

Research article

GraphicsAssist: Visual Artist's Assistant for Narration-to-Illustrative Graphic Composition

Suma Dawn  

Dept. of CSE & IT, Jaypee Institute of Information Technology, A-10, Sec-62, Noida-201309, India

Abstract:

Storytelling is an art that very few can carry with ease, putting in proper context the feelings and emotions of the characters, their setting, interactions, etc. This work attempts to recreate storytelling to form a graphic illustration for users so that the essence of the story is not lost in transference from one person to another, from storyteller to graphic illustration reader/user. 'GraphicsAssist' is an artist assistance system for graphic artists to help them generate certain options for usage as graphic frames. Text analysis has been exploited to understand contextual requirements or storytelling. Natural language-based learning has been applied to this requirement. The transfer-learning-based model is used to find suitable matches for the characters and visual environment settings. Illustrative drawings and their tags are considered a part of the frame database. Once the narration entities of the story are extracted, they are matched with the drawing tags and presented to the designer for framing and then panelling of graphic illustration. The chosen frames are displayed in panels in grid format. In its current form, manipulation of the selected drawings for the task of forming frames is, as yet, restricted. The aesthetic transition from frame to frame is yet lacking and may be integrated into later works. The work of this artist's assistance system is validated by the test group, which gave feedback on some pre-defined parameters. Though the system has its drawbacks, it is the first step towards such a system that provides a computational translation model for translating a textual narration to a set of visual composition frames. This work aims to capture the nuances, characters, emotions, story setting, etc., from textual narration as an input of the story and convert it to the graphic illustration in 2D form by incorporating illustrative drawings as its visual translation.

Keywords: Textual Narration; NLP, NER; Image Tagging; Narratology; Feature extraction; Graphic frames; Transfer Learning; Visual Narration



Article History: Received: 12 June 2024. Revised: 28 June 2024. Accepted: 29 June 2024. Published: 30 June 2024.

Copyright: © 2024 by the author/s. License Aesthetix Media Services, India. Distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Citation: Dawn, S. (2024). GraphicsAssist: Visual Artist's Assistant for Narration-to-Illustrative Graphic Composition. *Chitrolekha Journal* 8:1 <https://doi.org/10.21659/cjad.81.v8n109>

I. Introduction

Traditional literature is an incessant storehouse of stories. Indian folklore, with its diverse cultural bends, teachings, and environmental settings, consists of an amaranthine range of short stories and tales. Learning and entertainment are intertwined [1], [2]. While reading a story, the reader visualizes the story narration. This visualization imprinted on her mind allows for multiple interpretations [3 - 6]. These visuals include characters, their structures, the environment in which the story is set, the mood of the various characters, emotional changes, and others, each of which leaves a deep impact on the mind and affects the enjoyment of the story [7-9].

Traditional Indian graphics such as Pran's "Chacha Choudhury" and recent additions such as "Chota Bheem" have undergone a sea of change, starting from changing hand drawings to using illustrative designing tools, changes in panel depictions and others. "Graphic-Con" and studios such as Marvel's Graphic World, Disney, Pixar, Dreamworks and many more have brought graphics into high-grossing movies. Such narration visualizations are compelling the mapping from the narrative zone to presenting a computationally viable graphic engine. For example, in Voki.com, shy students use avatars to present their ideas and voice their thoughts. However, these avatars do not display many facial expressions, which may render them unfriendly to the users. Also, there are no automated or semi-automated systems or tools that can allow for the creation of a graphic frame that may allow for background settings and foreground characters along with the placement of relevant dialogue boxes. Further, there is no open-source software that considers a story narration as an input to form a caricature or a character, and nor are there any tools to extract an appropriate background by sensing the context of the story. This work proposes to semi-automate story narration and formalizes it as a graphic illustration to enhance the story experience. Such a graphic engine would encourage artists, writers, and designers to form and map their stories into a well-panelled graphic illustration formation for initial representations.

In this work, we propose to use textual narration as input to generate a graphic illustration. Firstly, Natural Language Processing (NLP) is used to extract the elements of the story and understand its various phases. Secondly, interpret the elements to feature descriptors. Thirdly, appropriate visual elements should be retrieved to map feature descriptors to match the tags of the illustrator's drawings. Finally, the selected illustrations are placed to form a visual graphic representation. The matching process uses transfer learning based on the feature space remapping function. After the initial query retrieval of all the images with the relevant tags, the similarity measure further fine-tunes the search result. The user can then choose any of the illustrations retrieved from the tagged image database.

The next section depicts the current related works with respect to theories of textual story narration, Part-of-speech tagging framework, and image tagging. Section III presents the proposed methodology for translating a story from textual narration to graphic illustration. The experimental results are detailed in Section IV. The feedback received from the test target group is aggregated in section V. Certain further improvements are mentioned in section VI. Section VII concludes the paper.

II. LITERATURE SURVEY

A short story may be considered to be a short piece of fiction. These are among the oldest forms of storytelling. Ancient texts and mythologies were considered to be a storehouse of stories and were passed on through oral language, so stories needed to be brief. Story narration is a fundamental form of communication. This is not only for entertainment but also for knowledge transmission through stories.

