The Thematic Modelling of Subtext

Charlie Hargood · David E. Millard · Mark J. Weal

Abstract Narratives form a key component of multimedia knowledge representation on the Web. However, many existing multimedia narrative systems either ignore the narrative qualities of any media, or focus on the literal depicted content ignoring any subtext. Ignoring narrative subtext can lead to erroneous search results, or automatically remixed content that lacks cohesion. We suggest that subtext can be computationally modeled in terms of Tomashevsky's hierarchy of themes and motifs. These elements can then be used in a semiotic term expansion algorithm, incorporating knowledge of subtext into search and subsequent narrative generation. We present two experimental applications of this technique. In the first, we use our thematic model in the automatic construction of photo montages from Flickr, comparing it to more traditional term expansion based on co-occurrence, and showing that this improves the perceived relevance of images within the montage. In the second, we use the thematic model in order to automatically identify Flickr images to illustrate short stories, where it dampened the perception of unwanted themes (an effect we describe as reducing thematic noise). Our work is among the first in this space, and shows that thematic subtext can be tackled computationally.

Charlie Hargood
Bournemouth University
E-mail: chargood@bournemouth.ac.uk

David E. Millard
University of Southampton
E-mail: dem@ecs.soton.ac.uk

Mark J. Weal
University of Southampton
E-mail: mjw@ecs.soton.ac.uk

1 Introduction

This work aims to solve the problem of a lack of thematic modelling in multimedia narrative systems by presenting an approach to term expansion powered by semiotics and informed by the literary theory of themes.

Narratives are central to the way that people communicate, from the brief conversational accounts that are exchanged everyday, to the historical and mythological stories that underpin our cultures [33]. However, while we have embraced digital technology in order to record and exchange our narratives - for example, via digital and social media - the narrative structures themselves are opaque to our machines, and our strategies for searching and managing content are therefore unable to take advantage of them.

Several projects have built machine-processable models of narrative. Drammar [32] has been used to create digital annotations of narrative that aid tools in the analysis and research of narrative media. Similarly the ArtEquAKT system [50] automatically generated artist's biographies on request, populating an adaptive story template from an ontology, and using a combination of crawled content and generated sentences to create the final text. However, the existing work in this area focuses on the primary structure, typically the plot (the order in which information is presented) and explicit content (the people, objects, and places that appear in the media), whereas secondary structures such as subtext are ignored. Subtext is the underlying meaning or ideas that the author of a piece of text or media wished to communicate to the reader/viewer. While research has been conducted into the emotional response to online multimedia [41], to our knowledge there are very few examples of *Themes* within media being explicitly modeled. Those systems that do address *Themes* [11] use the term to mean a form of classification that differs from the conventional narratology meaning [48], and other subtext modelling systems approach what could be considered smaller scale linguistic subtext such as sentiment or sarcasm rather than the broad narrative context of theme [22]. Addressing this gap in machine understandable models of narrative and multimedia analysis is the primary motivation of this work.

In this article we argue that *Themes* are one way in which this can be done. We have been inspired by Tomashevsky's work on themes and subtext [48] and our approach is based on a thematic model of themes and motifs that can be used to drive a semiotic term expansion during the search for narrative content [17].

We focus on two applications to demonstrate the potential of this model - the generation of themed photo narratives, and the automatic illustration of short stories. Both are search tasks, and in both cases the thematic model is used to retrieve images that the system considers to be thematically appropriate.

In the first task the images are assembled together into a visual collection around a particular theme, this could be considered a machine generated montage, or a resource for later manual editing (similar to the video aggregation and assembly work done by Kaiser [24]). In the second task the thematic model is used to find illustrations for a short story, with the aim of producing a more coherent overall narrative - something identified as a challenge for narrative generation [43].

In both of these cases the initial task of finding and describing resources is key. Assembling a set of resources that is appropriate to the original search (or seed terms) is critical if the final presentations are to be sensible and are to feel coherent. This means that the quality of the set *as a whole* is important, as the

resources are used and experienced together. The problem of diversity in search results is similar in that it considers the quality of the entire result set [2], but we actually have the opposite goal, in that rather than trying to return a varied set of results we need to return a set that is cohesive. This has been tackled in terms of coherence in time [26] and coherence in space [1], but we are interested with thematic relevance as a way of achieving coherence at the level of subtext. For example, Storyscope is an ontology-driven web-based environment for exploring narratives around museum collections, which uses setting and theme to select items that are relevant to the current storyline [51]. This is directly analogous to our illustration example but with a closed and annotated set of media items.

Our work therefore addresses the challenge of retrieving thematically coherent search results from open content for re-use in a narrative. In previous work we have described the model, and shown in an initial experiment that semiotic search (bases on a thematic model) is more effective than straightforward keyword search at producing a set of results that is thematically consistent [17]. In this article we build on this work, and evaluate its effectiveness in these two thematic coherence tasks.

In doing so we address three research questions:

1. Will using semiotic term expansion based on a thematic model produce image montages that are more thematically consistent than term expansion based on co-occurrence?
2. If we use these result sets for automatic illustration will it improve the perceived thematic coherence of a short story?
3. Does improving the thematic coherence of a short story also improve the perception of other coherence factors (such as logical coherence)?

This article is structured as follows:

Section 2 describes the theoretical background to our work, and how principles from structuralism, semiotics and narratology have been applied in information systems.

Section 3 sets out the underpinning thematic model, and how the theory of thematics and semiotics have been applied in the creation of our computational model.

Section 4 describes our first experiment to show whether using the model for term expansion can generate more relevant results for image montages than term expansion based on co-occurrence. This extends the work presented in [17] where we compared our approach to normal search in Flickr.

Section 5 then describes a second experiment to explore whether using semiotic term expansion in the automatic illustration of a short story improves the overall thematic cohesion of that story, and whether that subsequently has an impact on how cohesive that story is perceived to be in other ways. Finally in Section 6 we summarise our findings, and discuss potential avenues for future work.

2 Background

Our approach draws on structuralist work within narratology. Narratology refers to the theory of narrative that arises from literary theory, criticism, and philosophy. Structuralism is a philosophy concerned with identifying structures emergent

through language. It has been applied to many areas including literary theory and semiotics, for example in work by Barthes [6]. Our work is based on a structuralist analysis of narratives which assumes the existence of patterns and re-occurring forms. This is useful as it provides a framework in which we can work with defined entities and relationships.

2.1 Narratology and Structuralism

Structuralism has been criticised for its rigidity [46], and its critics observe that narratives do not always conform to a given explicit structure. Consequently it was philosophically followed by post-structuralism which favoured a less determinate theory of language. However, from the perspective of this research (which requires machine readable structures that are necessary simplifications of some richer reality) the discrete rules, elements, and relationships that structuralism offers are useful when beginning to build machine readable models of narrative. We acknowledge that not all narratives may adhere to structuralism's models, but also recognise the value in these models for identifying and creating structures within multimedia.

Unsurprisingly ideas of what comprise the elements within a narrative differ considerably. A classic distinction is between *what* is told in the narrative and *how* it is told, these were identified respectively by Russian Formalists as the '*Fabula*' and the '*Sjuzhet*'. This was adapted by French structuralists, particularly Roland Barthes [6], as '*Histoire*' and '*Discours*', which in turn is widely interpreted in English structuralism as '*Story*' and '*Discourse*', and later by others as '*Fabula*' and '*Discourse*'. The overloaded terminology here is confusing, but the essential important lesson is that a narrative may be modeled as a selection process where a wider corpus of candidate narrative elements (a '*Fabula*') has limited selections made from it which are structured together into a narrative (the '*Discourse*').

How story becomes discourse through the process of both authorship and consumption has been explored in literary theory through the notion of plot selection. As demonstrated in the Barthesian model of narrative, the conventional view is that the author selects story elements from the Fabula to be a part of the Sjuzhet. This concept was further explored by Musarra-Schroeder, based on Calvino's writings [40] as 'The Garbage Axiom' representing the process of the author deliberately omitting potential story elements.

Our own narrative system follows a similar process of '*Fabula*' and '*Discourse*' - as explained later in this article we build a corpus of images on a given topic (our '*Fabula*') and then select from this based on the rules of our model a montage of images (our '*Discourse*').

This selection of items to become part of a narrative is an important component of computational narrative systems. These systems include a diverse and sophisticated computational exploration of plot, but their structures are largely limited to the literal content of characters, actions, and settings. Little attention is made to the subtler notion of subtext. This can leave resulting stories lacking in cohesion or thematic depth. In our work we seek to go beyond the literal selection of content for a discourse, and to do that we need to model what those selections might mean to a reader. For this we turn to the field of Semiotics.

2.2 Semiotics, Thematics, and Subtext

Semiotics is the study of signs and how we extract meaning from them. Sassaure wrote that all signs are made up of a signifier and a signified; something we are observing and our understanding of it [44]. This literal interpretation is that of *denotation*; we see a specific football and to us this denotes the concept of a ball. Barthes expanded on this by describing the idea of *connotation*, that signs have a meaning beyond their literal expression, he wrote that the entire denotative sign becomes a signifier for a further signified; for example, we may connote from the ball the concept of competition [6].

