Automatic Conversion from E-content into Animated Storytelling

Kaoru Sumi[1] and Katsumi Tanaka[12]

[1]National Institute of Information and Communications Technology,
Interactive Communication Media and Contents Group
3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0289, Japan
Kaoru@nict.go.jp
http://www2.nict.go.jp/jt/a133/indexe.html
[2] Kyoto University, Graduate School of Informatics
Yoshida Honmachi, Sakyo, Kyoto 606-8501, Japan
ktanaka@i.kyoto-u.ac.jp

Abstract. This paper describes a medium, called *Interactive e-Hon*, for helping children to understand contents from the Web. It works by transforming electronic contents into an easily understandable "storybook world." In this world, easy-to-understand contents are generated by creating 3D animations that include contents and metaphors, and by using a child-parent model with dialogue expression and a question-answering style comprehensible to children.

1 Introduction

We are awash in information flowing from the World Wide Web, newspapers, and other types of documents, yet the information is often hard to understand; laypeople, the elderly, and children find much of what is available incomprehensible. Thus far, most children have missed opportunities to use such information, because it has been prepared by adults for adults. The volume of information specifically intended for children is extremely limited, and it is still primarily adults who experience the globalizing effects of the Web and other networks. The barriers for children include difficult expressions, prerequisite background knowledge, and so on. Our goal is to remove these barriers and build bridges to facilitate children's understanding and curiosity. In this research, we are presently considering the applicability of systems for facilitating understanding in children.

This paper describes a medium, called *Interactive e-Hon*, for helping children to understand difficult contents. It works by transforming electronic contents into an easily understandable "storybook world." *Interactive e-Hon* uses animations to help children understand contents. Visual data attract a child's interest, and the use of concrete examples like metaphors facilitates understanding, because each person learns according to his or her own unique mental model [1][2], formed according to one's background. For example, if a user poses a question about something, a system that answers with a concrete example in accordance with the user's specialization would be very helpful. For users who are children, an appropriate domain might be a

storybook world. Our long-term goal is to help broaden children's intellectual curiosity [3] by broadening their world.

Attempts to transform natural language (NL) into animation began in the 1970s with SHRDLU [4], which represents a building-block world and shows animations of adding or removing blocks. In the 1980s and 1990s, HOMER [5], Put-that-there [6], AnimNL [7], and other applications, in which users operate human agents or other animated entities derived from natural language understanding, appeared. Recently, there has been research on the natural behaviour of life-like agents in interactions between users and agents. This area includes research on the gestures of an agent [8], interactive drama [9], and the emotions of an agent [10]. The main theme in this line of inquiry is the question of how to make these agents close to humans in terms of dialogicality, believability, and reliability. Our research aims to make contents easier for users to understand, not for agent humanity. WordsEye[18] is text to scene generating system including special data. Our system generates animation but not scene.

Little or no attention has been paid to media translation from contents with the goal of improving users' understanding.

2   Interactive e-Hon

Figure 1 shows the system framework for Interactive e-Hon. Interactive e-Hon transforms the NL of electronic contents into a storybook world that can answer questions and explain of the answers in a dialogue-based style, with animations and metaphors for concepts. Thus, in this storybook world, easy-to-understand contents are created by paraphrasing the original contents with a colloquial style, by creating animations that include contents and metaphors, and by using a child-parent model with dialogue expression and a question-answering style comprehensible to children.
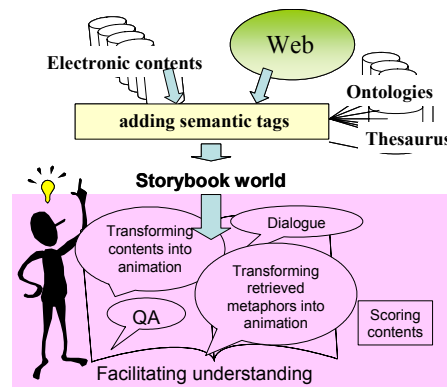

