# VizStory: Visualization of Digital Narrative for Fairy Tales

Chieh-Jen Huang
Department of Computer Science,
National Chengchi University,
Taipei 116, Taiwan
g9611@cs.nccu.edu.tw

Cheng-Te Li
Graduate Institute of Networking and Multimedia,
National Taiwan University,
Taipei 106, Taiwan
d98944005@csie.ntu.edu.tw

Man-Kwan Shan
Department of Computer Science,
National Chengchi University,
Taipei 116, Taiwan
mkshan@nccu.edu.tw

## ABSTRACT

Pictures can realize the impressions of texts for readers, especially for fairy tales. If we can present the fairy stories in the form of visual pictures, children will not only be more willing to concentrate their attentions on but also easily perceive the underlying messages. This work aims to present the fairy tales by transforming the texts to the visual form of pictures. To achieve such goal, we develop a system, VizStory, which consists of three steps. First, we investigate the narrative structure of the story to segment the whole story. Second, we select representative keywords for each segment. Third, through Web image search, we find suitable pictures to compose the visualization. Experimental results with human study show VizStory can present the narrative of stories with 75% accuracy.

***Index Terms*****—**Fairy Tales, Digital Narrative, Image Retrieval, Story Telling, Visualization

## 1. INTRODUCTION

A Picture is worth a thousand words. People tend to use different visual forms to help present what they want to express. With the development of digital technologies, more manners of visual presentation are created to accompany with diverse kinds of texts, such as newspaper, magazine, and weblog. For children, it is more obvious for the visual effects on enhancing their interests and understanding the contents of fairy tales [2], such as Snow White, Cinderella, and The Ugly Duckling. Due to the interestingness, the imagination, and the symbolization, the texts of fairy stories are usually interweaved with rich visual elements, including drawing, illustrations, and pictures. The visual clues can not only concretize the characters as well as their interactions in the fairy tale stories, but also realize the abstractive concepts of narrative structures [10]. This kind of visualization had been studied to be able to significantly benefit the process of cognition and learning for children [4].

In this study, to provide better reading experience for children, we aim at visualizing the fairy tales by automatically transforming and mapping the texts of fairy tale stories to pictures. Specifically, our goal is to create connections between fairy texts and pictures. We investigate the narrative structures of fairy stories and search for the suitable pictures from the Web, to accompany with the corresponding story paragraphs.

Finding suitable pictures for a story is challenging. Such task contains three critical parts. First, we are required to automatically understand the narrative structure of the story because different sections of the story depict diverse plots and convey various interactions between characters. Second, how can we determine and represent the central idea for each narrative section? Third, even with the key description of a paragraph section, where and how to find suitable pictures to accompany with? To provide reasonable and representative visualization for a story, such three problems are critical to be tackled.

We develop a novel story visualization system, *VizStory*, to present fairy tales by pictures, where the abovementioned three issues are carefully dealt with. As the user inputs the plain text of a fairy story, we first analyze the narrative structure by some text processing techniques. And then we select three types of representative keywords, leading keywords, theme keywords, and context keywords, to summarize and describe each story segment. Through Web image search engine, and by treating the selected keywords as the query, we are able to retrieve relevant pictures with respect to the input fairy tale. Finally, we determine the most representative pictures with visualization presentation to interweave with the corresponding story segment.

## 2. RELATED WORK

Some existing works find suitable images for diverse kinds of texts. Itabashi and Masunaga [5] consider that a series of sentences in a document can constitute a story scene to correlate images with the corresponding sentences. Joshi et al. [6] devise a story picturing engine to rank and select appropriate images for the description of short texts. Zhu et al. [13] generate pictures for general, unrestricted texts by defining a picturability measure to model the gist.