Conceptually this divides what might originally have been thought of as a single part of the narrative into two things: what the audience sees (the literal denotation), and what the audience understands (the connotation inferred from what they are presented). Contemporary structuralists have used this notion of connotation to begin to model the underlying meaning of a text, and we have used the same principle in our own work to begin to model subtext in terms of themes.

Thematics can be described as a structuralist approach to the concept of themes within narratives [48]. Tomashevsky deconstructs 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'). 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 be deconstructed into other themes or motifs whereas a motif may not be deconstructed. This builds a hierarchy with specific denoted motifs at the bottom and a tree structure of connoted themes above. Tomashevsky believed that themes were at the root of giving a narrative meaning and cohesion. Through themes an author can give a story purpose by presenting a coherent perspective rather than merely a report of events.

Computational work on themes seldom follows this narratological definition, and is often more simplistic. For example, Bischoff [7] looks at extracting themes from multimedia (music in this case) and tries to support thematic tagging of work. However Bischoff use of the word 'theme' refers more to its usage in media (such as 'traveling music') than its semiotic subtext. Similarly Joke-o-mat [11] presents successful work in the thematic tagging of sitcoms, however they have used the word theme to describe a type of scene section (such as 'dialogue' or 'punchline') rather than what narrative theorists such as Tomashevsky would have considered the thematic subtext. Harrell's use of 'thematic domains' [19] is closer to what we propose, but lacks any semiotic structure, while each domain represents a conceptual definition of the theme they are simply a collection of associated terms. Our model attempts to go beyond this, by including the denotation and connotation relationships between themes and motifs.

Computational work on subtext with a definition broader than just themes can also be found, typically this work is concerned with the sentiment of text. In the same way that our work seeks to cover narrative subtext rather than explicit narrative content such as plot, this work seeks to uncover the subtextual meaning of text (such as sentiment) rather than just its direct message. Examples of recent work in this space has been exploring the detection of sarcasm in text through a variety of approaches including rule based methods, NLP feature detection, learning and deep learning algorithms, and shared tasks approaches as

detailed in Joshi et al.'s recent survey of advances in the area [22]. Sarcasm is undoubtedly a form of subtext (and of significant importance to sentiment analysis where deceptive language can entirely reverse a sentiment) and there are examples of tag and metadata based approaches, such as that by Maynard et al. [36], which is similar to our own approach (in its reliance on metadata). However, where our work differ is the diversity and variety of intended message with thematic subtext (as opposed to the more discrete "is/isn't sarcastic" subtext of sarcasm), along with the specific type of subtext being addressed.

2.3 Term Expansion

The machine based expansion of connections between terms and concepts is something that is more typically known in the information retrieval field as *term expansion*. The idea is that by expanding the terms in user's queries or the candidate terms against which they are being matched a greater number of successful matches may be found. There are a variety of methods that can be used to achieve such expansion by assessing different relationships between a term and other terms.

2.3.1 Lexical Systems

Perhaps the most straightforward method of term expansion is to use a thesaurus, expanding a term using synonyms and other similar words. WordNet [39], a large general purpose thesaurus developed by Princeton University, provides a good basis for a system undertaking such an expansion with a large variety of terms and many different kinds of lexical relationships drawn between them. Voorhees conducted an initial investigation [49] on the generic effectiveness of lexical query expansion using WordNet as a basis for different lexical relationships and using the TREC collections[1] as test search data. However Voorhees' work shows that there is little advantage to such expansion, finding only minimal improvement on very small queries and no improvement on larger ones.

2.3.2 Co-Occurrence

Co-occurrence is a statistical method involving the analysis of the semantic similarity of two terms based on the frequency with which they occur together in a document. Co-occurrence can be used in automatic keyword extraction, such as in Matsuo's work [35] but can also drive query expansion as described by Kubek [25]. In such systems a corpus of potential results is analysed and terms attached to documents in the corpus that co-occur frequently with the terms used in the query are used to expand it. This method of expansion is automatic and has returned impressive results, and as Li's recent review of tag based image retrieval shows [27] co-occurrence continues to be regularly used in a range of systems as a measure of term similarity.

Co-occurrence appears to be an effective method for term expansion in improving relevance of queries. However it is a solely statistical basis for inferring

---

[1] A benchmark collection of documents, standard textual queries, and value judgements of relevance established by the TREC conference. Can be found at http://trec.nist.gov/data.html as of 12/2/18

what a users intentions were when using a term rather than based on any semantic understanding. As such it is vulnerable to query drift (expanding the terms in inappropriate ways) and its effectiveness is highly dependent on the quality of the corpus used to train it.

2.3.3 Ontological Approaches

One approach to solving the problem of query drift is to use models of expert knowledge as a basis for expansion for queries such as the work done by Fu [12], which uses an ontology to expand and improve geographical queries similar in objective to the co-occurrence work done by Buscaldi [10]. This tends to be most effective within a specific domain as ontologies are normally created for specific fields by a small group of experts, fully exploring a small group of concepts. For example, ontologies such as the Gene Ontology [4] are used to expand terms used in queries relevant to their subject or glean further meaning from terms used in media related to their subject. Ontological solutions for narrative media retrieval and composition have been used by the multimedia research community before, such as in more recent work by Kaiser in video assembly [24] and demonstrated persuasive results, if at the cost of the construction of detailed domain specific ontologies.

Our work is similar to this ontological approach in that it is term expansion based on an expert model, however in our case the model is a thematic one, and the relationships between concepts are all semiotic. However, as a thematic model, its rules are generalised and not tied to one specific domain or narrative type - rather they can be used in a range of instances. It does require the construction of instances of the model, but these are not limited to a single domain, as is the case with some other ontological approaches.

2.4 Multimedia Feature Extraction and Processing

In this particular work we are utilizing term expansion as a form of feature extraction, a common focus of multimedia analysis where features of a piece of multimedia are computationally inferred often to form assertions over the content. Typically this involves analysis of data on the content either extracted from the content itself or metadata included alongside the item. The former is often achieved through a hardware sensor such as in work on activity recognition by Liu et al. [30], or direct multimedia processing such as natural language processing as seen in the work by Preoiuc-Pietro et al. [42] on ideology analysis on social media. These techniques are not limited to text and feature extraction through image processing is common, including a variety of learning algorithms as demonstrated by Li et al.'s work [28], as is use of neural networks to classify varied media as seen in work by Shu et al. [45]. The alternative form of feature extraction does not use the direct multimedia itself but rather processing of metadata on the content already included such as tags. This includes applications seeking to refine metadata by adding or removing erroneous tags as seen in work by Tang et al. [47] and Li et al. [29], or tag and keyword processing as seen in Liu et al.'s [31] work on career trajectory analysis using occupation keyword analysis, or Kaiser's work [24] on multimedia aggregation through metadata.

Our own work is more in the later field as we use meta data as the basis of term expansion in order to infer the thematic features of images. However while existing approaches are often stochastic in nature, trained from co-occurrence or other observed associations in a large data set, ours is powered via semiotic relationships based on a thematic model which is itself based on fundamental literary theory and human captured denotations and connotations.

3 Thematic Model

In our work we assume a situation where a multimedia story is compiled from many small segments of content 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 which, depending on the granularity of the system, might be a single photo or paragraph, a sentence, or a fragment of an image. These are similar in definition and use to the 'Narrative Units' identified by the Drammar ontology [32] in that they are flexible, but effectively a single irreducible piece of media.

The content of these natoms is rich with information, however only some of it may be visible to a machine (such as generated meta data or authored tags on images.) We call these visible computable elements *Features*. Features might take any number of forms, in our work we commonly use tags but they might also be automatically detected through some computational analysis as mentioned previously in our discussion on feature extraction and processing. Features can each denote a *Motif*, a basic thematic object that has connotations within the story, for example the tag *cake* is a feature that 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 *celebration*. These themes, when combined with other themes or motifs could in turn connote broader themes, for example *wedding* might also connote *celebration*.

The model, shown in Figure 1, shows how the parts of the model map to Barthes' ideas of denotative signs as the signifiers for connotative signs. Features denote Motifs with themes being broader concepts communicated over the entirety of the narrative, typically by numerous motifs.

A set of rules augment the core components of the model (Natoms, Features, Motifs and Themes) with *Justifications*. When a connotation relationship is formed between a *motif* and a *theme* (or between sub-theme and theme), a justification for the connotation is also added explaining why one connotes the other; we added these rules to aid authorship, as no two themes should be connoted by motifs or themes with the same justification (we discuss authorship in more detail in Section 3.2 below. Justifications help the author consider the role of potential elements in connoting a theme and help them consolidate the wide variety of relevant features into motifs formed around the key roles.