Figure 1. Interactive e-Hon: This system transforms the natural language of electronic contents into a storybook world by using animation and dialogue expression.
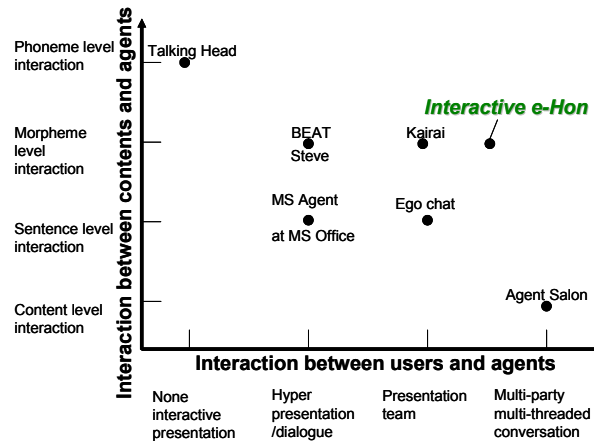
Figure 2: Interaction among users, agents and contents

Interactive e-Hon is a kind of word translation medium that provides expression through the use of 3D animation and dialogue explanation in order to help children to understand Web contents or other electronic resources, e.g., news, novels, essays, and so on.

For a given content, an animation plays in synchronization with a dialogue explanation, which is spoken by a voice synthesizer.

This processing is based on text information containing semantic tags that follow the Global Document Annotation (GDA) I tagging standard, along with other, additional semantic tags. Tags with several semantic meanings for every morpheme, such as "length," "weight," "organization," and so forth, are used. To provide normal answers, the system searches for tags according to the meaning of a question. To provide both generalized and concretized answers, after searching the tags and obtaining one normal answer, the system then generalizes or concretizes the answer by using ontologies. Recently, the Semantic Web [11] and its associated activities have adopted tagged documentation. Tagging is also expected be applied in the next generation of Web documentation.

In the following sub-sections, we describe the key aspects of Interactive e-Hon: the information presentation model, the transformation of electronic contents into dialogue expressions, the transformation of electronic contents into animations, and the expression of conceptual metaphors by animations.

I http://i-content.org/GDA

Internet authors can annotate their electronic documents with a common, standard tag set, allowing machines to automatically recognize the semantic and pragmatic structures of the documents.

2.1 Content presentation model

Our system presents agents that mediate a user's understanding through intelligent information presentation. In the proposed model, a parent agent (mother or father) and a child agent have a conversation while watching a "movie" about the contents, and the user (or users in the case of a child and parent together) watches the agents. In this model, the child agent represents the child user, and the

parent agent represents his or her parent (mother or father). For this purpose, the agents take the form of moving shadows of the parent and child. There are agents for both the user or users (avatars) and others (guides and actors), and the avatars are agentive, dialogical, and familiar [12]. Thus, we designed the system for child users to feel affinities with ages, helping them to deepen their understanding of contents. According to the classification scheme of Thomas Rist [13], a conversational setting for users and agents involves more cooperative interaction. This classification includes various style of conversation, e.g., non-interactive presentation, hyper-presentation/dialogue, presentation teams, and multi-party, multi-threaded conversation.

The horizontal axis of Figure 2 shows this classification. The vertical axis of Figure 2 shows the grain size of Interaction between contents and agents. Figure 2 shows the position of the agent related researches, in which there are Microsoft Agent at Microsoft Office, Talking Head[15], BEAT[8], Steve[10], Kairai[9], Ego chat[16], and Agent Salon[17].

With its agents for the users and for others, and with its process of media transformation from contents (e.g., question-answering, dialogue, and animation), Interactive e-Hon corresponds to between a multi-party, multi-threaded conversation and a presentation team. Interactive e-Hon has Morpheme level Interaction, because it has a close relationship with the contents being explained.

2.2 Transformation from contents into dialogue expressions

To transform contents into dialogues and animations, the system first generates a list of subjects, objects, predicates, and modifiers from the text information of a content. It also attempts to shorten and divide long and complicated sentences.