978-1-4799-2529-2/13 $31.00 © 2013 IEEE
DOI 10.1109/TAAI.2013.26

Balabanovic et al. [1] utilize the personal photo collection to tell the diary stories of users. Li et al. [7] convert words to images from Web by modeling the diverse semantic aspects of words. UzZaman et al. [11] summarize complex articles using Web images by detecting the events hidden in sentences. Mihalcea and Leong [9] find images for dictionary words as a kind of visual linguistic representations of machine translation. Though these studies are able to select pictures for texts, none of them consider the narrative structure of a story, which is what we target at and exploit for pictorially presenting the fairy tales.

## 3. THE PROPOSED METHOD

We give the system overview in Figure 1. We first decompose the story into segments which belong to the same scene or convey similar ideas. For each segment, we extract some keywords from both the whole story and the segment. And then we perform the keyword-based image search in Web to find and select suitable pictures for each text segment to present.

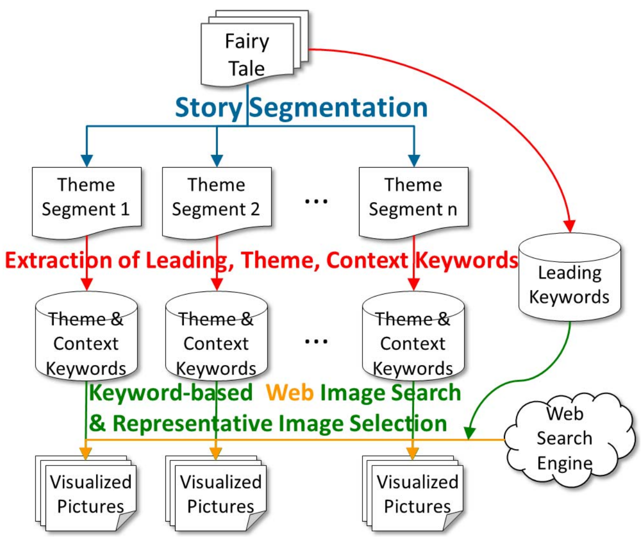


Figure 1: System Overview of *VizStory*.

### 3.1. Story Segmentation

The goal of story segmentation is to divide the whole story into several units, in which each unit talks about the similar theme. For the narrative structure of fairy tales, a story segment/theme usually consists of multiple topics/events. The authors tend to describe the occurrence of story events in a progressive manner. As a result, the endings of paragraphs do not mean the endings of a story segment/theme. But the characters, ideas, or textual description of things are inclined to appear across paragraphs for a certain theme. We leverage such progressive occurrence clues of terms to divide the whole story into segments with similar topics. A method considering the frequency and distribution of words is devised to perform the segmentation. Here we skip the method for brevity. We use Snow White as an example to illustrate the results of story segmentation, as shown in Figure 2.

### 3.2. Keyword Extraction

The second step is to extract keywords to represent the central plot in each theme segment. To accurately characterize the storyline, we categorize the keywords to three types. The first one is the *Leading Keyword* (KL), which aims to keep the characters, things, or events throughout the whole story. The second one is the *Theme Keyword* (KT), which targets at capturing the main plot in each theme segment. The third one is the *Context Keyword* (KC), which describes the supporting elements in each segment.

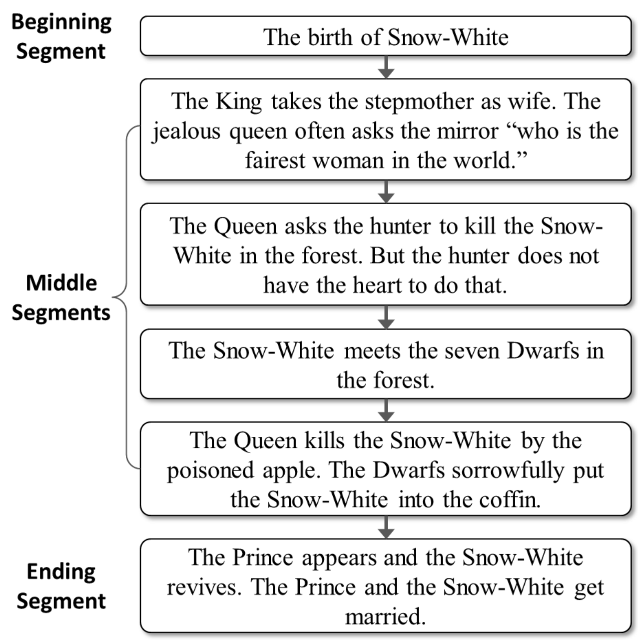


Figure 2: The segments of the storyline for Snow White.