In plain text these rules can be articulated as:

1. An element may be either a theme or a motif, not both, and all themes and motifs are considered elements.
2. A feature is not an element, nor can an element be considered a feature.
3. A denote relationship is always between a feature and a motif, and all motifs must be denoted by at least one feature.
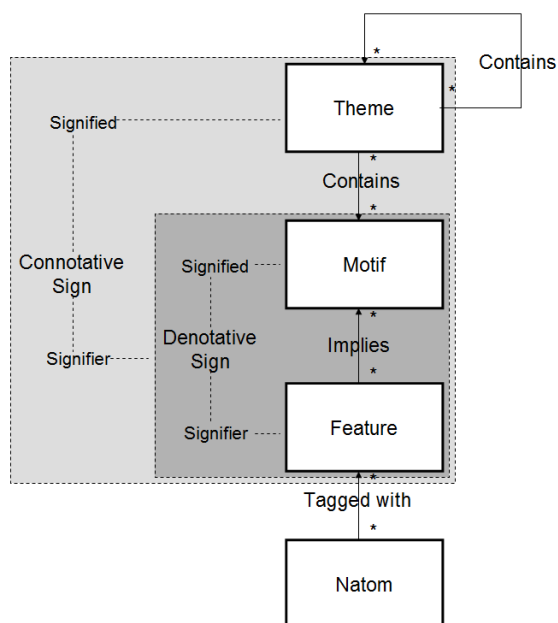
Fig. 1 The Thematic Model

4. A connote relationship is always between an element and a theme, and all themes must be connoted by at least one element.
5. All connote relationships must include a justification.
6. No two connote relationships may exist with the same theme and justification.

This forms the basis of our computational model of themes for narrative. Prior existing models in multimedia research such as Drammar [32], the recent work on transmedia by Jung [23], the video mash-up domain models by Kaiser [24], or the broad narrative ontology presented in OntoMedia [21] have all shown the advantages of machine readable models of narrative in search, media aggregation, annotation, navigation, and generation. However, prior models have nearly entirely focused on the literal content and plot of narrative and not its subtext - as ours does. As with Drammar [32] our model represents another instance of narrative theory realised as a computational model, in this case the theory of thematics. While there are other multimedia approaches to both feature extraction and subtext analysis, such as the work on sarcasm detection [22], these approaches do not address theme. Or, in the limited cases where they do, they address theme as genre or usage [7,11], or lack a semiotic structure [19].

3.1 Example

Figure 2 shows a simple example of how a collection of natoms connotes a theme in the terms of the model, in this case a passage of text, and two photographs

that could be interpreted as connoting the theme of winter. The features presented are present within the given natoms, it is feasible that the natoms would be tagged with them or that they might be automatically extracted from them. These features literally denote the motifs of snow, cold, and warm clothing. As snow demonstrates many different features might denote the device of snow but in this case thematically they serve the same effect. Finally in the context of each other these motifs connote the concept and theme of winter.
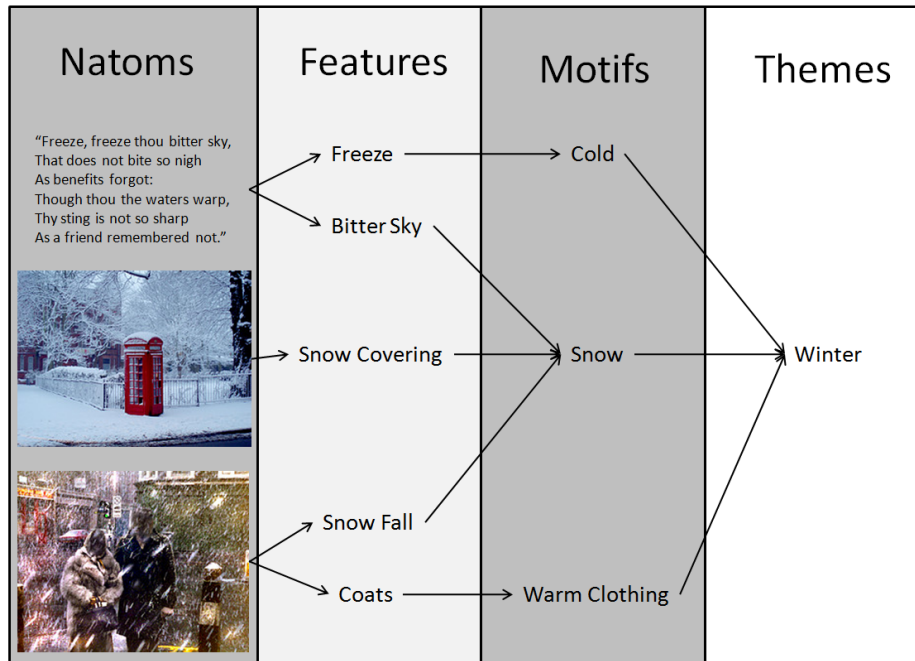

Fig. 2 A Worked Example

3.2 Authoring Method

In order for our approach to be practical it was necessary to have a systemic way for people to create valid instances of our thematic model. We deconstructed our own process for creating definitions and identified five stages for defining a given theme in the terms of the model:

1. *List Associated words*: The contributor spends some time expanding the seed theme into a list of associated words to get a list of related concepts.
2. *Classify as Themes or Motifs*: The contributor then makes two lists using the results of stage 1 based on the rules of the model classifying each as either a theme or a motif.
3. *Group elements*: The contributor groups together similar elements or those that share a similar purpose into a single element based around the shared purpose or a generalisation of the features they share.

4. *Expand Sub-Themes*: The contributor takes remaining theme elements and expands them as they have done the initial theme. Care is taken to consider stage 5 when doing this in order to save time.
5. *Remove associated elements*: The contributor removes each theme or motif that is not entirely relevant to the root theme.

This authoring process was refined into a guide, and has been described in depth and evaluated with users in our previous work [16]. The process is expensive in that it requires human authoring of definitions, however a majority of untrained users did create valid definitions, demonstrating that the method can be successfully used. A key area of future work will be how to better support the creation of thematic models, for example via a richer authoring tool, crowd-sourcing, collective intelligence, or part-automation of the process. The experiments described later in this article use valid thematic definitions created using this process by independent English undergraduate students at the University of Southampton, and later transcribed into XML for use in our systems by the developer.

4 First Task: Thematic Montages

To demonstrate the effectiveness of the narrative model in helping structure similar information our first experiment was devised to use the model in support of a retrieval and composition task for multimedia on the Web.

4.1 The Thematic Engine

The photo sharing system Flickr was used as a source of content (potential Natoms) due to the large amount of readily available tags (Features) that accompany the images. Tag folksonomies such as that made available by Flickr have been demonstrated to offer meta data on items of a higher semantic value as opposed to collections with automatically generated data [3].

The theme definitions were written in XML, with each file representing a thematic element (either a theme or a motif). Definitions for themes listed the motifs with which they shared a connotation relationship and definitions for motifs listed the features that denoted them. For this first experiment, four root themes were authored by hand following the defined authoring method described in Section 3.2. The themes selected for the initial experiment were *Winter*, *Spring*, *Family*, and *Celebration*.

The Thematic Engine generates montages by taking a desired montage size (number of images), a desired content (keyword subject), and a desired list of themes (comma separated list of keywords). The Thematic Engine searches Flickr for the desired content and forms a base corpus (in narrative terms a Fabula) using the top 30,000 images returned by the keyword search. The thematic quality of each image (its relevance to the requested themes) is then calculated and the top N images are returned where N is equal to the desired montage length.

The thematic quality of each image is calculated based on the features present. Each tag is considered to be a feature and using this, each image's *component coverage* and *thematic coverage* is calculated. How these are calculated and how
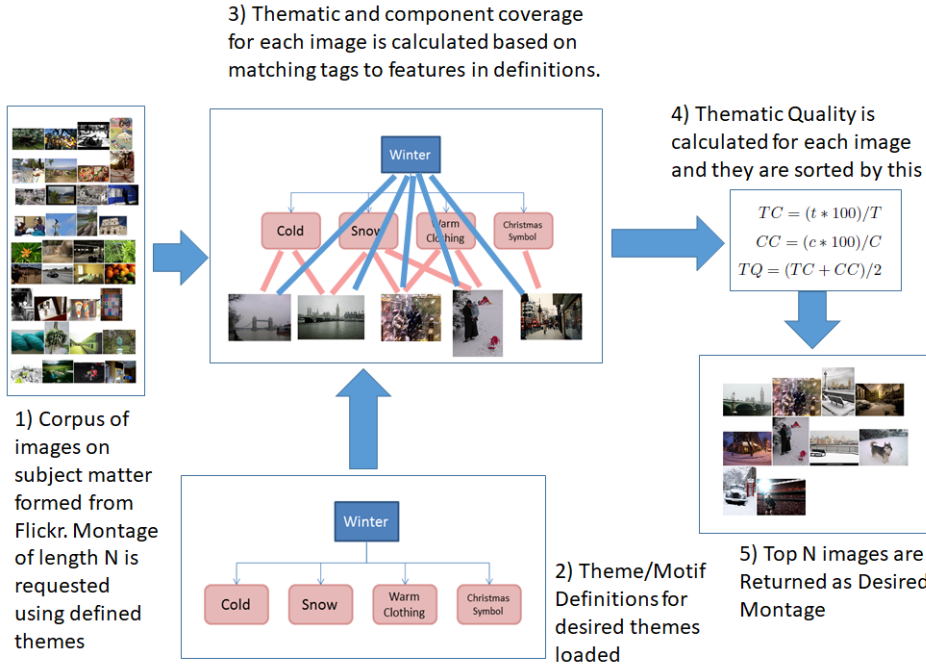
Fig. 3 The Process by which the TMB generates a montage

thematic quality is calculated from them is presented in equations 1, 2, and 3 below. TQ is thematic quality, TC is thematic coverage, CC is component coverage, T is the number of desired themes, C is the sum number of components (elements, themes or motifs, that directly connote a theme) of all desired themes, and t and c are the number of themes or components respectively for which the image has a relevant feature. A feature is considered relevant if it directly denotes a motif that is either a component or through a chain of connotation later indirectly connotes the component or theme requested.

$$TC = (t * 100)/T \tag{1}$$

$$CC = (c * 100)/C \tag{2}$$

$$TQ = (TC + CC)/2 \tag{3}$$

The final thematic quality is therefore expressed as a percentage and is based on how many of the desired themes the image is to connote to as well as how relevant it is to each theme's top level thematic components. The entire process is depicted diagrammatically in figure 3.