Then, by collecting these words and connecting them in a friendly, colloquial style, conversational sentences are generated. In addition, the system reduces the level of repetition for the conversational partner by changing phrases according to a thesaurus. It prepares explanations through abstraction and concretization based on ontologies, meaning that it adds explanations of background knowledge. For example, in the case of abstraction, "Antananarivo in Madagascar" can be changed into "the city of Antananarivo in the nation of Madagascar," which uses the ontologies, "Antananarivo is a city," and "Madagascar is a nation." Similarly, in the case of concretization, "woodwind" can be changed into "woodwind; for example, a clarinet, saxophone, or flute." These transformations make it easier for children to understand concepts.

In the case of abstraction, the semantic tag "person" adds the expression, "person whose name is"; "location" adds "the area of" or "the nation of"; and "organization"

Figure 3
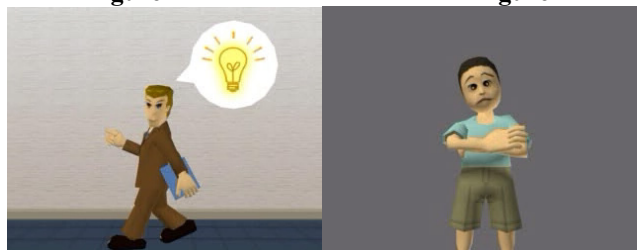
Figure 6

Figure 4

Figure 7

Figure 5

Figure 8

Figure3: *The public official sternly warned the presidents*
Figure4: *The company announced the fact of untruth*
Figure5: *It attracts the buyer's attention*
Figure6: *The company did the traffic accident hiding*
Figure7: *The Teddy bear crossed America*
Figure8: *The boy always have* a disadvantage

adds "the organization of". In the case of concretization, if a target concept includes lower-level concepts, the system employs explanations of these concepts

2.3 Transformation of contents into animations

Interactive e-Hon transforms contents into animations by using the word list described in the previous subsection. In an animation, a subject is treated as a character, and a predicate is treated as the action. An object is also treated as a character, and an associated predicate is treated as a passive action. One animation and one dialogue are generated for each list, and these are then played at the same time.

Many characters and actions have been recorded in our database. A character or action involves a one-to-many relationship. Various character names are linked to each character. Various action names are linked to each action, because often several different names indicate the same action. Actions can be shared among characters in order to prepare a commoditized framework of characters.

Representation of animation is from real motions, gestures, emphasized behaviors and cartoon like deformation. Figure 3 is from a human behavior which is close to real motions. Figure 4 is an emphasized behavior representing announcing untruth fact. The lighting fairy lamp of Figure 5 is cartoon like deformation. Figure 6 represent the traffic accident hiding by cartoon like deformation of that the man close the rolling door and hide the background of the traffic accident. Figure 7 is also cartoon like deformation by using the map. Figure 8 is representing expression of feeling from the text of *disadvantage*.

If there is a word correspondence between the name of a character and a subject or object in the list, the character is selected. If the subject or object is a company or government and other public offices, a man in suit is selected. If there is no word correspondence, in the case of the semantic tag "person," the system selects a general person character according to an ontology of characters.

The background file is selected by the order of priority of the semantic tag "location", "pur"(purpose), "sbj"(subject).

2.4 Searching and transformation of metaphors into animations

If a user does not know the meaning of a term like "president," it would be helpful to present a dialogue explaining that "a president is similar to a king in the sense of being the person who governs a nation," together with an animation of a king in a small window, as illustrated in Figure 9. People achieve understanding of unfamiliar concepts by transforming the concepts according to their own mental models [1][2]. The above example follows this process.