The computational transition from narration to visual depiction, especially as a graphic illustration, involves two basic dimensions of data. These dimensions must be processed and mapped the latter to get resultant data. One aspect of the system involves textual narration and story processing, while the other aspect involves mapping the illustrations to present an appropriate frame. Presentation of a story as a set of image frames in terms of graphic illustrations involves understanding the narration aspects as well as relevant computational requirements. The resultant frames should have simplicity and rhythm to generate a graphic illustration. While creating the characters and their environments, story details are important. For adequate visual effects, a proper panel of frames is of momentous importance for composition and page layout [10 - 18]. Apart from following the 12 principles of Animation [19], frame formation is also very important.

Recent studies in narration showcase the usage of various parameters for translation and transmission. These consist of setting, characters, and plots and point-of-views. 'Settings' gives an overview of the basics of the story's actions of time and place of happening, such as geographic locations, customs, beliefs, ideas and values, rituals, etc. These help give the story wholesomeness. These influence the narration, era, costumes, characters' reactions to various behaviours and protocols, mood and atmosphere. It sets the basic feeling of the story, climate conditions such as weather and temperature, populations and ancestral influences. Computationally, depicting setting parameters is very difficult, and its visualisation can only be through background and character behaviour. Characters play the plot so that the sequence of events can be advanced. Depending on the writer's choice, characters can be animals, people, or animate things. The story text provides information about various characters and their attributes, such as name, gender, and others. As the story progresses, various information such as what the secondary character wants from the protagonist, how the secondary character feels about the protagonist, where the story takes place – setting/description of the place – from adjectives, emotional adjectives, adjectives to describe objects; character descriptions; occasions; verbs and their associations with characters; nouns to compare characters; organization; time of the day; story theme, etc. A story usually processes through five stages of the plot, namely, exposition, rising action, climax, falling action and resolution. The aspects can be propelled via the conflicts, which may be emotional, intellectual, moral, etc. Plots may be presented chronologically, in flashbacks, or as media-res. Plot extraction may be presented computationally as a set of relational tuples based on neural conferencing and NLP.

Computational processing of textual narration and story processing involves the domain of NLP, which is augmented by Deep Learning approaches. These allow for the extraction of characters, backdrop, theme, organisation name, settings, etc. Named-Entity Recognition (NER) using Deep Learning having skip-gram and bag of words methods has given good results. On the other hand, tagging of illustrations involves associating multiple tags to these images, which can later be

applied to mapping from narration storyline to illustration frames. Initial image classification is performed using techniques used by authors in [20 - 25]. Further, automatic image annotation, to get better tagging data, is also performed using methods as shown by [26]. Relevant images are then processed for their associated sentiments using the method mentioned by authors [27], [28].

The limitations found in the reading include the fact that there is, currently, no automatic or semi-automatic system for assisting illustration artists in translating a textual narration to a graphic illustration setting. The methodology presented in this work is a step towards reducing this gap.

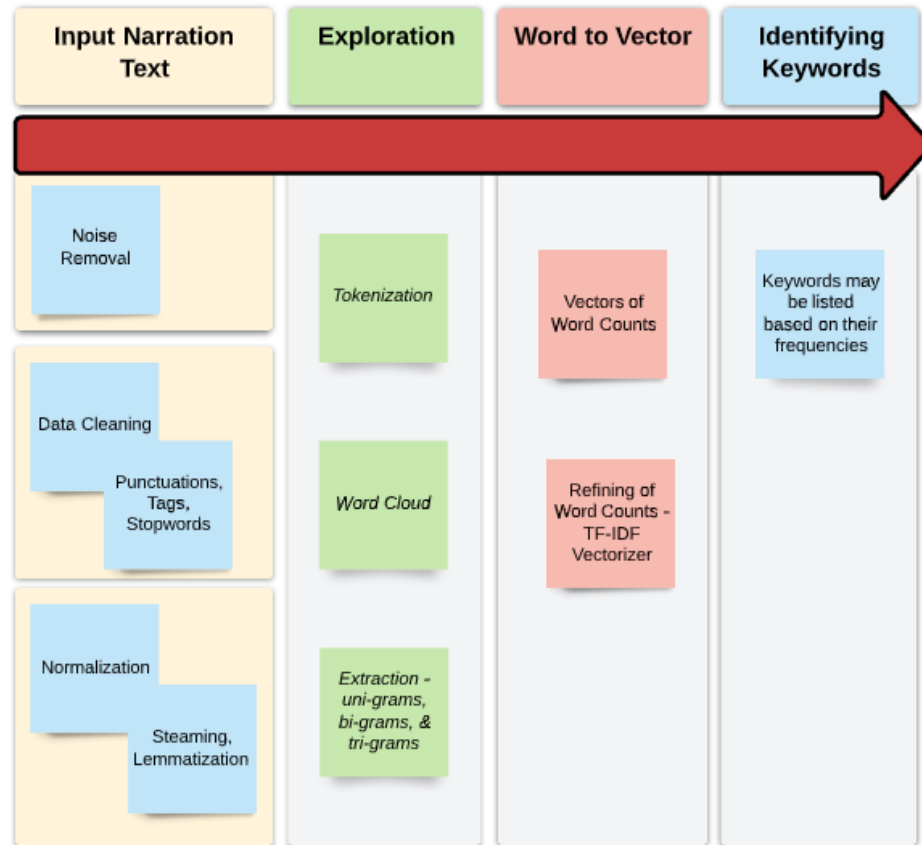
III. METHODOLOGY

- i. Extraction of nouns, named entities
- ii. noun – feature descriptors
- iii. extract appropriate characters/ objects for visual depiction.
- iv. list of various emotions
- v. list of body postures
- vi. sound words
- vii. dialogue and conversation elements extraction
- viii mapping various elements to characters.
- ix framing
- x. storyboarding

The intent, as mentioned in the sections earlier, is to form a graphic illustration from a textual narration. There is no automatic process for this, however, a semi-automatic procedure has been proposed in this work.