Initially we tested the effectiveness of the Thematic Engine as compared to a simple keyword search [15][17]. As the Thematic Engine is based in part on Flickr we elected to compare it to Flickr's keyword search. As well as comparing the thematic relevance of both approaches for individual images we were keen to see how well the thematic system performed in a more narrative context of many 'natoms'; in this case a photo montage. To summarise, our experiment showed

us with statistical significance that the the inclusion of themes produced images perceived to be more relevant then the Flickr keyword search, especially when images were presented in groups.

Having demonstrated that semiotic term expansion was effective we need to evaluate how it compares to existing term expansion methods, and how well it functions when used within a narrative context.

4.2 Comparison with Co-Occurrence

Our initial work demonstrated that semiotic term expansion is effective, but it is necessary to investigate the quality of that expansion as compared to existing techniques of term expansion.

Mandala's original review of a range of term expansion methods for query expansion [34] showed the strongest individual approach was co-occurrence, a method of term expansion that continues to be used as an effective means of measuring term similarity today in multimedia retrieval [27]. As such we identified co-occurrence term expansion as a suitable candidate for comparison.

In order to keep the comparison fair, the co-occurrence system would operate with the same rules as the Thematic Montage Builder (TMB) which used the Thematic Engine described above. A corpus on the subject of the montage would be compiled and the system would then expand the term representing the desired theme to identify the objects in the corpus with the highest thematic quality. The top N of these images, where N is the desired size of the montage, would then be returned as the montage.

The system rates the semantic similarity of two terms within the corpus based on how frequently they occur, and co-occur. For this system if the terms co-occurred as tags for a particular image in flickr this was recorded as a co-occurrence. Based on these two frequencies the semantic similarity of the two terms may be calculated in a number of different ways, we use the 'Mutual Information' measure as our similarity calculation which (while very similar to other similarity measures) has been shown to be slightly more effective [34].

Using these calculations the system can create a vector for a pseudo document (a model representing a theoretical ideal document with tags proportional to their similarity to the desired term). This is based on the semantic similarity of every term used as a tag in the corpus to the term for the desired theme, where each term is a dimension. The thematic quality of each image is then calculated as the Euclidean distance of a vector describing the image (where the frequency of each term comprises its distance along that dimension) from the vector describing the pseudo document. In the case where multiple themes are used the half-way point between the pseudo document for each theme is used. Also, when detecting the presence of a term, basic stemming is used so that plurals and other minor variations of the same term are all still detected.

This created a Co-Occurrence montage generator similar to the TMB in that a desired theme and content could be specified along with montage and corpus size and a montage would be returned that contained images relevant to the desired content that were also thematically relevant to the desired theme. The difference being one was using the semiotic expansions in the form of the thematic definitions and the other performing an automatic expansion based on co-occurrence.

Both of these applications were of O(n log n) complexity, the original scoring and co-occurrence detection being O(n) and the merge sort to order being O(n log n), and could not be used in real-time. It is possible however that the technical implementation of these algorithms might be improved, however as our contribution focuses instead on the relevance of images selected and not the efficiency this implementation is suitable for our needs.

4.2.1 Methodology

We ran an experiment to compare the performance of the TMB and the Co-Occurrence generator. The experiment displays images to participants under a title composed of both a content keyword and theme(s) such as London in Winter (images about London with the theme of winter). Both systems generate ten image montages for each title and participants view the images both individually and grouped together as a montage and rate their relevance to the titles. The experiment itself is divided into four tests; two tests for titles with a single theme, and two for titles including multiple themes to test to performance of the systems in both situations. For both sets of titles the first test displays the images individually at random under the title they were generated for and the users are asked to rate their relevance to the title from 1 to 5. The second test for each set of titles groups the images together in their montages, once again under the titles, and asks the participant to rate the relevance of the images as a group. Two base cases are used to give the results context, a low base case (BaseL) of ten randomly selected images which are taken from the most recent images uploaded to Flickr, and a high base case (BaseH) of ten images selected by a person compiling the best montage they can for the given titles from images in Flickr.

The titles were chosen to explore how the systems performed with titles including both single and multiple themes as well as titles with themes that complimented the content of the corpus or fabula as well as ones that clashed[2] with it. As such four single theme titles were used; two regular theme fabula pairings and two clashing theme fabula pairings, as well as two titles with multiple themes. In the tests requiring single theme titles, users were given one regular paired title and one clashing one alternating to the other two titles for the next participant. The titles were:

– Title 1: London in *Winter*
– Title 2: Earthquake and *Celebration*
– Title 3: *Family* Factory
– Title 4: *Spring* Picnic
– Title 5: *Family* in New York at *Winter*
– Title 6: *Celebration* of New House in *Spring*

We enforced a rule that no montage would contain more than one image by the same Flickr user as images uploaded as part of a set by a single user would often have strong inherent commonality. All montages were generated in the same afternoon to ensure they were using as similar a state of Flickr as possible. When the images were presented individually they were randomised so as to prevent the identification of which images belonged together in montages.

---

[2] by 'clashed' we mean themes that would not normally be associated with the subject matter - such as the theme of 'Celebration' with a corpus of images on 'Earthquakes'

4.2.2 Results

Recruitment to the experiment was through social media sites and received a total of 57 participants. Our findings were that the thematic system outperformed the co-occurrence based system both in individual images and with montages. Table 1 shows the frequency data and statistics for single images, and table 2 shows the same data but for the images grouped as montages (5=highly relevant, 1=not relevant). It is to be noted that in some cases a participant skipped or missed rating an image or montage and consequently the total frequencies are not identical (though are similar). The hypothesis that the TMB selects images rated more relevant for the given titles then the co-occurrence based system is true with a 0.0005 probability of error both for individual images and montages.

Table 1 Single image rating counts and statistics of TMB and Co-occurrence experiment

Set	1	2	3	4	5	Total	Set	Mean	SD	Variance
TMB	736	407	437	332	302	2214	TMB	2.57	1.42	2.03
Co-occur	989	438	380	232	162	2201	Co-occur	2.16	1.30	1.68
BaseL	1809	253	100	31	12	2205	BaseL	1.27	0.66	0.44
BaseH	313	430	565	451	459	2218	BaseH	3.14	1.33	1.77

t=10.221, df=4413, p=0.0005

Table 2 Grouped Images Rating Count and Statistics of TMB and Co-Occurrence Experiment

Set	1	2	3	4	5	Total	Set	Mean	SD	Variance
TMB	5	39	74	76	30	224	TMB	3.39	1.00	0.99
Co-occur	32	75	64	40	14	225	Co-occur	2.68	1.11	1.24
BaseL	180	38	4	0	1	223	BaseL	1.22	0.52	0.27
BaseH	0	7	32	77	106	222	BaseH	4.27	0.82	0.68

t=7.053, df=447, p=0.0005

While this improvement might seem slight it is important to view it in the context of both base cases. Figure 4 shows the mean relevance ratings of the four different methods of selecting the images. Standard error was calculated but is too small to display on these graphs. Both graphs show the thematic system outperforming the co-occurrence system. The margin of improvement, which at first might seem small, is more impressive considering the margin between entirely random images and images purposefully selected to make the best montage possible.

We note that images are rated higher when presented as a montage (with the exception of BaseL). As shown in table 3, the average improvement in relevance rating from rating given as single image to rating given as a montage however is higher for images selected by the TMB then those selected by the co-occurrence based system. The hypothesis that the TMB experiences a stronger improvement from individual images to grouped images is true with a less than 0.0005 probability of error.

We also recorded how both systems performed for titles that contained a single theme as well as those with multiple themes. This is shown in table 4. We recorded
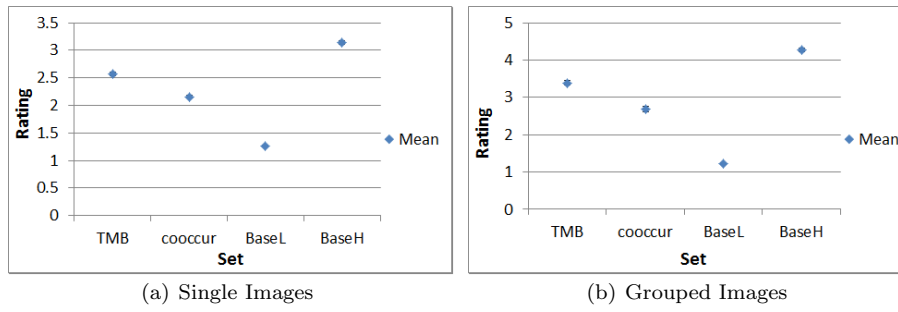
(a) Single Images   (b) Grouped Images

Fig. 4 Single and Grouped Image Rating Mean of TMB and Co-Occurrence Experiment

Table 3 Grouped Images Improvement Statistics of TMB and Co-occurrence Experiment

Set	Mean	SD	Variance
TMB	0.81	1.46	2.12
Co-occur	0.53	1.48	2.20

t=6.297, df=4413, p=0.0005

how each system performed for titles with a clashing theme keyword pairing as well as those with a regular pairing. This is displayed in tables 5.