The dialogue explanation depends on the results of searching world-view databases. These databases describe the real world, storybooks (with which children are readily familiar), insects, flowers, stars, and so on. The world used depends on a user's curiosity, as determined from the user's input in the main menu. For example, "a company president controls a company" appears in the common world-view database, while "a king reigns over a country" appears in the world-view database for storybooks, which is the target database for the present research. The explanation of "a company president" is searched for in the storybook world-view database by utilizing synonyms from a thesaurus. Then, the system searches for "king" and obtains the explanation, "A company president, who governs a company, is similar to a king, who governs a nation." If the user asks the meaning of "company president," the system shows an animation of a king in a small window while a parent agent, voiced by the voice synthesizer, explains the meaning by expressing the results of the search process.

In terms of search priorities, the system uses the following order: (1) complete correspondence of an object and a predicate; (2) correspondence of an object and a predicate, including synonyms; (3) correspondence of a predicate; and (4) correspondence of a predicate, including synonyms.

Commonsense computing [14] is an area of related research on describing world-views by using NL processing. In that research, world-views are transformed into networks with well-defined data, like semantic networks. A special feature of our research is that we directly apply NL with semantic tags by using ontologies and a thesaurus.

For example, there is one of explanations of *a house servant,* in which "a house servant is similar to a pupil of a wizard in the sense of obeying someone" together with an animation of that in a small window. There is one of explanations of *a bear,* in which "a bear is similar to a wolf in the sense of attacking someone" together with an animation of that in a small window.

3 Application to Web contents

Web contents can easily be created by anybody and made available to the public. These contents differ from publications, which are written by professional writers and edited by professional editors, in that they are not always correct or easy to understand. Because these contents may include errors and unknown words (like neologisms, slang words, and locutions), they tend to be ill-defined. In this section, we thus discuss practical problems and solutions in transforming Web contents into a storybook world.

For example, we might try to transform the actual content, "the origin of the *teddy bear*'s name," from a Web source into an animation and a dialogue (Figure 9) Web contents can easily be created by anybody and made available to the public. These contents differ from publications, which are written by professional writers and edited by professional editors, in that they are not always correct or easy to understand. Because these contents may include errors and unknown words (like neologisms, slang words, and locutions), they tend to be ill-defined. In this section, we thus discuss practical problems and solutions in transforming Web contents into a storybook world.

For example, we might try to transform the actual content, "the origin of the *teddy bear*'s name," from a Web source into an animation and a dialogue (Figure 9)

3.1 Transformation of Web contents into dialogues

As described above, the system first generates a list of subjects, objects, predicates, and modifiers from a content's text information; it then divides the sentences in the

Figure. 9. A sample view from Interactive e-Hon

In this case, e-Hon is explaining the concept of a "president" by showing an animation of a king. The mother and child agents talk about the contents. The original text information can be seen in the text area above the animation. The user can ask questions to the text area directly.

The original text:

"President Roosevelt went bear hunting and he met the dying bear in autumn of 1902. However, the President refused to shoot to death and helped the bear. With the caricature of Clifford Berryman, the occurrence was carried by Washington Post as a heartwarming story."

The following is a dialogue explanation for this example:

Parent Agent: President Roosevelt carried out to bear hunting. Then, it met with dying small bear.
Child Agent: The President is having met the small bear which is likely to die.

A real child: *What is a president, mummy? (Then, his mother operate e-Hon system by clicking the menu)*
 (Here, an animation using the retrieved metaphor is played.)
Parent agent: A president is similar to a king as a person who governs a country.(With king's animation in a small window)
A real parent: *A president is a great man, you know?*

Parent Agent: Do you know what did it carry out after this?
Child Agent: No. what did it carry out after this?
Parent Agent: the President refused to shoot to death. And, the President helped the small bear.
Child Agent: The President assisted the small bear.
Parent Agent: The occurrence was carried by Washington Post as a heartwarming story,
with the caricature of Clifford Berryman.
Child Agent: The episode was carried by News paper as a good story.

text. For example, it might generate the following lists from the long sentences shown below:

(Original Sentence 1)

"It is said that a confectioner, who read the newspaper, made a stuffed bear, found the nickname "Teddy," and named it a "Teddy bear.""