If the input is a textual narration consisting of short sentences, the basic steps may be enumerated as follows:

- (i) Extraction of story elements –
 - (a) Extraction of Keywords



In this stage, all keywords can be extracted. These keywords will help in extracting the story elements. Input to the system is textual narration or a short story. This is fed to the POS-NER module [29] for the extraction of characters, places, time, and other entities. This stage is followed by establishing relationship rules. These entities are mapped to the story elements, i.e., Characters - their names, their associated adjectives, gender, person/animal; Association between primary and secondary characters; Setting, i.e., physical look-and-feel of the place where events are taking place; Verbs and elements that are associated with activities; Events and their timings. The nouns extracted through the NLP parser help select the images to be used by matching with their associated tags [30 - 34], [38]. In any given sentence, each word belongs to a definite grammatical class such as noun, verb, adjective, adverb, article, etc [29], [35], [36], [37]. A POS tag set is a set of terms representing POS classes, including a simple tag set and complex morphological information. Most words have a preferred class. For English, Brown tag set and Penn tag set are widely known POS tagging sets. Initially, POS taggers assign one POS tag to each input token. There exist multiple ways of breaking down POS tagging [29], [26]. Apart from POS tagging, Named-Entity-Recognition (NER) is also important.

(b) Extraction of Narrative events.

In a textual narration, an event may be represented as a tuple constituting the expressions or in terms of what, who, why, where, when, and how (5W1H) extracted from the parts of speech tags. The Parts-of-Speech (POS) tags and Named-Entity Relationships (NERs) can depict the components and their relationships, such as subjects, verbs and objects. An event may be

expressed as a tuple <Character: Verb, Dependency>. Dependency depicts the role of a character in relation to the verb or preposition. If the 5W1H can be extracted, it would present the elements such as subject, predicate, object, location, time, etc. These are then used to find the grammatical relationship between words and sub-divide phrases to extract the action verb and its relationship to other components. Standardization method, such as lemmatization, is used to reduce the complexity of sentences and make analysis much more efficient by restoring morpheme units to their basic forms [44]. NER may also be used to depict the relationship.

Example:

Narrative text: Shivika was walking along a road

Extracted Events: $e_s(\text{Shivika})$, $e_v(\text{walk})$, $e_o(\text{road})$

(c) Extraction of Narrative chain.

A Narrative chain is said to be a partially ordered set of narrative events that share a common character. These chains can contain a set of events and a binary relationship between the event set. These POS tags, Narration events and chains are then converted to feature descriptors.

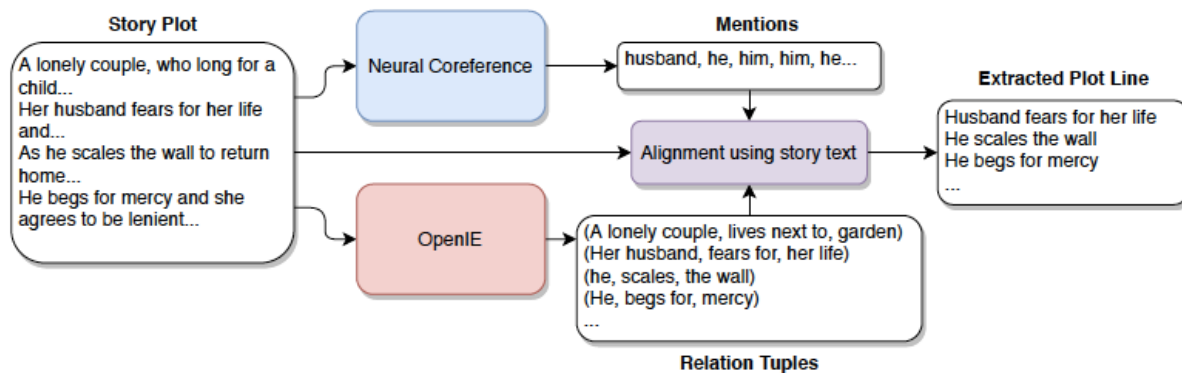


Figure 1: An illustration of high level plot point extraction.

- (ii) Next, dialogues from the narrative structure are to be extracted. Text in the narration may contain dialogues, either stated directly or indirectly. Dialogues are the words spoken by characters in the narration, whereas action sequences help give direction to how the characters perform or react. A dialogue may be said to consist of two parts – characters who are uttering the dialogue and the dialogue text itself. Dialogues are extracted using the context of the sentences themselves. Verbs like 'said' and 'told' and their various forms, when used with proper nouns, may give rise to dialogue, especially if the said text is enclosed within double quotations.
- (iii) The third stage consists of a mapping feature descriptor to image tags.
- (iv) The fourth stage is that of choosing passable illustrations pertaining to image tags.
- (v) The fifth stage sets the arrangement of the image as required by the story narrative and associating the dialogues with the characters.

The flow between these steps can be depicted as shown in Figure 1.