We devise the *Keyphrase Extraction Module*, denoted by *KEM*, as the basis to extract three kinds of keywords. The goal of *KEM* is to determine whether or not a term/word is a keyword. We design *KEM* to be a supervised phrase learning, which consists of the training stage and the extraction/testing stage. In the training stage, *KEM* learns the keyphrase decision classifier from manually-annotated data. And we apply the classifier to determine the unlabeled keywords in the extraction stage. We extract the features suggested by [14] to train the classifier. Especially, we find the following two features are the most effective. The first is *Term Frequency*. The importance of a word/term is in proportional to the occurrence frequency in a document. Given a document $d$ (i.e., a segment), we compute the *Term Frequency* (*TF*) of a term $t$ by $TF(t,d) = freq(t,d)/size(d)$, where $freq(t,d)$ is the occurrence count of term $t$ in document $d$, and $size(d)$ is the total count of terms in document *d*. The second is *Term Position***.** Terms which are located at the starting and ending positions of a document tend to be the important ones. We compute the relative position of a term to be the feature: $TP(t,d) = first(t,d)/length(d)$, where where $first(t,d)$ is the first appeared position of term *t* in document *d*, and $size(d)$ is the length of document *d*.

We utilize the proposed *KEM* method to extract the leading and theme keywords. For either the level of the whole story or segments, all noun phrases with n-gram words in the document are feed into the training stage. The only difference between extracting the leading and the theme keywords lies on using diverse rages of texts for the testing/extraction stage, the whole story text for the leading keywords while the story segments for the theme keywords. Based on the learned classifier, we select words with highest probabilities to be the keywords. Note that we exploit the Naïve Bayes Classifier [15] as the classification method.

For context keywords, they are used to represent the supporting characters, contextual scenes, and other auxiliary things in a theme segment. Since context keywords are usually not obvious and infrequent in the story, we leverage the co-occurrence relationships between words to discover context keywords. The central idea is to take advantage of Web search engine to expand the theme keywords as the context ones. For example, for a segment with words "hunter" and "forest", because "forest" could be infrequent, it is hard to be regarded as a theme keyword. But "forest" could be frequently co-occurred with "hunter" in Web search results. Based on this idea, we devise a two-step method to extract the context keywords. First, to derive the expanded words, we input the theme keywords into Web search engine, and perform the page search result clustering [16], in which those search results with similar themes are clustered. Second, we find the most frequent terms from top large clusters of webpages as the context keywords. Note the selected frequent terms must occur in the theme segment.

Taking the example of Snow White, the extracted leading keywords KL={mirror, snow-white, and queen}. For the seven segments in Figure 2:

- Segment 1: KT={blood, frame}, KC={};
- Segment 2: KT={fairest}, KC={wall, queen};
- Segment 3: KT={huntsman}, KC={forest, heart};
- Segment 4: KT={dwarfs, woman}, KC={world, house, day, tale, queen};
- Segment 5: KT={cheeks, dwarfs}, KC={snow, queen};
- Segment 6: KT={prince}, KC={fairest}.

### 3.3. Query Formation

Based on the discovered leading, theme, and context keywords, we propose to use keyword-based image search from Web to select suitable pictures for each theme segment. Our method consists of two tasks: (1) how to compose the query for Web image search, and (2) how to select representative images as final pictures. We deal with these two tasks in the following. We deal with these two questions by Query Formation (this subsection) and Representative Image Selection (Subsection 3.4). Query Formation aims to construct the query for Web image search based on the discovered keywords in each theme segment. We design two kinds of queries: *Theme Query* and *Event Query*. The former aims to capture which main elements appear in a theme segment while the latter targets at detecting what happens to the characters in a segment.

