NLP Story Maker

Takako Aikawa, Lee Schwartz, Michel Pahud

Microsoft Research
One Microsoft Way, Redmond, WA 98052, USA
{takakoa, leesc, mpahud}@microsoft.com

Abstract

This paper explores a novel approach to linking Graphics and Natural Language Processing (NLP). Our tool, Story Maker, lets users illustrate their stories on the fly, as they enter them on the computer in natural language. Our goals in creating Story Maker are twofold: to explore the use of NLP in the dynamic generation of animated scenes, and to explore ways to exploit users' input in order to obviate the necessity of having a large database of graphics. With our NLP technology, users can input unrestricted natural language. Story Maker provides users with direct visual output in response to their natural language input. The tool can potentially impact both the way we interact with computers and the way we compose text.

1. Introduction

Story Maker was originally invented to motivate children to write stories using the computer. We wanted to create an environment in which children would enjoy writing stories and thereby enhance their reading and writing abilities. The tool was intended to make unnecessary the distracting chore of searching for just the right picture to illustrate a story.[1] Story Maker is fun to use. It provides children with instant gratification, while encouraging them to read and write.

The novelty of our tool is in its linking of Graphics and NLP. Under our approach, natural language input is analyzed by our NLP engine (Heidorn, 1998), which passes on to the graphics component all the information necessary to render appropriate graphics, i.e., those that match the story that is being entered.

An important feature of the tool is that it can be extended by users both in terms of the number of graphics available for illustrating stories and in terms of the link between words and graphics. For instance, users can drag and drop 2D images of their family members onto the tool, associate names with these images, and have the images displayed automatically in 3D space when they use these names in a story. In this way, users personalize the graphic environment and use the vocabulary they want. Story Maker relieves the burden of having to build up a large repository of graphics before writing begins. At the same time, it addresses the problem of the unlimited nature of natural language. The integration of NLP and Graphics makes it possible to have a series of animated graphics generated dynamically based on a user's story line.

[1] Some studies have been done to tackle the problem of searching images using natural language parsers/World Knowledge database (Liberman et al. (2001), Lieberman and Liu (2002), among other). The work presented in this paper, however, does not address this issue.

The organization of this paper is as follows: Section 2 provides a brief description of Story Maker. Section 3 presents an overview of our NLP technology, where we focus on how the interface issues between Graphics and NLP are handled. Section 4 focuses on the graphics component, presenting the basic structure of the graphics component and describing the mechanism by which users can extend and customize the repository of graphics. Section 5 provides our conclusion and future directions.

2. Story Maker Architecture

Story Maker consists of two main components: (i) the NLP component and (ii) the graphics component. The user enters a story in the tool one sentence at a time. When a sentence has been completed, the NLP component analyzes it. Based on its analysis, the component passes on information that the graphics component will need to generate an appropriate animated graphic in 3D space. In this prototype, the NLP output includes information on the *actor*, *action*, *object*, *background*, etc. in the input sentence. For instance, Figure 1 is the output from the NLP component for the sentence, "the man kicked a ball on the beach."

Sentence 1: The man kicked a ball on the beach.

```
<ACTOR>man</ACTOR>
<ACTION CONTINUOUS=0>kick</ACTION>
<BACKGROUND>beach</BACKGROUND>
<OBJECT>ball</OBJECT>
```

Figure 1: The NLP output for Sentence 1

The XML-formatted output of the NLP component, as illustrated in Figure 1, becomes the input to the graphics component. From its input, the graphics component generates an animated scene that includes the appropriate graphics for the actor, object, and background specified, along with the appropriate action/behavior of the actor. Figure 2 illustrates the overall process of the tool. Figure 3 is a screenshot of the animated scene for Sentence 1 above.