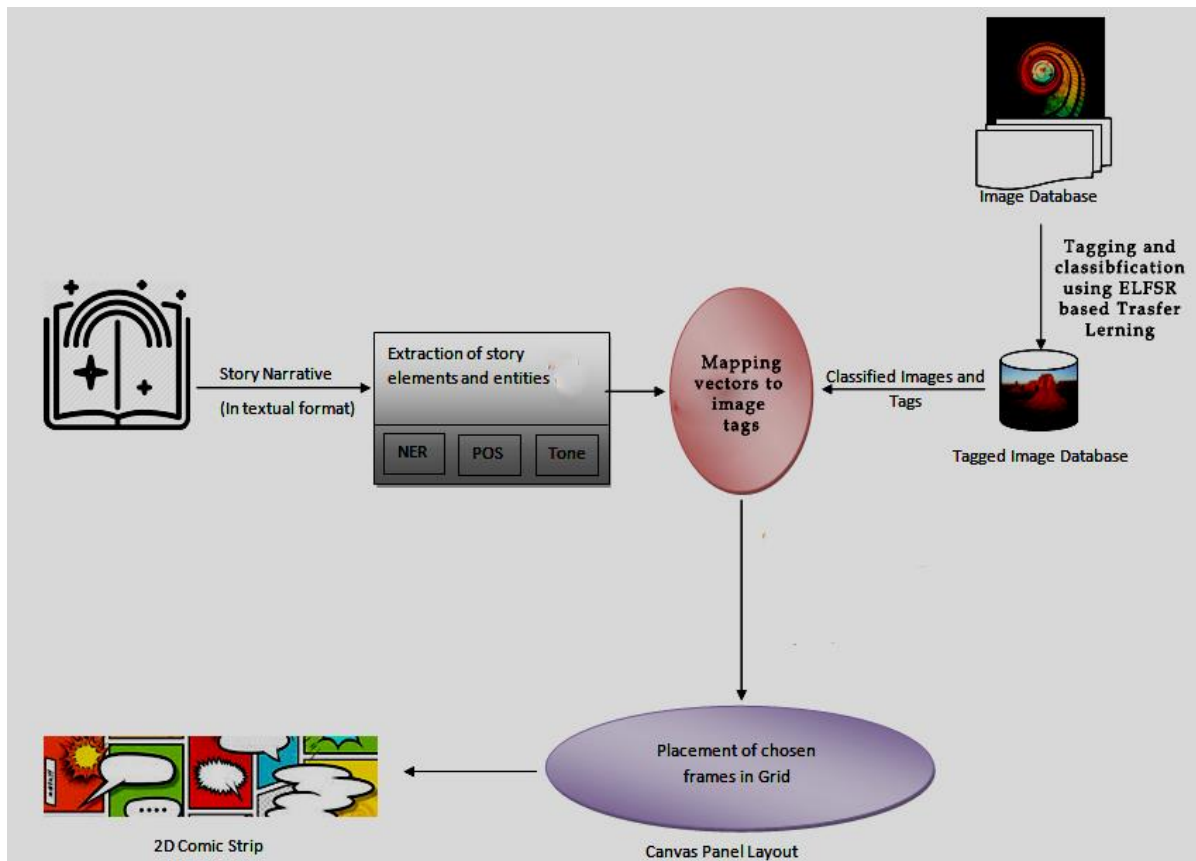


Figure 3: "GraphicsAssist" Module Diagram

A. Narration to POS-NER tags

B. Associating Emotions to Characters

The adjectives play an important role in establishing a relationship between objects and/or nouns by demonstrating ownership. In this case, possessive adjectives are most suitable. Descriptive adjectives associated with their noun counterparts present their different quantifiers. These present the associated noun's opinion, action, state or quality. These quantifiers may be colour, size, taste, and texture, among others. These adjectives are used for the portrayal of emotion related to the main narration of the characters. The six emotions of happy, sad, angry, surprise and indefinite adjectives give some idea of the count – some, few, many, etc. In our work, these have been used alongside nouns for the depiction of various narration characters. These are usually used in the representation of secondary characters.

The WordNet list of adjectives is used (WordNet Database) as it also gives their associated sentiments. These basic sentiments and emotions are part of the matching keywords that are used to retrieve characters.

C. Selection of Characters from Illustrative Drawing/Illustrations.

The illustration dataset is created by crawling various sites. Each of these collected illustrations is then mapped to image tags. Image classification and tagging are implemented using the transfer learning approach.

Convolution Neural Network (CNN) has emerged as a leading architecture for image recognition and classification. Other methods, too, have shown promising results for tagging illustrative images [20], [21], [22]. CNN model can be controlled by varying its depth and breadth. Deep CNN has further fueled its usage. The input image is fed to the network that undergoes several stages of convolution and pooling. Then, the representations from these operations feed the connected layer, which finally outputs the class label [24], [25]. Feature extraction is performed by the convolution layers. This can be represented as: $Y_k = f.(W_k * x)$ where Y_k is the k^{th} output feature map, x is the input image, W_k is the convolution filter related to the k^{th} feature map, $*$ is the 2D convolution operator, and $f(.)$ represents the nonlinear activation. The pooling layers help reduce the spatial resolution of the feature maps and attain spatial invariance to input distortions and translations. The aggregation layers are used to transmit the maximum value within each receptive field to the next layer. This is represented as $Y_{k ij} = \max_{(p,q) \in \mathfrak{R}_{ij}} x_{k pq}$ where $Y_{k ij}$ denotes the output of the pooling operation associated with the k^{th} feature, $x_{k pq}$ represents the element at the location (p, q) contained by the pooling region \mathfrak{R}_{ij} . This region is the receptive field around the position (i, j) . Multiple convolutional and pooling layers are then stacked on top of each other to extract more abstract feature representations. They are followed by fully connected layers to perform high-level classification. This is performed by implementing the softmax operator or support vector machine. After this, CNN training takes place to adjust its biases and weight parameters. The training is implemented using back-propagation that determines the gradient of the performance function. The initial matching involves using dictionary-based searching for particular tags. However, there are many cases where proper tagged data is not available, or the coarse selection presents a huge set of data for depiction. The selection in such cases is done using transfer learning with a feature-space remapping function mechanism.

