecting photos based directly on their tags.

writing themes

Some early work has been done towards formalising the process by which an instance of the thematic model is written. The TMB works by using instances of the thematic model to interpret whether narrative objects successfully connote a theme. It is comprised of Themes and Motifs, Themes are intangible concepts have a list of other themes and motifs that successfully connote it where as motifs are more tangible devices that have a list of features that more literally denote them.

For example to write the theme of winter we would follow this series of rules to find the contents of the theme. Winter might be connoted, amongst other things, by “Christmas”, “Snow”, and “Snowflake”. “Christmas” is a high level concept, it is not anchored to any one thing within the narrative, it is a subtle and complex thing connoted by a great many ideas as well as things; it is a theme. “Snow” is more tangible, snow relates to a direct device within the story and can be anchored to a specific object within the narrative; it is a motif. “Snowflake” again is tangible but it connotes winter in the same way “Snow” does, it serves to denote the device of “snow” in the same way “snowman” might, it is a specific instance that may exist within a narrative element that serves to use “snow” to connote “winter”; it is a feature.

Rules:

List Connotations: List all concepts, objects, and words that to you connote the idea of the desired theme. List everything that you associate with it in anyway and to you helps build the idea of the theme in your head.

Divide Tangible Objects and Concepts: Divide the listed connotations into those that are anchored to specific objects and devices that could be included in a narrative element and those that are broader concepts connoted by many things and less tangible. These broader concepts become Themes.

Group motifs: Group similar tangible objects together. Consider the relationship the object has with the desired theme and group together objects that belong to the same narrative device. For example, in the theme of picnic “chicken” “sandwich” and “scotch egg” all serve the same purpose of denoting “food”. These grouped together objects become your themes motifs.

Iteratively write the contents for sub-themes and motifs: For each theme and motif repeat step 1. For motifs this will be slightly different, as you are not considering a desired high level concept but a much more tangible object you will be listing denotations not connotations, for example list every specific object that might exist in a narrative element that would lead to denoting this concept. Be careful only to list things that directly denote the motif, not associated words, these are the motifs features. For themes the process is identical as step 1 was before. Repeat this step until all sub themes and motifs have been iterated through and written.

Identify associated themes and motifs: Check the components of every sub-theme and motif of the desired theme, and in turn every sub-theme and motif of each of those. Ensure the entire contents of a sub-theme or motif is relevant to the parent theme and in turn connotes the parent theme. A sub-theme (or motif) that contains elements irrelevant to the parent theme becomes an associated theme and is removed from the model.

results and contributions

The most important results and contributions in this area so far are of the initial evaluation of the thematic model. The results for the pilot study of 22 individuals are presented here, the full evaluation of 100 individuals is nearly complete, and early analysis shows that the results are similar to the pilot study.

evaluation

For the evaluation it was important to measure what advantage there was in using a thematic system for natom selection over a keyword search system, but we also wanted to see whether themes emerged more strongly from groups of images than with individuals.

The evaluation asked participants to rate images individually and in sets according to how they matched a given subject and theme (for example, 'London in Winter'). The images and sets were generated in four different ways:

TMB: Using the TMB and Flickr API to search by subject and select by component coverage

Flickr: Using Flickr to search by subject and theme, filtered by relevance

BaseL(ow): Selecting images from Flickr at random

BashH(igh): Using Flickr to search tags by subject and filter manually

In this way we hoped to compare the performance of the TMB with keyword search on Flickr, and place both of these methods in context by comparing them to random and hand-picked samples. For each test the user would be presented with two titles and under each the images for the test (depending on the test either individually or in groups) and asked to rate them 1-5 on their relevance to the title. To ensure the data was representative we chose titles composed of contrasting themes and fabulas as well as well matched themes and fabulas. We also included titles that included more than one theme in separate tests.

In order to make the evaluation fair we presented the single image text first (so participants would not already have associated them with a group). The images on the single image test were also randomly shuffled and for the group tests we randomised the order in which sets appeared. We also added a restriction on image groups that no more than one image would be allowed per author - this is because image sets published by an author naturally flow and would artificially seem to be stronger montages. Finally users were only allowed to take the evaluation once, a unique evaluation link for each user was given out per email address.