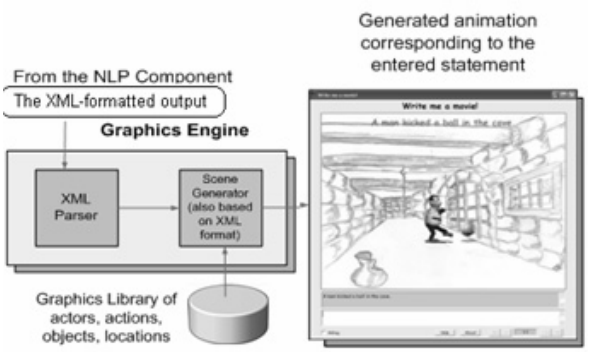

Figure 2: Tool Description


Figure 3: Screenshot of the animated scene for Sentence 1

3. NLP Component

3.1. Overview

The NLP component "understands" the basic semantic structure of a given sentence, including WHO (subject) DID WHAT (verb) to WHAT (object) WHERE (location). For instance, if a user types in Sentence 1 (i.e., "The man kicked a ball on the beach."), the system knows: (i) the subject = *man*; (ii) the verb = *kick*; (iii) the object = *ball*; and (iv) the location = *beach.* Figure 4 is a screenshot of the analysis that our NLP engine produces for the sentence.[2]

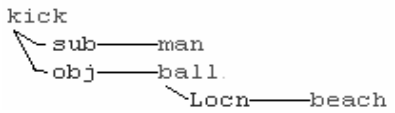

Figure 4: The NLP analysis of Sentence 1

Based on the analysis in Figure 4, the NLP component generates the XML formatted output provided in Figure 1. This becomes the input to the graphics component of Story Maker.

3.2. Word-Graphic Association

In this prototype, we started out with a small set of words, listed in Table 1, which correspond to our initial set of pre-rendered graphics.

Actors: woman, man, dragon
Actions: walk, run, jump, stumble, fly, fall, drink, kick, eat
Objects: ball, cheese, lemon, pear, hot dog, tomato, banana, chair, can
Locations: house, cave, forest, town, road, beach

Table 1: List of the pre-rendered graphics

Obviously, in unrestricted text, we are likely to find other words used for these actions, objects and scenes. Since natural language is infinite, it is impossible for us to manually associate every word that could possibly be associated with an existing graphic with that graphic. One way to address this problem, though not to solve it, is with the use of synonyms. To illustrate this point, we give sets of examples (1) and (2) below. In each set, the (a) and (b) sentences express propositions that could be illustrated in the same way. There is no need to generate different animated scenes for the (a) and (b) sentences if '*jump* and *hop*' and '*fly* and *hover*', respectively, can be considered as synonyms of one another.

(1) a. The man was *jumping* on the beach.

b. The man was *hopping* on the beach.

(2) a. The man was *flying* over the beach.

b. The man was *hovering* over the beach.

To reduce the burden of associating words with graphics, once a word has been associated with a graphic, all synonyms of that word are associated with the same graphic. As a first cut at finding groups of synonymous words, we extracted all the synonyms of the words in Table 1 from WordNet (Fellbaum, 1998). and associated each set of synonyms with the same graphic by introducing an intermediate meta lexicon (what we call "MetaLex"). For instance, since we associate the jumping graphic with the word "*jump*", we assign the MetaLex, JUMP to the synonyms (e.g., *hop*, *bounce*) of "jump" as well as to "jump" itself. Whenever possible, the XML output passed by the NLP component to the graphics component contains MetaLexes, not input words. The same principle applies to graphics for actors, objects, backgrounds, etc. In this way, we minimize the labor of associating words with graphics, while providing users with flexibility of word choice.

3.3. Various Linguistic Issues: the Power of NLP

In writing a story, it is inevitable that the writer will use pronouns (at least, in English), rather than use the same word over and over again. So, for example, you would not expect to find a well-written story proceed in the manner below, in which "*princess*" is repeated in each sentence.