Table 4 Single and Grouped Images Single/Multiple Themes in Title Contrast Statistics of TMB and Co-Occurrence (CoOc) Experiment

Single				Grouped			
Set	Mean	SD	Var	Set	Mean	SD	Var
TMB Single Theme	2.74	1.52	2.30	TMB Single Theme	3.43	1.00	1.00
CoOc Single Theme	2.56	1.44	1.07	CoOc Single Theme	3.16	1.13	1.28
BaseL Single Theme	1.26	0.67	0.45	BaseL Single Theme	1.15	0.41	0.17
BaseH Single Theme	3.16	1.42	2.01	BaseH Single Theme	4.21	0.87	0.75
TMB Multi Theme	2.40	1.31	1.70	TMB Multi Theme	3.35	0.99	0.99
CoOc Multi Theme	1.75	0.98	0.97	CoOc Multi Theme	2.21	0.86	0.74
BaseL Multi Theme	1.28	0.65	0.43	BaseL Multi Theme	1.30	0.60	0.35
BaseH Multi Theme	3.12	1.24	1.53	BaseH Multi Theme	4.33	0.78	0.60

Table 5 Single and Grouped Images clashing/regular theme keyword pairing in title performance statistics of TMB and Co-occurrence (CoOc) experiment

Single				Grouped			
Set	Mean	SD	Var	Set	Mean	SD	Var
TMB Regular Pair	2.71	1.45	2.09	TMB Regular Pair	3.56	0.96	0.93
CoOc Regular Pair	2.16	1.27	1.62	CoOc Regular Pair	2.60	1.09	1.19
BaseL Regular Pair	1.27	0.66	0.44	BaseL Regular Pair	1.23	0.52	0.28
BaseH Regular Pair	3.37	1.28	1.64	BaseH Regular Pair	4.48	0.71	0.50
TMB Clashing Pair	2.18	1.28	1.64	TMB Clashing Pair	2.88	0.92	0.84
CoOc Clashing Pair	2.15	1.37	1.87	CoOc Clashing Pair	2.93	1.15	1.32
BaseL Clashing Pair	1.27	0.65	0.43	BaseL Clashing Pair	1.20	0.49	0.24
BaseH Clashing Pair	2.5	1.24	1.53	BaseH Clashing Pair	3.61	0.81	0.66

4.2.3 Analysis

Our data shows that semiotic term expansion driven by our thematic models is a more effective means of expanding thematic keywords than co-occurrence. The relevance of TMB images was rated higher for both single and grouped images than the co-occurrence images, and the improvement from single presentation experienced by images presented as a montage was also greater for the TMB, all to a degree that can be considered statistically significant. While the improvement experienced may at first seem slight the standard error on the means shown is very small (0.027 - 0.074) and in the context of the two base cases the improvement is more impressive. The improvement from entirely random images to purposefully selected images by hand is 1.872 for single images and 3.046 for group images, the improvement from co-occurrence to TMB is 0.419 and 0.703 for single and grouped respectively.

Semiotic term expansion also showed it was more capable of selecting images for titles containing multiple themes; this can be attributed to the way thematic score is calculated emphasising images relevant to both themes and looking for common shared motifs. As before the TMBs weakest performance was when it was required to produce montages for titles with a clashing theme fabula pairing in the title, this is to be expected due to the fact that the features representing the specific desired motifs will rarely found within the corpus. However, in this case the co-occurrence system also struggled and performed comparably badly.

The lower performance of the co-occurrence system may be explained by query drift as discussed in [53]. This is to some extent born out by examining the image sets generated by co-occurrence, for example we can see it has drifted from winter to snow to snowdrop (the flower). It has also been noted in work such as that by Xu [52] that the best results from co-occurrence come when it is trained using a local corpus that is known to be relevant to the query being expanded. While we were training using a local corpus it was not specifically relevant to the element we were expanding, for that to be the case the corpus would (as an example) have to be populated with a Flickr search for 'London in Winter' rather than just 'London'. If this is the case it is possible co-occurrence is less effective for expansion of terms for which it is more difficult to acquire a training corpus of ascertained relevance such as a theme.

There is the possibility that the TMB may be particularly well suited to a particular title and was therefore having its average taken higher by an individual case. In order to analyse this a little further table 6 displays the mean rating for each title from both the TMB and the co-occurrence systems for single images whereas table 6 does the same for montaged images. Both tables also show the improvement in relevance made by the TMB (negative numbers representing instances where co-occurrence performed better).

The TMB has scored significantly higher for titles 1 and 5, which were 'London in Winter' and 'Family in New York at Winter'. However, if we remove the mean ratings for both titles including winter entirely we find the TMB still has a higher mean than co-occurrence for both single and montaged images, showing 2.380 for the TMB and 2.267 for co-occurrence for single images and 3.243 for the TMB and 2.992 for co-occurrence for grouped images. It is also still statistically significant, even excluding the winter titles; the TMB performed better than the co-occurrence

Table 6 Mean rating by title for grouped images of TMB and Co-occurrence

	Set	Title 1	Title 2	Title 3	Title 4	Title 5	Title 6
	TMB	3.89	2.25	2.10	2.68	2.37	2.44
Single	Co-occur	2.69	2.56	1.71	3.25	1.60	1.90
	Difference	1.20	-0.31	0.39	-0.57	0.77	0.54
	TMB	4.17	3.00	2.74	3.78	3.35	3.35
Grouped	Co-occur	2.79	3.60	2.19	4.04	1.86	2.55
	Difference	1.38	-0.60	0.56	-0.26	1.49	0.80

system with a t of 2.247 (p=0.01, df=2741) for single images and a t of 1.952 (p=0.05, df=278) for grouped images.

To summarise our findings:

– It is possible to use definitions created in terms of a thematic model to generate simple photo montages relevant to a desired theme.
– A system using thematic definitions creates montages rated more relevant than those offered by either basic keyword search or co-occurrence term expansion.
– The thematic system is still effective in situations demanding multiple themes but less effective if the desired content and theme clash.
– While all systems are more effective at finding themed montages rather than single images, the improvement experienced by the thematic system is greater.

5 Second Task: Illustration and Thematic Cohesion

Our second objective was to assess the impact of the thematic model on the automatic illustration of a short story. In particular, is it better than regular search in terms of thematic cohesion? In order to do this some tangible ways of measuring the cohesion of a narrative must first be established.

5.1 Cohesion Variables

By narrative cohesion we mean the extent to which the various parts of a narrative successfully work together to produce some overall effect in the reader. There are a number of different ways in which a narrative can be considered to be cohesive.

Genre is a common classification of narrative based upon a set of reoccurring features that position a narrative culturally within the context of other narratives. Tomashevsky suggested that the genre of the narrative was what limited the motifs available [48]. The Coh-Metrix project [14] worked towards creating a system for analysing the coherence of texts through several metrics (including latent semantic analysis, term frequency and density, and concept clarity.) The measuring of these metrics however was intrinsically based upon the genre of the narrative, which, they identified as important to coherence [38]. In his work identifying key features of narrative Bruner [9] also highlights the importance of genre to cohesion. Under his discussion on 'Genericness' he explains how genre is a way of 'comprehending narrative.' By conforming to convention the narrative guides the audience to subconsciously fill in gaps in the presentation and make sense of the content.

In work by Booth [8] there is a description of the importance of the concept of narrator in narrative. As the narrator is core to the telling of the story, coherence

in how the narrator is presented is also important to the cohesion of the story itself. McAdams explains from the perspective of modern psychology that people become narrators in order to make sense of a series of events or stories, thus it is the presence of a narrator that leads to coherence in a story [37].

We have already discussed how the logical use of language may affect the coherence of a narrative however there are other linguistic choices made in the telling of a story that might also affect its coherence. Earlier we discussed how structuralists such as Barthes [6] and Bal [5] consider narrative to be comprised of layers, often of story and discourse, where story stands for content and discourse for how the story is told. Features of discourse have already been identified here; themes, genre, narrator, but these cannot be said to completely account for the language choices made in presenting a narrative. The use and style of language can have an effect on its coherence. Style can be said to be a composite of attitude, tone, and mood of a narrative, representing decisions made on the presentation of elements at the discourse level. The stylistic cohesion of a narrative could be said to be in part the extent to which an author sets out and then abides by their own linguistic conventions.