**Theme Query.** We believe the most important query terms are about the leading characters in the story. In fairy tales, since the leading roles usually occur in both the beginning and ending of the story [4], we examine and consider those leading keywords occurring in both beginning and ending segments as the leading characters. For sentences in each theme segment, we combine the corresponding leading characters with the theme and context keywords using the AND operation to compose the theme query. For example, the content of the second theme segment of White Snow is given, as the following five sentences in Figure 3, in which the words with underline are the detected theme keywords while the words with bold style are the detected context keywords. Based on our query formation method, the queries generated for each sentence are shown in Figure 4. We can find there are totally three queries generated. "mirror **AND** fairest **AND** wall", "queen **AND** fairest", and "mirror." Note that "null" indicates no query is generated for the sentence because neither leading keywords nor theme keywords appear.

(1) And at the same time her mother died.
(2) About a year afterwards the King married another wife, who was very beautiful, but so proud and haughty that she could not bear anyone to be better-looking than herself.
(3) **Queen** owned a wonderful <u>mirror</u>, and when she stepped before it and said: "<u>Mirror</u>, <u>mirror</u> on the **wall**, who is the fairest of us all?"
(4) It replied: "The **Queen** is the <u>fairest</u> of the day."
(5) Then she was pleased, for she knew that the <u>mirror</u> spoke truly.

Figure 3: The sentences in the second theme segment of Snow White.

(1) null
(2) null
(3) queen **AND** mirror **AND** fairest **AND** wall
(4) queen **AND** fairest
(5) mirror

Figure 4: The composed query of sentences in Figure 3.

**Event Query.** In fairy tales, the progress of storyline is significantly affected by the happening of a series of events involved by the leading characters [18][19]. An event in fairy tales refers to the co-occurrence of some story elements. Therefore, to represent the progress of a fairy tale, we aim to extract the events and treat them as the queries. An event usually consists of some characters (i.e., noun) and the actions taken by them (i.e., verb). To make children easily understand, the authors of fairy tales use a basic sentence pattern to the description of an event: *Subject*+*Verb*+*Object*, in which *Subject* and *Object* are characters while the *Verb* captures the interaction between *Subject* and *Object*. Thus we represent an event as *Verb*(*Subject*, *Object*). For example, in the sentence "A pumpkin becomes a carriage", "pumpkin" interacts with "carriage" through "become", and the event is represented by "*become*(*pumpkin*, *carriage*)." We exploit Marneffe et al.'s method [17] to detect such kind of sentences and compose events for sentences in a theme segment. Finally, if both *Subject* and *Object* are identified keywords, we use AND operation to combine them with *Verb* to construct the event query. For example in Figure 3, two events are identified, as highlights by gray: *marry*(*king*, *wife*) and *own*(*queen*, *mirror*). Since *wife* does not belong to any types of keywords, the final event query would be "queen **AND** own **AND** mirror."

### 3.4. Representative Image Selection

*Representative Image Selection* aims at finding the most suitable pictures from the keyword-based Web search resulting images for the final presentation of each segment. The Google Image Search is employed. This task consists of two steps, filtering and ranking, as shown in Figure 5. By using the extracted keywords as query, we retrieve the top-*k* images from the image search engine. These results are treated as the preliminary image set. Note that each keyword query will have a preliminary image set. Then, we leverage the semantic correlation between images to filter out some irrelevant ones, and derive the candidate image set. The next ranking procedure aims to select representative image with respect to the query, with the consideration of visual coherence among determined images.

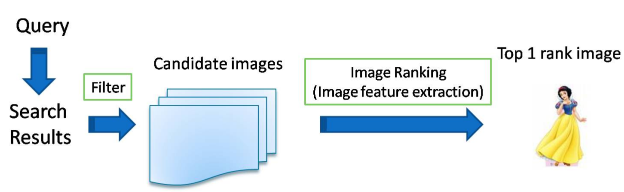


Figure 5: Flowchart to select the representative image.