Story 1: "*Once upon a time, in a sleepy little town, there lived* ***a princess****.* ***The princess*** *loved to walk in a nearby forest. One day,* ***the princess*** *saw a bird flying…..*"

The writer would undoubtedly use the pronoun "*she*" in Story 1 to refer to the princess mentioned in the first sentence. Part of the NLP component's "understanding" of natural language input is determining exactly who "*she*" refers to (i.e., anaphora resolution). If users cannot use pronouns in writing a story, they cannot write naturally. It is one of the goals of the tool to allow users to use natural language input for generating graphics. Therefore, our linguistic understanding of the text, which includes anaphora resolution, is essential to the animation task.

[2] Note that for this sentence, the attachment ambiguity of the locative "beach" is not important for successful animation of the scene.

Linguistic issues such as negation, ellipsis, etc. are also taken care of by the NLP component. For instance, if a user enters a sentence like (3), we will generate the graphics of a woman, not a man, jumping on the beach. In spite of the fact that the sentence does not explicitly state that the woman was jumping, the NLP component makes it clear that that is what she was doing. Figure 5 is a screenshot of the animated graphics of the sentence in (3).

(3) The man was not jumping on the beach but the woman was.

Figure 5: Screenshot of the animated scene based on (3)

Because of our NLP technology, Story Maker can generate animated scenes appropriate for sentences with negation, ellipsis, and pronominal reference, as well as for standard sentences of varied degrees of structural complexity. With fine-grained NLP analysis, we can allow the user a high degree of freedom of expression without burdening the graphics component.

4. Graphics Component

This section describes the graphics component of Story Maker. We first discuss the basic structure of the graphics engine and then describe how the graphics component allows users to customize/create graphics.

4.1. Basic Structure of the Graphics Engine

The graphics component consists of two modules: the XML parser module and the scene generator module. The XML output from the NLP component, as exemplified in Figure 1, is the input to the graphics component. The XML parser module reads the actor, action, object, location information and calls the scene generator module. The scene generator uses a graphic library of actors, actions, objects, locations, etc. to generate an animated scene for the current statement. In this prototype, the graphic library contains pre-rendered actors, actions, objects, locations, etc. in a 3D environment. In an advanced prototype, however, we plan to have actor and object skeletons rendered dynamically.

4.2. Customizing/Creating Graphics

One of the goals of Story Maker is to allow users to customize/create their own graphic environment. To accomplish this goal, we have pursued a variety of approaches. The first was to incorporate Pen/Ink technology into the tool. We added a simple sketch pad user interface (UI) to Story Maker so that users can create their own actors by sketching a face using the Pen/Ink technology. The second approach was to let users select their own 2D images, such as photos, so that they can expand the graphics library using their favorite images. The Sketch Pad UI enables users to do this by simply copying (i.e., drag and drop) their images onto it. Users name their new graphics as they please. Figure 7 provides a screenshot of the Sketch Pad UI in which the user sketched a face on the existing man graphic and named the resulting graphic "Toto".

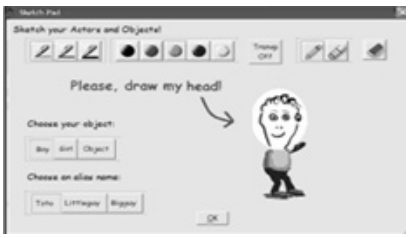

Figure 7: A Screenshot of the Sketch Pad UI

When the user is finished with the Sketch Pad UI, the graphics component adds the new face to the graphics library along with the actor's name (i.e., *Toto*)[3]. When the user enters a statement that includes the new actor's name, the NLP component associates the name (*Toto*) with an identical new MetaLex, *Toto*, and sends that information to the graphics component. The graphics component automatically loads the new face and places it on the top of a default 3D body. This body executes the action specified in the user's input.[4] Figure 8 is the screenshot of Story Maker for the statement, "Toto kicks a ball in the forest".

Figure 8: Display of the actor "Toto" on Story Maker

Figure 9 shows the picture-based actor named "*Michel*" on the left, and the same actor as part of the animated scene for the statement "Michel stumbled on the road".