Transfer learning works on the basic principle of learning from a pre-trained model and transferring or applying this knowledge to the non-trained smaller dataset. For objects classified with the earlier used CNN, the initial convolution layers of the network are frozen, and the latter remaining layers are used for making a prediction. The convolutional layers are used to extract low-level features such as strong edges, regular patterns, and gradient functions. The final layers are used for extracting specific features or patterns that help in finer classification. This allows for classifying illustrations that may have an affine transformation mapping or a limited degree of non-affine transformations, such as skew or wrap functions associated with them [40], [41]. An ensemble classifier is then built by defining the mapping function between the target domain and source domain features. This classifier, called "Ensemble Learning via Feature-Space Remapping" (ELFSR), is used to predict the class of illustrative drawings.

D. Visualization – Canvas Panel Frames

The formalism of textual narration to graphic illustration depiction is not yet streamlined; rather, it follows an intuitive and simplistic learning model. Initially, a database of illustrations is created so as to cater to the requirement of selecting appropriate characters for depiction. Presently, the set of backgrounds is very constrained. First, all the elements of the narration (Story Elements, as mentioned in Figure 1) have been identified. The image database is a pre-classified, priority-tagged set of illustrations. Since the aim is to formalize the narration as a graphic illustration, illustrative drawings have been considered in the database. The illustrations are of the size of 3" x 3", 72 PPI, JPEG images with sizes ranging from 3KB to 56MB. Another design constraint is to use the *mediumblob* datatype for easy storage.

Story elements are then converted to entities, i.e., features and their feature vectors. Each feature has a feature type and its relevant feature vector set. Feature types are assumed to be of 4 categories – (i) characters, (ii) settings, (iii) adjectives to describe a particular character and (iv) association map. Character feature type has the following associated parameters – type (person/animal/robot/others), gender (male/female), and order (protagonist / secondary). The association map is an adjacency matrix that describes the relationship between various characters, such as feelings and emotions, and their strength. It uses the concept of a bag of words, sentiment labelling, and its associated strength with the characters. This is an n-dimension matrix. The story element description is then mapped with the tags of images.

For our current scenario, we have used only characters as foreground expressions and their tone/emotion. Constraints dictated that background depiction be delayed for this phase of the work. Further, dialogues have also been withheld. Though the placement of the speech balloon is currently non-consequential, it would be important while developing a book. Onomatopoeic words, i.e., words associated with noises such as human sounds, animal sounds, vehicle sounds, and others, are also depicted in speech balloons, and placement is near to characters so as to redesign the facial expression of the characters. Motions and their adverbs are extracted and depicted using minor motion lines and curves [16], [17]. Most extractions are done using the Named-Entity-Recognition module and stored in vectors.

IV. EXPERIMENTATION AND RESULTS

Since the latency of the system should be kept as low as possible to have a decent flow in the working of the application, for our current experimentations, a few constraints have been maintained: (a) the textual narrations have not been classified to any particular themes; (b) the narrations closely follow the POS [29], [35 - 37], (c) only textual narration has been used for the depiction of frames. No dialogues, sound words or exaggerations are incorporated; (d) it has been assumed that the narrations follow Freytag's pyramid for dramatic structure depiction; (e) frame modification is restricted, and the chosen image is assumed to be for foreground object only.

The image database contains images and tags. Image data extraction is based on mining their tags. Initially, a non-tagged illustration image database is formed. Currently, the database consists of illustrative drawings only, as the initial aim is to form frames of cartoon-type illustrations. As of now, the database has almost 3000 images tagged. These tagged images are mined to get

approximate images for the placement within the frames. Currently, only the emotions of happy, sad, and neutral are considered. For animals, no such markers are yet used. At the current state, each tag has only one to two synonyms associated with it. The final application structure is shown in Figure 1.

In this paper, we have depicted details of 2 such case studies. The details of these examples are discussed below. For each narration, POS tags are extracted. Further, enhanced universal dependencies are found, too. These have been extracted using the NLTK libraries. Entities and the association between them are also found. Stanford's NLP libraries [42], [43] have been implemented for the same. The narration text is also passed through NER to distinguish noun type entities of only person and location and numerical data of number and duration only. The tone of the narration at the document level, as well as sentence level, has been detected. Based on these extractions, the tags from the illustrative drawings are matched. The first screen presents only the top 10 matches to the user.

Table 1: Textual Narration; POS Tags; Entities and their associations.