Each test contained two titles composed of different subjects and themes from the four the TMB prototype was able to use, in each test one title paired the theme with a complementing fabula, the other title paired the theme with a contrasting fabula to observe performance under different conditions. The titles chosen for single themes were London in Winter, Celebration and Earthquake, Spring Picnic, and Family Factory and for multiple themes My Family in New York at Winter, and Celebrating the New House in Spring.

Our pilot study was performed with 22 users. While this is a relatively low number of people it still gave us a large amount of data, as each user was asked to rate 40 images and 4 groups for each of the 4 sources. This resulted in 880 data points for single images and 88 for groups, enough for quantitative significance to emerge (which we measured with a t test).

evaluation results

The data from the pilot evaluation show some significant results. The mean rating of natoms from the TMB is higher than that for a keyword search (Flickr) in both single and group images. Figure 3 and Tables 1 and 2 show the data and t-tests for single images. Figure 4 and Tables 3 and 4 show the data and t-tests for grouped images. The hypothesis that the TMB selects natoms more relevant to the title than a keyword search is true with only a 2.5 percent probability of error for both group and single images.

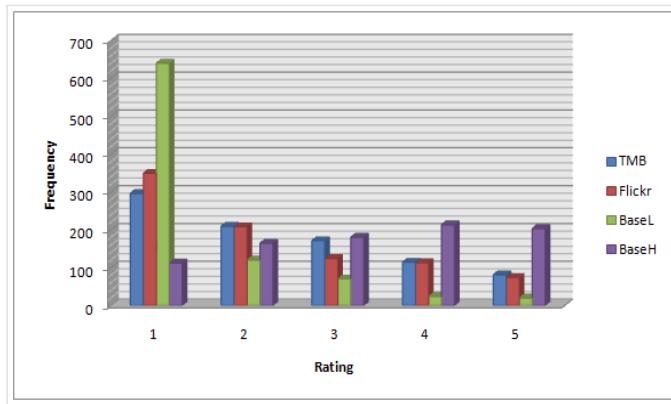


Figure 3: Single Image Rating Frequency

Set	1	2	3	4	5	Total
TMB	296	209	171	114	81	871
Flickr	349	208	124	113	74	868
BaseL	640	120	70	24	20	874
BaseH	112	164	180	213	203	872

Table 1: Single Images Rating Frequency

Set	Mean	SD	Variance
TMB	2.397	1.319	1.741
Flickr	2.257	1.330	1.769
BaseL	1.471	0.926	0.850
BaseH	3.264	1.345	1.809

$t=2.207$, $df=1737$, $p=0.025$

Table 2: Single Images Rating Statistics

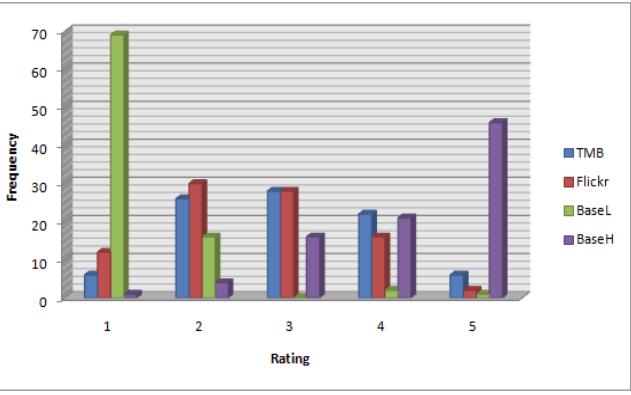


Figure 4: Grouped Image Rating Frequency

Set	1	2	3	4	5	Total
TMB	6	26	28	22	6	88
Flickr	12	30	28	16	2	88
BaseL	69	16	0	2	1	88
BaseH	1	4	16	21	46	88

Table 3: Grouped Images Rating Frequency

Set	Mean	SD	Variance
TMB	2.955	1.049	1.101
Flickr	2.614	1.011	1.021
BaseL	1.295	0.697	0.486
BaseH	4.216	0.976	0.953

$t=2.182$, $df=172$, $p=0.025$

Table 4: Grouped Images Rating Statistics

At first glance the difference between the TMB and Flickr only appears to be slight however it must be seen in the context of the difference in results between a best case scenario (human selection: BaseH) and a worst case scenario (random selection: BaseL). The ranges between these are rather smaller than we might expect, and in this context the improvement given by TMB is rather more impressive.