From the literature we have thus identify five key variables for narrative cohesion [18]:

– **Logical Sense:** the connective language used to explain the content of the narrative.
– **Themes:** the concepts communicated implicitly throughout the narrative.
– **Genre:** the conformance to conventions that culturally contextualise the narrative.
– **Narrator:** the presence of a consistent perspective communicating the narrative.
– **Style:** the way narrative elements are presented within the discourse.

Measured appropriately, and considered together, we propose to use these cohesion variables as a basis to understand the level of cohesion within a narrative that has been automatically illustrated.

5.2 The Illustrator Experiment

Having decided upon these metrics for measuring narrative cohesion we can now address our second and third research questions, and look at how illustrations selected by our semiotic term expansion method alter the perceived thematic cohesion of a narrative, and whether this subsequently impacts the perceived cohesion of the narrative as a whole.

5.2.1 Methodology

For this experiment participants filled in a web questionnaire on the perceived narrative cohesion of three short stories with illustrations. The three short stories selected had three different methods of generating illustrations for the stories, thus nine possible combinations, with each user seeing the three stories with illustrations generated from different methods. The illustration method to story pairings were

rotated using the principle of latin squares to get a spread of data for each method on each story.

The stories used in the experiment were divided into logical sections with each section given an illustration. To facilitate this the stories were stored as xml allowing them to be marked up where the different sections began and ended. The xml model for each story stored a content keyword for each section as well as a theme for the whole story. These keywords and themes were used in the selection of the images.

The stories used were selected from Steve Ersinghaus'[3] contributions to the 2009 100 days project where he wrote 100 short stories. This was an ideal resource for the experiment with a large collection of stories with suitably complex themes, strong imagery that lent themselves to illustration, and an author that was happy to engage with the experiment. Fifteen of the stories were reviewed for their suitability for the experiment. The stories that were picked were the ones which logically fell into 3-5 sections (each of which could receive an illustration) and were of an appropriate length for the planned experiment (took less than 10 minutes to read). Also, to ensure the spectrum of naturally occurring coherence in the plot was covered, a story that was distinctly abstract (and arguably authored with deliberately low cohesion) was selected, as well as a story that was more deliberately strongly coherent, and a third that fell somewhere between. The three stories selected were:

– **Story 1 - The Point:** An abstract story about two people meeting.
– **Story 2 - The Night:** A dark story about a boy and unseen terrors with strong visual imagery.
– **Story 3 - Computer Leon:** A more conventional, dialogue based story about competition between computing professionals.

The illustrations for the stories were generated by one of three methods:

– **Method 1 - Content and Theme:** Illustrations were generated based on a content keyword for each section and a theme selected for the story. This was done using the TMB with a corpus based on the content keyword from Flickr and the theme designated for the story.
– **Method 2 - Content only:** Illustrations were generated based on a chosen keyword describing the content in each section. This would be done using a Flickr search for the keyword.
– **Method 3 - Human Selected:** A high base case was created made up of illustrations selected from Flickr by a literature expert after due consideration of the stories, the expert was also the source of both the content and theme keywords for Methods 1 and 2.

A comparison between methods 1 and 2 would show whether thematic cohesion had increased due to the themed images and also whether this had resulted in a change in other cohesion variables. Method 3 on the other hand gave our results context with an intended best case scenario. The expert for method 3 was an English Masters graduate from Cambridge University with a history of involvement in both literary criticism and computer science research communities, and was independent of the research team.

---

[3] http://www.steveersinghaus.com/

In generating the meta data necessary for the experiment attempts were made to be as fair and impartial as possible. Before selecting images for method 3 our expert was asked to identify a keyword to describe the literal content of each defined section of the stories and also to list the themes that they felt were present within each story. They were also asked to identify from their lists of themes for each story which they felt was the strongest theme. The strongest themes went into the story models as the listed theme for each story and the keywords for content identified were entered for the content keyword for each relevant section.

Having completed this the newly identified strongest themes were modelled into definitions for use with the TMB. To keep the definitions of the identified themes impartial three volunteers were asked to follow the thematic definition guide explained in an earlier section to define the themes. During this process an expert in the model was present to collaboratively help in forming these definitions to ensure the models created were valid, creative control of the definitions was left solely to the volunteers and all the themes and motifs comprising the model were identified by them. The stories and their identified themes are displayed in table 7.

Table 7 Identified themes for the cohesion experiment

Story	Strongest Theme	Other Identified Themes
The Point	Relations	Pun, Simile Description, Innuendo, Geometry, Microcosm, Macrocosm
The Night	Transgression	Discovery, Fear, Authority, Experience, The Unknown
Computer Leon	Competition	Business, Reputation, Service, Technology, Ignorance, Marketing, Cheating, Sabotage

Having completed our models of the stories, illustrations were generated for them using the various methods and added to the models. In the case of our own approach this followed the same procedure as dictated in section 4.1 and figure 3. As Flickr is a user-generated collection it is possible that individual images might be incorrectly tagged. While the effect of individual images was reduced in the previous experiments by the large volume of images involved, the number of illustrations viewed in this experiment is much less and as such the effect of a single anomalous image is potentially increased. To reduce the effect of individual images each system selected their top five images instead of one for each illustration and when participants viewed the illustrations a random image from this montage of 5 would be selected to be the actual displayed illustration.

The images selected obeyed similar rules to our previous experiment in that illustrations for a single story may not contain more than one image per Flickr user (as images from the same set may inherently be cohesive). Selected images were reviewed with the intention of removing any potentially offensive images, or images with impractical height to width ratios, however, ultimately no images needed to be removed.

The experiment was advertised through social media and 66 participants took part. Participants were emailed a link to a brief introduction and a glossary of terms to ensure they knew what was meant by terminology such as themes, genre,

narrator, etc. Participants were asked when reading the story to also consider the illustrations. Once they had begun the participants were shown the first story with its illustrations and then asked to answer a short questionnaire (explained below). This process was repeated for all three stories.

The questionnaire was designed to measure the perceived cohesion based on the five variables we had identified as related to narrative cohesion. Each question was answered using a single Likert scale of 1-5 (5 being the very positive response) with the exception of question 2 which asked the users to rate each theme on a list of 23 themes (the entire list of themes identified by the independent expert for all stories). The questions were:

1. How logical was the story? E.g. did the story make causal sense to you?
2. Please rate the strength of the presence of the following themes in the story. E.g. how apparent was it that these themes were present? Were they subtle or overt? (Followed by a list of themes)
3. How strongly do you feel this story fits into an established genre?
4. How strong and consistent was the presence of an identifiable storyteller? E.g. Was the story told from a perspective you could easily identify?
5. Is the style, presentation, and language used to express the story consistent? E.g. is the story throughout presented in the same way or does it frequently change tone?

Stories were displayed in a deliberately plain format on a single page. While this could lead to a long page, navigating can break immersion when evaluating a narrative [13] and as we were measuring cohesion we were keen to avoid this. A screen shot of a narrative displayed through the system can be seen in figure 5.

Fig. 5 A screenshot of 'The Night' as displayed by the system.

5.2.2 Results

The results for different story and method pairings can be found in table 8 and the graph in figure 6. For Logic, Genre, Narrator, and Style the mean of the rating for the relevant question was used, for theme however our question was more complicated and this warranted a more sophisticated scoring system. Thematic cohesion has been divided into three scores; Theme(S) representing the mean score for the strongest theme (as identified by our independent expert) for that story, Theme(I) representing the mean score for all the other included or present themes identified in that story, and Theme(E) representing the mean score for all the themes not identified by our expert for that story.

Table 8 Cohesion ratings for stories by illustration methods

Story	Method	Logic	Theme(S)	Theme(I)	Theme(E)	Genre	Narrator	Style
The Point	TMB	1.96	3.21	2.66	1.66	1.63	2.13	2.67
The Point	Keyword Search	2.00	3.05	2.70	1.89	1.86	2.86	2.86
The Point	Manual Selection	2.14	3.24	2.64	1.77	1.71	2.29	2.48
The Night	TMB	4.62	3.76	4.26	1.66	4.29	3.76	4.38
The Night	Keyword Search	4.21	3.13	4.10	1.79	4.21	3.54	4.33
The Night	Manual Selection	4.33	2.67	3.91	1.79	4.05	4.38	4.57
Computer Leon	TMB	4.80	4.95	4.27	2.14	3.48	4.00	4.29
Computer Leon	Keyword Search	4.52	4.86	4.15	2.19	3.10	3.19	4.14
Computer Leon	Manual Selection	4.58	4.96	4.23	2.25	3.38	3.54	4.13
All	TMB	3.79	3.97	3.73	1.82	3.32	3.30	3.78
All	Keyword Search	3.58	3.68	3.65	1.96	3.18	3.20	3.78
All	Manual Selection	3.69	3.62	3.59	1.94	2.95	3.40	3.72