<i>Narration- Storyline</i>	<i>POS Tagging</i>	<i>Entities and their associations</i>
Shivika is a 10-year-old girl. She was walking along a road. Bhola, her dog, was walking with her. They were going to the market.	Shivika/NNP is/VBZ a/DT 10/CD year/NN old/JJ girl/NN ./. She/PRP was/VBD walking/VBG along/IN a/DT road/NN ./. Bhola/NNP ./, her/PRP\$ dog/NN ./, was/VBD walking/VBG with/IN her/PRP ./. They/PRP were/VBD going/VBG to/TO the/DT market/NN./.	Character 1: Shivika (person) Associated Nouns – girl, age-10 yrs, road Verb: walking Character 2: Bhola (animal) Associated Noun – dog. Associated pronoun – her (previous root). Verb: Walking Pronoun: They, market Verb: going Location: Market. Emotion: Happy Final Verb: Walking
Radhika is a small girl of about 5 years. Bhavesh is her elder brother. She was eagerly waiting for her Bhavesh, near the door. Bhavesh was going to bring a toy-	Radhika/NNP is/VBZ a/DT small/JJ girl/NN of/IN about/RB 5/CD years/NNS ./. Bhavesh/NNP is/VBZ her/PRP\$ elder/JJ brother/NN ./.	Character 1: Radhika (person) Associated Nouns – girl, age-5 yrs. Character 2: Bhavesh (person), elder Associated Pronoun – her (previous root), bother. Pronoun: She, door

<p>car for her. Radhika was very excited.</p>	<p>She/PRP was/VBD eagerly/RB waiting/VBG for/IN Bhavesh/NNP ./, near/IN the/DT door/NN ./.</p> <p>Bhavesh/NNP was/VBD bringing/VBG a/DT toy-car/NN for/IN her/PRP ./.</p> <p>Radhika/NNP was/VBD very/RB excited/JJ ./.</p>	<p>Verb: waiting</p> <p>Case: near- associated noun- door</p> <p>Character: Bhavesh (person), toy-car</p> <p>Associated Pronoun – her (person)</p> <p>Associated Verb - bringing</p> <p>Emotion: Happy</p>
---	--	--

Also, the tones of the narration at the document level and sentence level were found. This is required to gauge the general mood of the storyline. Table 2 depicts the tones at various narration levels for the considered test cases. The placing of the frames in the timeline follows the progress of the story narration itself without consideration of the time periods that may have been included within the narration itself.

Narration 1:

Frame 1: A girl in the foreground (background not included)

Frame 2: A dog,

Frame 3: Road

Frame 4: Marketplace.

Narration 2:

Frame 1: A girl in the foreground (background not included)

Frame 2: Brother

Frame 3: Toy Car




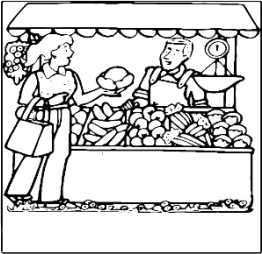




Table 2: Tones at various narration levels

Example	Document-level	Sentence-level
Narration 1	Joy: 0.56	<p>0.86 They seemed very happy.</p> <p>0 Shivika is a 10-year-old girl.</p> <p>0 She was walking along a road.</p> <p>0 Bhola, her dog, was walking with her.</p> <p>0 They were going to the market.</p>
Narration 2	Joy: 0.83 Confident: 0.68	<p>0.99 Radhika was very excited.</p> <p>0.57 Radhika is a small girl of about 5 years.</p> <p>0.55 Bhavesh was going to bring a toy-car for her.</p> <p>0 Bhavesh is her elder brother.</p> <p>0 She was eagerly waiting for her Bhavesh, near the door.</p>

Images tags and classification is presented in table 2. For our experimentation, we currently have 19271 images that were tagged and with their listed entities. These were used for the extraction of the illustrations by matching the keywords extracted from the story entities.

Depiction of the illustrations retrieved by search keywords with respect to the tags associated with the images is depicted. The keywords extracted are used as search terms. After the retrieval of illustrations from the database as per their tags, the user can choose a particular image as per his requirements. Currently, manipulation of the images is not allowed. Once the user chooses the image, it is showcased as a single frame in the graphic illustration canvas. In the current experiment, the grid layout is used. Since only a horizontal illustration depiction is required, vertical divisions in the horizontal frameset are used. For the examples explained in this work, image frames used in the canvas layout are depicted in Table 3.

Table 3: Outcomes of “GraphicsAssist” system - Graphic illustration from Story Narration.

<p>Narration 1: Shivika is a 10-year-old girl. She was walking along a road. Bhola, her dog, was walking with her. They were going to the market.</p>	
<p>Images that can be used in frames:</p>	
	  
<p>No. of scenes: 1</p>	
<p>Frames</p>	
<p>Narration 2: Radhika is a small girl of about 5 years. Bhavesh is her elder brother. She was eagerly waiting for Bhavesh, near the door. Bhavesh was going to bring a toy-car for her. Radhika was very excited.</p>	
<p>Images that can be used in the frames:</p>	
	  
<p>No. of scenes: 1</p>	
<p>Frame</p>	

V. VALIDATION OF THE PROPOSED ARTIST ASSISTANCE SYSTEM

This graphic illustration designer tool was given to a target closed group for validating its usage. The feedback received was good, considering that this is an initial system as an artist and designer's assistant. The dataset was created earlier by prior crawling. The metric used included the following parameters as shown in table 4.

Table 4: Metric for subjective evaluation of the "Narration-to-Graphic" system.

Metric	Percentage based on user satisfaction
Character detection	100%
Environment setting detection	62%
Extraction of relevant characters	100%
Frame panel formation	36%

This proposed system is unique as it one of its kind. There is no other online or freeware system that can help as an artist's assistant by assimilating various illustrations for a given storyline.