(List 1) MS: modifier of subject; S: subject; MO: modifier of object; O: object; MP: modifier of predicate; P: predicate.

- S: confectioner, MS: who read the newspaper, P: make, O: stuffed bear;
- S: confectioner, P: find, O: nickname "Teddy," MO: his;
- S: confectioner, P: name, MP: "Teddy bear";
- S: it, P: said.

(Original Sentence 2)

"But, the president refused to shoot the little bear and helped it."

(List 2)
- S: president, P: shoot, O: little bear;
- S: president, P: refuse, O: to shoot the little bear;
- S: president, P: help, O: little bear.

The system then generates dialogue lines one by one, putting them in the order (in Japanese) of a modifier of the subject, the subject, a modifier of an object, the object, a modifier of the predicate, and the predicate, according to the line units in the list. To provide the characteristics of storytelling, the system uses past tense and speaks differently depending on whether the parent agent is a mother or a father.

Sometimes the original content uses reverse conjunction, as with "but" or "however" in the following example: "but…. what do you think happens after that?"; "I can't guess. Tell me the story." In such cases, the parent and child agents speak by using questions and answers to spice up the dialogue. Also, at the ending of every scene, the system repeats the same meaning with different words by using synonyms.

3.2 Transformation of Web contents into animations

In generating an animation, the system combines separate animations of a subject as a character, an object as a passive character, and a predicate as an action, according to the line units in the list.

For example, in the case of Original Sentence 2 above, first,
- president (character) shoot (action)
- little bear (character; passive) is shot (action; passive)

are selected. After that,
- president (character) refuse (action)

is selected. Finally,
- president (character) help (action)
- little bear (character; passive) is helped (action; passive)

are selected.

This articulation of animation is used only for verbs with clear actions. For example, the be-verb and certain common expressions, such as "come from" and "said to be" in

English, cannot be expressed. Because there are so many expressions like these, the system does not register verbs for such expressions as potential candidates for animations.

3.3 Handling errors and unknown words

One problem that Interactive e-Hon must handle is dealing with errors and unknown words from Web contents, such as neologisms, slang words, locutions, and new manners of speaking. The text area in the system shows original sentences. Erroneous words and unknown words are thus shown there, but they are exempt from concept explanation by metaphor expression.

In generating dialogue expressions using such words, the resulting dialogues and animations may be strange because of misunderstood modification. In the case of a subject or predicate error, an animation cannot be generated. In the Interactive e-Hon system, if an animation is not generated, the previous animation continues to loop, so errors may prevent the animation from changing to match the expressions in a dialogue. If both the animation and the dialogue work strangely, the text area helps the user to guess the original meaning and the reason for the problem. In addition, new or unknown words can be registered in the NL dictionary, the animation library, and the ontologies.

In fact, our example of "the origin of the teddy bear's name" from the Web may exhibit some errors in Japanese, such as the equivalent of "Teodore Roosevelt" or "Othedore Roosevelt". In such cases, since the original text is shown in the text area, and most of the variant words corresponding to "Roosevelt" are related to "the president," this was not a big problem.

4 Experiment using subjects

We conducted an experiment using real subjects to examine whether Interactive e-Hon's expression of dialogue and animation was helpful for users. We again used the example of "the origin of the *teddy bear*'s name.". Three types of content were presented to users and evaluated by them: the original content read by a voice synthesizer (content 1), a dialogue generated by Interactive e-Hon and read by a voice synthesizer (content 2), and a dialogue with animation generated by Interactive e-Hon and read by a voice synthesizer (content 3). The subjects were Miss T and Miss S, both in their 20s; child K, five years old; and child Y, three years old. Both women understood content 2 as a dialogue but found content 1 easier to understand because of its compaction. They also thought content 3 was easier to understand than content 2 because of its animation. T, however, liked content 1 the best, while S favored content 3. As T commented, "Content 1 is the easiest to understand, though content 3 is the most impressive." In contrast, S commented, "Content 3 is impressive even if I don't hear it in earnest. Content 1 is familiar to me like TV or radio." She also noted, "The animations are impressive. I think the dialogues are friendly and may be easy for children to understand."