5.2.3 Analysis

The results lead to some interesting observations. First of all, as might have been expected, the overall cohesion scores of the deliberately selected abstract story 'The Point' were lower than the other two stories (a total average of 2.351, as supposed to 3.702 for 'The Night', and 3.864 for 'Computer Leon'.) The story selected for deliberately high cohesion scored generally higher. This helps supports the general notion that our questionnaire was able to record cohesion scores. However, conclusions based on the different methods for presentation are not straightforward with no method significantly and consistently raising cohesion above other methods.
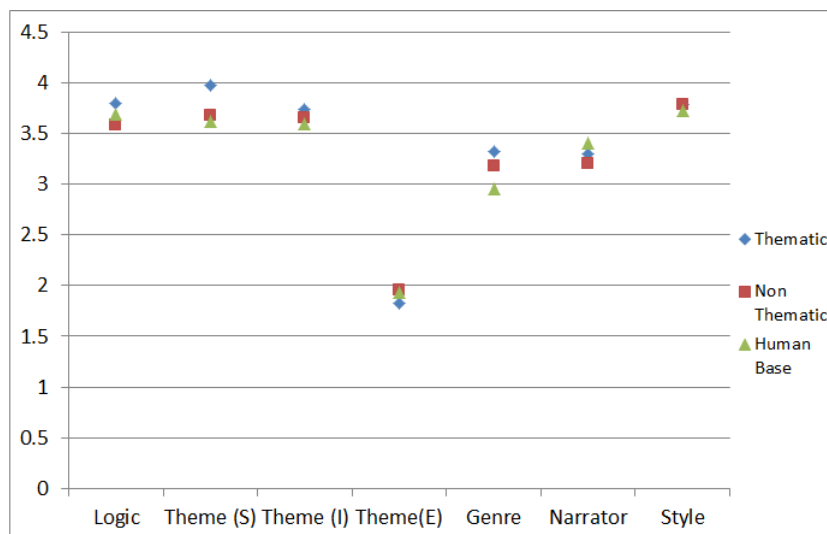
Fig. 6 Cohesion scores for all stories (average)

Our research question was whether thematic illustrations selected by a thematic system improved the perceived thematic cohesion of the narrative. To answer this we need to consider how an improved thematic cohesion would manifest within the scores. As a story becomes more thematically coherent its stronger deliberate themes would be identifiable throughout and false or unintended themes (what we might refer to as 'thematic noise') would become less detectable. As such, in our thematic scores we would expect to see Theme(S) rise and Theme(E) decrease for a successful increase in thematic cohesion.

Analysing the overall data for the range of stories we find that the thematic approach (TMB) has increased Theme(S) and decreased Theme(E) over the generative approach not using themes (Keyword Search). However, when putting this through a t test the hypothesis 'TMB scores Theme(S) higher than Keyword Search' scores a t of 1.181 (df=130, p=0.2) whereas 'TMB scores Theme(E) lower then Keyword Search' scores a t of 2.607 (df=2010, p=0.005) showing that while the decrease in Theme(E) is statistically significant with only a 0.005 probability of error, the increase in Theme(S) is not statistically significant with a 0.2 probability of error. Thus we can conclude that while the images selected by semiotic term expansion have improved thematic cohesion, they have done this only by reducing thematic noise, rather than increasing the presence of a specific theme.

The style of the story may well be a factor in the ability of the Thematic Illustrator to improve thematic cohesion. Our results (as shown in table 8) show that for the thematic approaches, improvement of Theme(S) over the keyword approach is much more substantial for Story 2 ('The Night') than for other stories. Also to be noted is the relatively minor or negative effect on cohesion of thematic emphasis in Story 1 ('The Point'). This could be attributed to the relatively abstract style of story making it difficult to automatically generate relevant or effective illustrations and as such reducing the effect of illustrative emphasis.

To answer our other research question, whether an increase in thematic cohesion leads to an improvement in overall cohesion, we performed a Pearson's correlation between Theme(S) and each of the other non-thematic metrics. The results are presented in table 9. What we find is a moderate correlation with Logic (p = 0.005), and a weak correlation with Genre (p = 0.05). There is also a weak but non-significant correlation with Narrator (p = 0.1), and almost no correlation at all with Style.

Table 9 Pearson's correlation between Theme(S) and other non-thematic metrics

Metric	Correlation Coef	p
Logic	0.30	0.005
Genre	0.19	0.05
Narrator	0.13	0.1
Style	0.06	none

These results suggest that a system capable of improving thematic cohesion could see an improvement in other cohesion variables, in particular Logic and Genre. This would provide a strong argument for pursuing methods of thematic emphasis as it might be used to raise the coherence of generated or adaptive narratives. However further work is needed to establish the ways in which these variables are dependent on each other.

Within this work we have begun to understand how narrative cohesion may be modelled and captured. The experiments contained have also shown that it is potentially viable to alter the coherence of the narrative through thematic emphasis using illustrations. While more work is necessary to build a complete understanding of the effect of thematic emphasis, significant steps have been made here to establish metrics, the effect on thematic cohesion (in particular thematic noise), and the relationship between different variables of cohesion.

6 Conclusions

We began this work by noting that the research on using narrative concepts for information retrieval and the automated generation of content often tends to ignore subtext or at least does not explore narrative themes. We have suggested that a way in which subtext can be explored is by modelling themes based on thematic structuralist theory [48], and have used these thematic models as the basis for a semiotic term expansion.

Our goal has been to see if a search strategy based on this semiotic term expansion will yield more thematically coherent results and lead to better automated remixing of online materials, in particular the automatic construction of photo montages, and illustration of short stories. We outlined three specific research questions:

Question 1: Will using semiotic term expansion compose themed image montages that are more thematically consistent than term expansion based on co-occurrence?

In previous work we had shown that semiotic term expansion works, and is more effective than keyword search [17]. However, the thematic models required

to drive semiotic term expansion are expensive to create and it was therefore important to show how our method compared to more established methods of term-expansion, in particular using co-occurrence. Our first experiment shows that our system using semiotic term expansion outperformed term-expansion based on co-occurrence with statistical significance ($p=0.0005$). While the scale of the improvement is small in objective terms when considered relative to the high and low base cases in our experiment it represents a more sizable improvement. We acknowledge that our conclusions here are limited to our own specific implementations (which we detail) and while co-occurrence remains the basis of many state of the art approaches that minor technical refinements might be made to both implementations. However, our results still demonstrate the value and potential in our approach.

Question 2: If we use these result sets for automatic illustration will it improve the perceived thematic coherence of a short story?

Improving thematic cohesion can be broken down into two parts: improving a chosen theme, and dampening unwanted themes. In our second experiment we have shown that using semiotic term expansion dampened unwanted themes significantly ($p=0.005$), but did not necessarily improve the perceived cohesion of the chosen theme. This may indicate that there is a certain ceiling to what can be achieved in terms of promoting a theme, but does show that thematic noise can be effectively reduced.

Question 3: Does improving the thematic coherence of a short story also improve the perception of other coherence factors (such as logical coherence)?

We have presented a number of coherence factors drawn from the literature and have been able to look at the correlations in the improvement of the different factors to see if making a change in one actually has an impact on the rest. We have shown using Pearson's correlations that improving perceived theme correlates moderately with perceived logical coherence ($r=0.30$, $p=0.005$), and weakly with genre cohesion ($r=0.19$, $p=0.05$). This is evidence that improving thematic cohesion gives readers the perception that the story is more coherent in other ways.

Our research has therefore shown that semiotic term expansion based on a thematic model is effective at making search results more thematically relevant and we believe that it might be utilised in conjunction with other models of narrative to improve narrative generators or other re-mixing systems.

The success of our semiotic term expansion is reliant on the quality of the thematic definitions built for it. Due to the subjective nature of the model this is in turn reliant on human authors. In previous work we have shown that it possible to provide a guide that leads to the creation of effective models, and this does provide some systematic structure to the creation of thematic models. However, more work is needed to explore whether models could be constructed in an automatic way, for example by using clustering techniques to derive coherent terms and concepts from social media streams [20].

Our work is unusual in that it focuses on the subtext and narrative themes, rather than the primary media content or structural elements. Our results show that subtext, in particular thematic subtext, can be successfully manipulated by a machine.

Semiotic approaches provide a way for us to model the underlying meanings and intentions of authors and creators, leading to opportunities for improving both search and automatic content generation. This work represents a contribution to-

wards those goals, but requires further development in how semiotic structures could be created and how they could be applied. The ultimate goal is that systems will begin to understand and utilise the subtler aspects of narrative in as meaningful a way as we do, and that their ability to search, analyse, or generate narratives becomes subsequently more powerful. Future work in this space might seek ways to accelerate the construction of thematic definitions - which is time expensive, explore the application of the model for coherence in other domains, or explore the limitations of the approach in other manners such as with other mediums (e.g. video), a broader selection of stories from a wider selection of genres, or with even larger more varied collections of themes.