[3] The current system currently stores the new face with the corresponding actor's name. We cannot have 2 actors with the same name. However, our next version will be able to support having several actors with the same name.

[4] In this prototype, we only have a default body for the 'man' and 'woman' actors but in the future the tool will be able to load an unlimited number of actors, objects, locations, etc.

Figure 9: Creation of the picture-based actor on Story Maker

Users can not only add a new actor, but they can add new objects as well. Whenever NLP passes to graphics a word that is not associated with a MetaLex, the Sketch Pad UI can open to allow the user to draw or import a picture of the object, name it, and enter synonyms for it. When the name is entered, NLP also proceeds to dynamically add the same MetaLex for all the synonyms it can find for that name.

5. Conclusion

Story Maker makes the animating of natural language easy and fun. It is a powerful tool that allows natural language to guide graphic presentation. Without NLP technology, it would be very difficult for a user to have a series of animated scenes generated automatically and directly from a story that s/he enters on the computer in unrestricted natural language. The novelty and the power of the tool lie in the fact that the two technologies, Graphics and NLP, coexist and collaborate in the same application.

The current prototype has much room for growth. Currently, NLP is not passing information about attributes of objects (e.g., 'a *big*/*red*/*small*/etc. car') to the graphics component. However, the linguistic analysis contains such information, and much more. Similarly, it only passes information about action verbs to the graphics component. We plan to enable the tool to handle sentences that don't express an action, but a change of state (e.g., the man *became fat*/*become taller*/etc.). NLP already contains information on the nature of the verb; it is just a matter of equipping the graphic component to handle such information. In addition, we will work on the translation of prepositions (e.g. *on*/*under*/*besides*) into graphics (e.g., the man put the book under/on/besides the table.).[5] Of course, understanding and representing the semantics of spatial expressions in natural language are very difficult problems. Initially, we will only be able to pass along fairly simple spatial (and temporal) information to the graphics component

From the point of view of the graphics component, we would like to allow users to animate a stick figure (most likely in 2D for easy use), or perhaps a robot, when a particular action/behavior requested from users is not in the library. The animation described by the user on the stick figure would then be added to the behavior library and applied to all actors in future stories. Another interesting extension for Pen/Ink technology would be to allow users to draw the actor body (in 2D) over a stick figure template on which each body part can be easily recognized by the graphics engine and animated. Of course, no matter how a new action is added, the user would be able to name the action, add his own synonyms for that name, and have the NLP system automatically extract synonyms as well.

With our plans to enable users to customize/create their own graphics and name them, the tool will increase in power. We also plan to integrate speech technology into the tool so that users can tell and hear their stories while seeing them on the tool. We see great potential in Story Maker, not only for linking Graphics and NLP but also for integrating technology from different fields into one platform.

ACKNOWLEDGMENTS

We thank the people in the NLP group for their feedback and support; especially, Lucy Vanderwende, Deborah Coughlin, and Gary Kacmarcik.

References

Fellbaum, C. (ed.), 1998. WordNet: An Electronic Lexical Database, Cambridge, MA, MIT Press. Available at: http://www.cogsci.princeton.edu/~wn/.

Heidorn, G. E., 1988. Intelligence Writing Assistance. In Dale R., Moisl H., and Somers H. (eds.), *A Handbook of Natural Language Processing: Techniques and Applications for the Processing of Language as Text*. Marcel Dekker, New York, 181-207.

Liberman, H., Rosenzweig, E., and Singh, P. Aria (2001). An Agent for Annotating and Retrieving Images. IEEE Computer, July 2001, 57-61.

Lieberman, H., Liu, H., 2002. Adaptive Linking between Text and Photo Using Common Sense Reasoning. MIT Media Lab.

Winograd, T., 1999. A Procedural Model of Language Understanding. *In D. Cummins and R. Cummins (eds.) Minds, Brains, Computers: The Foundations of Cognitive Science*, Blackwell.

[5] See Winograd (1999) for related work.