The filtering step is to rule out irrelevant images to have a candidate image set. We exploit the technique of Weighted AND (WAND) [3] to filter images. For each returned image, we examine whether the corresponding webpage satisfies WAND using the extracted keywords. WAND technique can be written in the following formula:

$$WAND(x_1, w_1, \ldots, x_k, w_k, \theta) = true$$
$$\text{if and only if } \sum_{1 \le i \le k} x_i w_i \ge \theta$$

where $x_i$ is a keyword and $w_i$ is the frequency of $x_i$ in the webpage. We include the image into the candidate image set which will enter into the selection step, if the corresponding webpage satisfies WAND. In the ranking step, for each theme segment, we select some representative images considering visual features of candidate images. The idea is to select images containing the most *visual words*, where a visual word is a group of similar features of images. Specifically, we extract the local image feature, *scale invariant feature transform* (SIFT) [8], for each candidate image. By clustering the SIFT feature points based on the Euclidean distance, we can derive several groups of feature points, and each group is regarded as a visual word. Here we only choose the most frequent visual words. And we select the most representative images according to the numbers of visual words they contain. In other words, if an image contains more visual words, it will have higher potential to be selected as the representative ones for a theme segment.

## 4. EXPERIMENTS

We conduct experiments to evaluate the effectiveness of the proposed VizStory system. The goal is to test if the visualized pictures generated by VizStory are able to accurately convey the plots or impressions of the fairy tales. We collect 200 stories from Andersen's Fairy Tales and Grimms' Fairy Tales. We invite five adults to estimate the familiarity of the stories. Each person is asked to annotate "familiar" or "unfamiliar" for every story. We pick 25 stories which are annotated with "familiar" by more than three persons, and pick the other 25 "unfamiliar" ones annotated by more than three persons. And then we randomly select 5 out of 25 "familiar" stories as well as 5 out of 25 "unfamiliar" ones for the evaluation. Such 10 selected stories are listed in Table 1.

Table 1. The familiar and unfamiliar fairy tales to be evaluated in the experiments.

| familiar | unfamiliar |
|---|---|
| Hansel and Gretel | Cat and Mouse in Partnership |
| Little Red Riding Hood | Gossip Wolf and the Fox |
| Sleeping Beauty | The Bell |
| The Little Match Girl | The Dog and the Sparrow |
| The Ugly Duckling | The Wise Little Girl |

We further invite another 10 persons to help evaluate the performance of VizStory. Each person is allowed to view the generated pictures of VizStory. And for each generated picture sequence, each person is also asked to find the corresponding fairy story from the total 50 fairy tales. We adopt the accuracy as the evaluation criterion by comparing the volunteer-determined results with the ground truth. The results are shown in Figure 6. For theme query, the average accuracy of the 10 testing fairy tales is 71%, in which 76% for familiar stories and 66% for unfamiliar ones. For event query, the average accuracy of the 10 testing fairy tales is 75%, in which 78% for familiar stories and 72% for unfamiliar ones. Event query has better performance than theme query. We believe it is due to that event query captures the interactions between characters while theme query lists only the characters and other story elements. Besides, the accuracy of familiar stories is higher than unfamiliar ones. We think such accuracy could result from the richness of Web resources about different fairy stories. For familiar stories, VizStory can easily find more suitable pictures. Although few images about people-unfamiliar stories in Web, our VizStory is still able to find relevant pictures conveying similar idea with 66% accuracy, to help understand the stories.

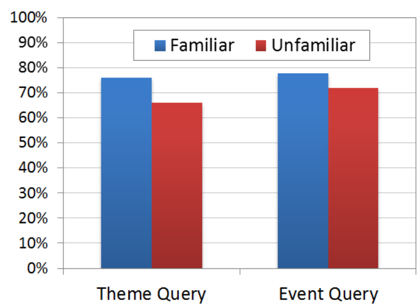


Figure 6: The accuracy of story identification.