VI. CONCLUSION

The undertaking considered for presenting this literature starts with textual narration as an input to the system and getting a set of probable parameters for presenting an assortment of illustrative visuals. With the text, the systems dissect it into a set of entities - characters, emotions, and associations. These entities then form the basis of the search keyword. On the other end of the spectrum, an image database of illustrations is considered. After being tagged, these are to be used for the visual depiction. The process, though not perfect, is able to do certain cognitive justification to the narration storyline.

The various modules of this translation, from textual narration to graphic illustration visualization, are mentioned in detail. The initial phase consists of dissecting the narration to extract the story elements. These elements form the search vectors. These include the ramifications of each sentence, associating them with their corresponding parts of speech, and then applying NER formations and RACK structure approaches. The emotion is also extracted. These attribute sets are used as tags to match the illustration database. Further, a tagged illustration database is also maintained, wherein each illustration is tagged using multiple text tags of shape, size, colour, gender, and other classification attributes. The retrieval of images from the database is built on the query of the search keywords. The query then applied a finer matching technique based on transfer learning using the technique of ensemble learning via feature-space remapping. The user can choose any of the illustrations, as retrieved, to assemble the frames of the graphic illustration panel.

The "GraphicsAssist" application is a semi-automatic system for the generation of graphic frames from a given storyline. It is a very novel application that considers the multidisciplinary widely

different domains of literature (narration) for computational prediction of frames for the depiction of a graphic illustration. Though this is a semi-automatic procedure and a finer approach may be required to generate a better graphic frame, the non-existence of such systems makes this application very novel in this domain. Numerous shortcomings are the basis of future enhancements that will optimize the system as well as make it aesthetically more acceptable to graphic artists.

This work is useful for graphic artists as it can help in creating an initial phase of short graphic-illustrations. "GraphicsAssist" is a semi-automatic tool for finding entities that can depict the textual narration. Later, it may also be extended to accommodating graphic books and small animated series using appropriate animation principles and techniques, virtual/augmented environments, and games for use in entertainment, education, cultural and traditional literature preservation, amongst others. Numerous requirements pertaining to information extraction are being concluded satisfactorily, yet cognitive depiction and their evaluations are a huge hurdle. This may lead to reliability issues with respect to the cognition requirements of the narration.

References

- [1] Salovey, Peter, and John D. Mayer. 1990. "Emotional intelligence." *Imagination, cognition and personality* 9, no. 3: 185-211.
- [2] Grimm, Michael, Emily Mower, Kristian Kroschel, and Shrikanth Narayanan. 2006. "Combining categorical and primitives-based emotion recognition." In *Signal Processing Conference, 2006 14th European*, pp. 1-5. IEEE.
- [3] Alm, Cecilia Ovesdotter, and Richard Sproat. 2005. "Perceptions of emotions in expressive storytelling." In *Ninth European Conference on Speech Communication and Technology*.
- [4] Batliner, Anton, Kerstin Fischer, Richard Huber, Jörg Spilker, and Elmar Nöth. 2003. "How to find trouble in communication." *Speech communication* 40, no. 1-2: 117-143.
- [5] Ericsson, Karl Anders, and Herbert Alexander Simon. 1993. *Protocol analysis: Verbal reports as data*, Rev. the MIT Press.
- [6] Hoque, Mohammed E., Mohammed Yeasin, and Max M. Louwerse. 2006. "Robust recognition of emotion from speech." In *International Workshop on Intelligent Virtual Agents*, pp. 42-53. Springer, Berlin, Heidelberg.
- [7] Cohn, Jeffrey F., and Takeo Kanade. 2007. "Use of automated facial image analysis for measurement of emotion expression." *Handbook of emotion elicitation and assessment*. 222-238.
- [8] Craig, Scotty, Sidney D'Mello, Barry Gholson, Amy Witherspoon, Jeremiah Sullins, and Arthur Graesser. 2004. "Emotions during learning: The first steps toward an affect sensitive intelligent tutoring system." In *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, pp. 264-268. Association for the Advancement of Computing in Education (AACE).
- [9] Hudlicka, Eva, and Michael D. Mcneese. 2002. "Assessment of user affective and belief states for interface adaptation: Application to an Air Force pilot task." *User Modeling and User-Adapted Interaction* 12, no. 1: 1-47.