References

1. Adams, B., McKenzie, G., Gahegan, M.: Frankenplace: Interactive thematic mapping for ad hoc exploratory search. In: Proceedings of the 24th International Conference on World Wide Web. pp. 12–22. WWW '15, ACM (2015)
2. Agrawal, R., Gollapudi, S., Halverson, A., Ieong, S.: Diversifying Search Results. In: WSDM '09 Proceedings of the Second ACM International Conference on Web Search and Data Mining. pp. 5–14. ACM (2009)
3. Al-Khalifa, H., Davis, H.: Folksonomies versus automatic keyword extraction: An empirical study. IADIS International Journal On Computer Science And Information Systems (IJCSIS) 1, 132–143 (2006)
4. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G.: Gene ontology: tool for the unification of biology. Nature Genetics 25(1), 25–29 (May 2000)
5. Bal, M.: Narratology: Introduction to the Theory of Narrative. University of Toronto Press (January 1998)
6. Barthes, R., Duisit, L.: An introduction to the structural analysis of narrative. New Literary History 6, 237–272 (1975)
7. Bischoff, K., Firan, C.S., Nejdl, W., Paiu, R.: How do you feel about "dancing queen"?: deriving mood & theme annotations from user tags. In: Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries. pp. 285–294. JCDL '09, ACM, New York, USA (2009)
8. Booth, W.: The Rhetoric of Fiction, chap. Types of Narration, pp. 69–74. University of Chicago Press (1974)
9. Bruner, J.: The narrative construction of reality. Critical Inquiry 18, 1–21 (1991)
10. Buscaldi D., Rosso P., S.E.: A wordnet-based query expansion method for geographical information retrieval. In: CLEF 2005 Working Notes (2005)
11. Friedland, G., Gottlieb, L., Janin, A.: Narrative theme navigation for sitcoms supported by fan-generated scripts. Multimedia tools and applications 63(2), 387–406 (2013)
12. Fu, G., Jones, C.B., Abdelmoty, A.I.: Ontology-based spatial query expansion in information retrieval. In: In Lecture Notes in Computer Science, On the Move to Meaningful Internet Systems. vol. 3761, pp. 1466–1482 (2005)
13. Gee, K.: The ergonomics of hypertext narrative: Usability testing as a tool for evaluation and redesign. ACM Journal of Computer Documentation 25(1), 3–16 (2001)
14. Graesser, A., McNamara, D., Louwerse, M., Cai, Z.: Coh-metrix: Analysis of text on cohesion and language. Behavior Research Methods 36, 193–202 (2004)
15. Hargood, C., Millard, D., Weal, M.: Using a thematic model to enrich photo montages. In: Proceedings of the 20th ACM conference on Hypertext and hypermedia. pp. 135–140. ACM (2009)
16. Hargood, C., Millard, D., Weal, M.: Capturing the semiotic relationship between terms. New Rev. Hypermedia Multimedia 16(1-2), 71–84 (2010)
17. Hargood, C., Millard, D., Weal, M.: A semiotic approach for the generation of themed photo narratives. In: HT '10: Proceedings of the 21st ACM conference on Hypertext and hypermedia. pp. 19–28. ACM (2010)

18. Hargood, C., Millard, D., Weal, M.: Measuring narrative cohesion: A five variables approach. In: Narrative and Hypertext at Hypertext '11. ACM (2011)
19. Harrell, D.: Shades of computational evocation and meaning: The griot system and improvisational poetry generation. In: Proceedings, Sixth Digital Arts and Culture Conference. pp. 133–143 (2005)
20. JafariAsbagh, M., Ferrara, E., Varol, O., Menczer, F., Flammini, A.: Clustering memes in social media streams. Social Network Analysis and Mining 4(1), 1–13 (2014)
21. Jewell, M., Lawrence, F., Tufield, M.: Ontomedia: An ontology for the representation of hetrogeneous media. In: Multimedia Information Retrieval Workshop (2005)
22. Joshi, A., Bhattacharyya, P., Carman, M.J.: Automatic sarcasm detection: A survey. ACM Computing Surveys (CSUR) 50(5), 73 (2017)
23. Jung, J.E., Lee, O.J., You, E.S., Nam, M.H.: A computational model of transmedia ecosystem for story-based contents. Multimedia Tools and Applications 76(8), 10371–10388 (2017)
24. Kaiser, R., Hausenblas, M., Umgeher, M.: Metadata-driven interactive web video assembly. Multimedia Tools and Applications 41(3), 437–467 (2009)
25. Kubek, M., Unger, H.: On n-term co-occurrences. Recent Advances in Information and Communication Technology 265, 63–72 (2014)
26. Kuzey, E., Setty, V., Strötgen, J., Weikum, G.: As time goes by: Comprehensive tagging of textual phrases with temporal scopes. In: Proceedings of the 25th International Conference on World Wide Web. pp. 915–925. WWW '16, International World Wide Web Conferences Steering Committee (2016)
27. Li, X., Uricchio, T., Ballan, L., Bertini, M., Snoek, C.G., Bimbo, A.D.: Socializing the semantic gap: A comparative survey on image tag assignment, refinement, and retrieval. ACM Computing Surveys (CSUR) 49(1), 14 (2016)
28. Li, Z., Liu, J., Tang, J., Lu, H.: Robust structured subspace learning for data representation. IEEE transactions on pattern analysis and machine intelligence 37(10), 2085–2098 (2015)
29. Li, Z., Tang, J.: Weakly supervised deep matrix factorization for social image understanding. IEEE Transactions on Image Processing 26(1), 276–288 (2017)
30. Liu, Y., Nie, L., Liu, L., Rosenblum, D.S.: From action to activity: sensor-based activity recognition. Neurocomputing 181, 108–115 (2016)
31. Liu, Y., Zhang, L., Nie, L., Yan, Y., Rosenblum, D.S.: Fortune teller: Predicting your career path. In: AAAI. vol. 2016, pp. 201–207 (2016)
32. Lombardo, V., Damiano, R.: Semantic annotation of narrative media objects. Multimedia Tools and Applications 59(2), 407–439 (2012)
33. Lugmayr, A., Sutinen, E., Suhonen, J., Sedano, C.I., Hlavacs, H., Montero, C.S.: Serious storytelling - a first definition and review. Multimedia Tools and Applications 76(14), 15707–15733 (2017)
34. Mandala, R., M, R., Tokunaga, T., Tanaka, H.: Combining multiple evidence from different types of thesaurus for query expansion (1999)
35. Matsuo, Y., Ishizuka, M.: Keyword extraction from a single document using word co-occurrence statistical information (2003)
36. Maynard, D., Greenwood, M.A.: Who cares about sarcastic tweets? investigating the impact of sarcasm on sentiment analysis. In: Lrec. pp. 4238–4243 (2014)
37. McAdams, D.P.: The problem of narrative coherence. Journal of Constructivist Psychology 19(2), 109–125 (April 2006)
38. McNamara, D., Louwerse, M., Graesser, A.: Coh-metrix: Automated cohesion and coherence scores to predict text readability and facilitate comprehension. Tech. rep., Institute for Intligent Systems, University of Memphis, TN (2002)
39. Miller, G.: Wordnet: An on-line lexical database. International Journal of Lexicography 3(4) (1990)
40. Musarra-Schroeder, U.: Il labirinto e la rete, vol. 519. Bulzoni (1996)
41. Orellana-Rodriguez, C., Diaz-Aviles, E., Nejdl, W.: Mining affective context in short films for emotion-aware recommendation. In: Proceedings of the 26th ACM Conference on Hypertext and Social Media. pp. 185–194. ACM (2015)
42. Preoţiuc-Pietro, D., Liu, Y., Hopkins, D., Ungar, L.: Beyond binary labels: political ideology prediction of twitter users. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). vol. 1, pp. 729–740 (2017)

43. Riedl, M.O., Young, R.M.: From linear story generation to branching story graphs. IEEE Computer Graphics and Applications 26(3), 23–31 (2006)
44. Saussure, F., Bally, C., Sechehaye, A., Riedlinger, A.: Course in General Linguistics. McGraw-Hill (1966)
45. Shu, X., Qi, G.J., Tang, J., Wang, J.: Weakly-shared deep transfer networks for heterogeneous-domain knowledge propagation. In: Proceedings of the 23rd ACM international conference on Multimedia. pp. 35–44. ACM (2015)
46. Sturrock, J.: Structuralism and Since: from Levi Strasuss to Derrida. Oxford University Press (1979)
47. Tang, J., Shu, X., Qi, G.J., Li, Z., Wang, M., Yan, S., Jain, R.: Tri-clustered tensor completion for social-aware image tag refinement. IEEE transactions on pattern analysis and machine intelligence 39(8), 1662–1674 (2017)
48. Tomashevsky, B.: Russian Formalist Criticism: Four Essays, chap. Thematics, pp. 66–68. University of Nebraska Press (1965)
49. Voorhees, E.M.: Query expansion using lexical-semantic relations. In: SIGIR '94: Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval. pp. 61–69. Springer-Verlag (1994)
50. Weal, M., Alani, H., Kim, S., Lewis, P., Millard, D., Sinclair, P., Roure, D.D., Shadbolt, N.: Ontologies as facilitators for repurposing web documents. International Journal of Human-Computer Studies 65, 537–562 (2007)
51. Wolff, A., Mulholland, P., Collins, T.: Storyscope: Using theme and setting to guide story enrichment from external data sources. In: Proceedings of the 24th ACM Conference on Hypertext and Social Media. pp. 79–88. ACM (2013)
52. Xu, J., Croft, W.B.: Query expansion using local and global document analysis. In: SIGIR '96: Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval. pp. 4–11. ACM (1996)
53. Zhou, X.S., Huang, T.S.: Unifying keywords and visual contents in image retrieval. IEEE MultiMedia 9(2), 23–33 (2002)