As expected the results also show that the TMB proves significantly better in a montage context where it can build themes over a group of natoms, a t-test shows this hypothesis to be true with only a 0.05 percent probability of error. In addition the data shown in table 5 reveals that while both a keyword search and TMB improved when their natoms were presented as a group the TMBs improvement was much more significant, the hypothesis that the TMBs improvement was greater than the improvement of a keyword search in a group context is shown with this data to be true according to a t-test with a 0.5 percent probability of error.

These results offer encouraging observations towards two of our evaluation objectives. The TMB seems to be performing better than a keyword search with some significance and further more it seems the TMB is very strong within a group context, this could lead us to believe it could perform similarly strongly within a narrative context however a full evaluation would be necessary to confirm these initial quantitative findings as well as answer further evaluation objectives to refine the process of calculating thematic quality.

Conclusion and future work

The early results show that the TMB performs strongly in comparison to simple keyword searching, and that our thematic model can successfully be used to connote themes in a simple montage. This suggests that it is worth exploring its effect with a more sophisticated narrative generation system.

The full evaluation experiment for this research is nearly complete and the results continue to show promise. The next step after completing a full evaluation of the model is to experiment with integrating this system with full narrative generation and research has already begun on exploring how to perform such an integration. Work is also underway to formalise the process of authoring themes to assure the quality of other instances of the model. Thematics could well lead to improving the quality of narrative generation which in turn could lead to adaptive hypermedia systems that are both uniquely relevant to a user's requests as well as engaging.

References

-
- [1] H. Al-Khalifa and H. Davis. Folksonomies versus automatic keyword extraction: An empirical study. *IADIS International Journal On Computer Science And Information Systems (IJCISIS)*, 1:132-143, 2006.
 - [2] H. Alani, S. Kim, D. Millard, M. Weal, W. Hall, P. Lewis, and N. Shadbolt. Automatic ontology-based knowledge extraction and tailored biography generation from the web. *IEEE Intelligent Systems*, 18:14-21, 2003.
 - [3] M. Alberink, L. Rutledge, and M. Veenstra. Sequence and emphasis in automated domain-independent discourse generation. In *Information systems*, pages 1{10, 2003.
 - [4] R. Barthes and L. Duisit. An introduction to the structural analysis of narrative. *New Literary History*, 6:237-272, 1975.
 - [5] M. Bernstein. Card shark and thespis: exotic tools for hypertext narrative. In *Proceedings of the twelfth ACM conference on Hypertext and Hypermedia*, 2001.
 - [6] P. DeBra, A. Aerts, B. Berden, B. de Lange, B. Rousseau, T. Santic, D. Smits, and N. Stash. Aha! the adaptive hypermedia architecture. In *Proceedings of the fourteenth ACM conference on Hypertext and hypermedia*, pages 81{84, 2003.
 - [7] C. Hargood, D. Millard, and M. Weal. A thematic approach to emerging narrative structure. In *Web Science at Hypertext08*, 2008.
 - [8] M. Mateas and A. Stern. Facade: An experiment in building a fully-realized interactive drama. In *Game Developers Conference*, 2003.
 - [9] M. McQuillan. *The Narrative Reader*. Routledge, London, 2000.
 - [10] M. Theune, S. Faas, A. Nijholt, and D. Heylen. The virtual storyteller: Story creation by intelligent agents. In *TIDSE 2003: Technologies for Interactive Digital Storytelling and Entertainment*, 2003.
 - [11] B. Tomashevsky. *Russian Formalist Criticism: Four Essays*, chapter Thematics, pages 66{68. University of Nebraska Press, 1965.
 - [12] M. Tuffield, S. Harris, D. P. Dupplaw, A. Chakravarthy, C. Brewster, N. Gibbins, K. O'Hara, F. Ciravegna, D. Sleeman, Y. Wilks, and N. R. Shadbolt. Image annotation with photocopain. In *First International Workshop on Semantic Web Annotations for Multimedia (SWAMM 2006)* at *WWW2006*, Edinburgh, United Kingdom., 2006.
-