- [10] Spisak, April. 2008. "Drawing Words & Writing Pictures: Making Graphics: Manga, Graphic Novels, and Beyond." *Bulletin of the Center for Children's Books* 62, no. 2: 62-63.
- [11] Amheim, Rudolf. 1974. "Art and Visual Perception, a psychology of the creative eye. The New Version."
- [12] Barber, John. 2005. "The phenomenon of multiple dialectics in graphics layout." *Retrieved October 16 (2002)*.
- [13] Cohn, Neil. 2013. "Visual narrative structure." *Cognitive science* 37, no. 3: 413-452.
- [14] Cohn, Neil, Philip Holcomb, Ray Jackendoff, and Gina Kuperberg. 2012. "Segmenting visual narratives: evidence for constituent structure in graphics." In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 34, no. 34.
- [15] Dobel, Christian, Gil Diesendruck, and Jens Bölte. 2007. "How writing system and age influence spatial representations of actions: A developmental, cross-linguistic study." *Psychological Science* 18, no. 6: 487-491.
- [16] Goodbrey, Daniel. 2015. "The Sound of Digital Graphics." *Writing Visual Culture*.
- [17] Drucker, Johanna. 2008. "Graphic devices: narration and navigation." *Narrative* 16, no. 2: 121-139.
- [18] Holmqvist, Kenneth, and Constanze Wartenberg. 2005. "The role of local design factors for newspaper reading behaviour-an eye-tracking perspective." *Lund University Cognitive Studies* 127: 1-21.
- [19] Johnston, Ollie, and Frank Thomas. 1981. *The illusion of life: Disney animation*. New York: Disney Editions.
- [20] Lu, Dengsheng, and Qihao Weng. 2007. "A survey of image classification methods and techniques for improving classification performance." *International journal of Remote sensing* 28, no. 5: 823-870.
- [21] Quattoni, Ariadna, Michael Collins, and Trevor Darrell. 2009. "Transfer learning algorithms for image classification." PhD diss., Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science.
- [22] Quattoni, Ariadna J. 2005. "Object recognition with latent conditional random fields." PhD diss., Massachusetts Institute of Technology.
- [23] Guo, Tianmei, Jiwen Dong, Henjian Li, and Yunxing Gao. 2017. "Simple convolutional neural network on image classification." In *Big Data Analysis (ICBDA), 2017 IEEE 2nd International Conference on*, pp. 721-724. IEEE.
- [24] Ciregan, Dan, Ueli Meier, and Jürgen Schmidhuber. 2012. "Multi-column deep neural networks for image classification." In *Computer vision and pattern recognition (CVPR), 2012 IEEE conference on*, pp. 3642-3649. IEEE.
- [25] Rawat, Waseem, and Zenghui Wang. 2017. "Deep convolutional neural networks for image classification: A comprehensive review." *Neural computation* 29, no. 9: 2352-2449.
- [26] Yang, Yang, Zi Huang, Heng Tao Shen, and Xiaofang Zhou. 2011. "Mining multi-tag association for image tagging." *World Wide Web* 14, no. 2: 133-156.
- [27] You, Quanzeng, Jiebo Luo, Hailin Jin, and Jianchao Yang. 2015. "Robust Image Sentiment Analysis Using Progressively Trained and Domain Transferred Deep Networks." In *AAAI*, pp. 381-388.

-
- [28] Yuan, Jianbo, Sean Mcdonough, Quanzeng You, and Jiebo Luo. 2013. "Sentribute: image sentiment analysis from a mid-level perspective." In *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining*, p. 10. ACM.
- [29] Cutting, Doug, Julian Kupiec, Jan Pedersen, and Penelope Sibun. 1992. "A practical part-of-speech tagger." In *Proceedings of the third conference on Applied natural language processing*, pp. 133-140. Association for Computational Linguistics.
- [30] Pantic, Maja, and Leon JM Rothkrantz. 2003. "Toward an affect-sensitive multimodal human-computer interaction." *Proceedings of the IEEE* 91, no. 9: 1370-1390.
- [31] Shafran, Izhak, and Mehryar Mohri. 2005. "A comparison of classifiers for detecting emotion from speech." In *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, vol. 1, pp. I-341. IEEE.
- [32] Finney, Nathaniel, and Jordi Janer. 2010. "Soundscape generation for virtual environments using community-provided audio databases." In *W3C Workshop: Augmented Reality on the Web*. Barcelona.
- [33] Gong, Yunchao, Yangqing Jia, Thomas Leung, Alexander Toshev, and Sergey Ioffe. 2013. "Deep convolutional ranking for multilabel image annotation." *arXiv preprint arXiv:1312.4894*
- [34] Geng, Yanyan, Guohui Zhang, Weizhi Li, Yi Gu, Ru-Ze Liang, Gaoyuan Liang, Jingbin Wang, Yanbin Wu, Nitin Patil, and Jing-Yan Wang. 2017. "A novel image tag completion method based on convolutional neural network." *arXiv preprint arXiv:1703.00586*.
- [35] Kupiec, Julian. 1992. "Robust part-of-speech tagging using a hidden Markov model." *Computer Speech & Language* 6, no. 3: 225-242.
- [36] Owoputi, Olutobi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider, and Noah A. Smith. 2013. "Improved part-of-speech tagging for online conversational text with word clusters." Association for Computational Linguistics.
- [37] Charniak, Eugene, Curtis Hendrickson, Neil Jacobson, and Mike Perkowitz. 1993. "Equations for part-of-speech tagging." In *AAAI*, vol. 93, pp. 784-789.
- [38] Wu, Jiajun, Yinan Yu, Chang Huang, and Kai Yu. 2015. "Deep multiple instance learning for image classification and auto-annotation." In *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, pp. 3460-3469. IEEE.
- [39] WordNet Database. Accessed from URL <https://wordnet.princeton.edu/wordnet/frequently-asked-questions/database/> as on 13th Feb 2018.
- [40] Feuz, K. D., & Cook, D. J. (2015). Transfer learning across feature-rich heterogeneous feature spaces via feature-space remapping (FSR). *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(1), 3.
- [41] Chattopadhyay, R., Sun, Q., Fan, W., Davidson, I., Panchanathan, S., & Ye, J. (2012). Multisource domain adaptation and its application to early detection of fatigue. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 6(4), 18.
- [42] Chen, Danqi, and Christopher Manning. 2014. "A fast and accurate dependency parser using neural networks." In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 740-750.

- [43] Manning, Christopher, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. "The Stanford CoreNLP natural language processing toolkit." In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pp. 55-60.
- [44] Yu Hye-Yeon, and Kim Moon-Hyun, "Automatic Event Extraction method for Analysing Text Narrative Structure", *2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, 2021, pp. 1-4.