We further show the accuracy of theme query and event query for each user, as exhibited in Figure 7 and Figure 8 respectively. For both queries on familiar and unfamiliar stories, we can find that most users can reach at least 60% accuracy, except for few individuals. Users could feel more acceptable for event query than theme query, especially for the User 3. In addition, we show the accuracy of both queries for each story, as exhibited in Figure 9. Overall speaking, the average accuracy of theme query and event query is 70% and 75% respectively. The event query has better accuracy than the theme query. All stories can have at least 60% accuracy, except for Story 10, which is “The Wise Little Girl.” Such tale states about two brothers fight for a stallion. It is very easy to be confusing because the paragraphs of the first half are all about the brothers and horses and the wise girl appear from the latter half. In short, experimental results show the visualized pictures our VizStory can effectively help users understand the ideas conveying by fairy tales (especially for using the event query), except for some stories with special narratives.

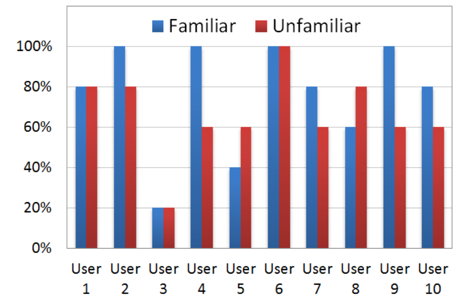


Figure 7. User accuracy of **theme query** for both familiar and unfamiliar story sets.

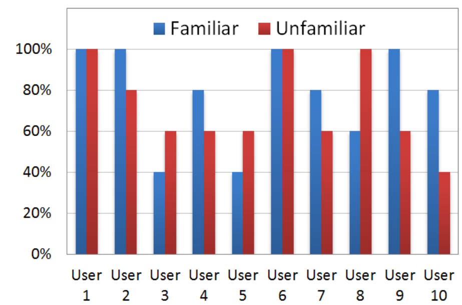


Figure 8: User accuracy of **event query** for both familiar and unfamiliar story sets.

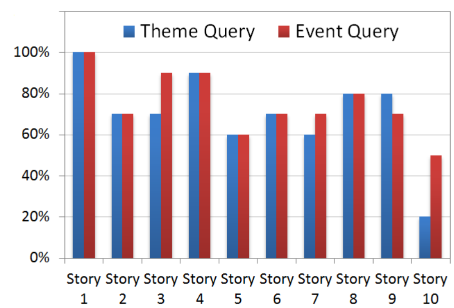


Figure 9: Story accuracy of users.

## 5. DEMONSTRATION OF VISUALIZED PICTURES

We demonstrate some visualized pictures generated by VizStory. Five pictures with the corresponding nearby story sentences are presented. The pictures are shown in Figure 10 while the corresponding sentences are listed in Table 2. We can observe our VizStory is able to successfully find the suitable picture to accompany with each story segment.

(a) (b) (c) (d) (e)

Figure 10: The visualized pictures generated by VizStory for the fairy tale Snow White.

Table 2: The corresponding nearby sentences for the pictures in Figure 10 (the story is Snow White).

| | |
|---|---|
| (a) | A long time ago, a child was born to a queen and king and she was called Snow-White. |
| (b) | Queen owned a wonderful mirror, and when she stepped before it and said: "Mirror, mirror on the wall, who is the fairest of us all?" |
| (c) | Snow-White fled into the woods where Seven little dwarfs lived. |
| (d) | The Queen discovered where Snow-White was living and disguising herself as a witch, took a poisoned apple and set out for the Dwarfs cottage. |
| (e) | The Prince took Snow-White to his palace where they were married and lived happily ever after. |

## 6. CONCLUSION

We develop a novel system, *VizStory*, to present the fairy tales with suitable pictures by considering the narrative structures and the semantic contents of stories. The central idea is to extract representative keywords from the segmented narratives and exploit Web image search to select visually-suitable pictures. Experimental results on human studies show the promising effectiveness of VizStory. Ongoing work aims to detect the turning points to enhance the understanding of the story. Besides, we are discovering the social relationships between characters and visualizing the social network to learn how the characters interact with each other.

## 7. REFERENCES


[1] M. Balabanovic, L. L. Chu, and G. J. Wolff. Storytelling with Digital Photographs. *Proc. of ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'00)*, 564-571, 2000.

[2] S. Benford, B. B. Bederson, K. P. Akesson, V. Bayon, A. Druin, P. Hansson, J. P. Hourcade, R. Ingram, H. Neale, C. O'Malley, K. T. Simsarian, D. Stanton, Y. Sundblad, and G. Taxen. Designing Storytelling Technologies to Encouraging Collaboration between Young Children. *Proc. of ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'00)*, 556-563, 2000.

[3] A. Z. Broder, D. Carmel, M. Herscovici, A. Soffer, and J. Zien. Efficient Query Evaluation using a Two-level Retrieval Process. *Proc. of ACM International Conference on Information and Knowledge Management (CIKM'03)*, 426-434, 2003.

[4] A. Druin. The Role of Children in the Design of New Technology. *Behaviour and Information Technology*, Vol. 21, No. 1, 1-25, 2002.

[5] Y. Itabashi, and Y. Masunaga. Correlating Scenes as Series of Document Sentences with Images – Taking The Tale of Genji as an Example. *Proc. of IEEE International Conference on Data Engineering Workshops (ICDE'05)*, 1235, 2005.

[6] D. Joshi, J. Z. Wang, and J. Li. The Story Picturing Engine – A System for Automatic Text Illustration. *ACM Transaction on Multimedia Computing, Communications and Applications (TOMCCAP)*, 2(1), 68-89, 2006.

[7] H. Li, J. Tang, G. Li, and T. S. Chua. Word2Image: Towards Visual Interpreting of Words. *Proc. of ACM International Conference on Multimedia (MM'08)*, 813-816, 2008.

[8] D. Lowe. Distinctive Image Features from Scale-invariant Keypoints. *International Journal of Computer Vision*, 60(2), 91-110, 2004.

[9] R. Mihalcea and C. W. Leong. Toward Communicating Simple Sentences Using Pictorial Representations. *Machine Translation*, Vol. 22, No. 3, 153-173, 2008.

[10] K. E. Steiner and T. G. Moher. Graphic StoryWriter: An Interactive Environment for Emergent Storytelling. *Proc. of ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'92)*, 357-364, 1992.

[11] N. UzZaman, J. P. Bigham, and J. F. Allen. Multimodal Summarization of Complex Sentences. *Proc. of ACM International Conference on Intelligent User Interfaces (IUI'11)*, 43-52, 2011.

[12] I. H. Witten, G. W. Paynter, E. Frank, C. Gutwin, and C. G. Nevil-Manning, Kea: Practical Automatic Keyphrase Extraction. *Proc. of ACM International Conference on Digital Libraries*, 129-152, 1999.

[13] X. Zhu, A. B. Goldberg, M. Eldawy, C. R. Dyer, and B. Strock. A Text-to-Picture Synthesis System for Augmenting Communication. *Proc. of AAAI Conference on Artificial Intelligence (AAAI'07)*, 1590-1595, 2007.

[14] F. Sebastiani. Machine Learning in Automated Text Categorization. *ACM Computing Surveys (CSUR)*, 34(1), 1-47, 2002.

[15] I. Rish. An Empirical Study of the Naive Bayes Classifier. *Proc. of IJCAI Workshop on Empirical Methods in Artificial Intelligence*, 2001.

[16] S. Osiński, D. Weiss. A Concept-Driven Algorithm for Clustering Search Results. *IEEE Intelligent System*, 20(3), 2005.

[17] M. de Marneffe, B. MacCartney, and C. D. Manning, Generating Typed Dependency Parses from Phrase Structure Parses. *Proc. of the International Conference on Language Resources and Evaluation (LREC'06)*, 2006.

[18] 陳正治，童話寫作研究，五南圖書出版公司，1990。

[19] 蔡尚志，童話創作的原理與技巧，五南圖書出版公司，1996。