K, who is five years old, said that he did not understand content 1. He first felt that he understood content 2 a little bit, but he did not express his own words about it. He found content 3, however, entirely different from the others, because he felt that he understood it, and he understood the difficult word *kobamu* in Japanese, which means "refuse" in English. Child Y, who is three years old, showed no recognition of contents 1 and 2, but he seemed to understand content 3 very well, as he was able to give his thoughts on the content by asking (about President Roosevelt), "Is he kind?".

In this experiment, we observed that there was a difference between the results for adults and children, despite the limited number and age range of the subjects. At first, we thought that all users would find it easiest to understand content 3 and would like it and be attracted by it. In fact, the results were different.

We assume that contents that are within a user's background knowledge are easier to understand by regular reading, as in the case of the adults in this experiment. In contrast, for contents outside a user's background knowledge, animation is expected to be very helpful for understanding, as in the case of the children. Further experiments may show that for a given user, difficult contents outside the user's background knowledge can be unde rstood through animation, regardless of the user's age.

5 Discussion

Interactive e-Hon's method of expression through dialogue and animation is based on NL processing of Web contents. For dialogue expression, the system generates a plausible, colloquial style that is easy to understand, by shortening a long sentence and extracting a subject, objects, a predicate, and modifiers from it. For animation expression, the system generates a helpful animation by connecting individual animations selected for the subject, objects, and predicate. The result is expression through dialogue with animation that can support a child user's understanding, as demonstrated by the above experiment using real subjects.

In the process of registering character data and corresponding words, or an action and its corresponding words, which are one-to-many correspondences, certain groups of words that are like new synonyms are generated via the 3D contents. These groups of synonyms are different from NL synonyms, and new relationships between words can be observed. This can be considered for a potential application as a more practical thesaurus based on 3D contents, as opposed to an NL thesaurus.

Reference terms (e.g., "it," "that," "this," etc.) and verbal omission of a subject, which are open problems in NL processing (NLP), still remain as problems in our system. As a tentative solution, we manually embedded word references in the GDA tags. A fully automatic process knowing which words to reference will depend upon further progress in NLP.

As for the process of transforming dialogues, Interactive e-Hon generates all explanations of locations, people, and other concepts by using ontologies, but granular unification of the ontologies and user adaptations should be considered from the perspective of determining the best solution for a given user's understanding.

6 Conclusion

We have introduced Interactive-e-Hon, a system for facilitating children's understanding of electronic contents by transforming them into a "storybook world." We have conducted media transformation of actual Web contents, and demonstrated the effectiveness of this approach via an experiment using real subjects. We have thus shown that Interactive e-Hon can generate satisfactory explanations of concepts through both animations and dialogues that can be readily understood by children.

Interactive e-Hon could be widely applied as an assistant to support the understanding of difficult contents or concepts by various kinds of people with different background knowledge, such as the elderly, people from different regions or cultures, or laypeople in a difficult field,.

As future work, we will consider expanding the databases of animations and words, and applying Interactive e-Hon to several other kinds of contents.

References

1. Philip N. Johnson-Laird: Mental Models, Cambridge: Cambridge University Press. Cambridge, Mass.: Harvard University Press (1983).

2. D. A. Norman, *The Psychology of Everyday Things*, Basic Books (1988).

3. Hatano and Inagaki: Intellectual Curiosity, Cyuko Shinsho (in Japanese) (1973).

4. Terry Winograd, Understanding Natural Language, Academic Press (1972).

5. Vere, S. and Bickmore, T: A basic agent. Computational Intelligence, 6:41-60 (1990).

6. Richard A. Bolt: "Put-that-there": Voice and gesture at the graphics interface, International Conference on Computer Graphics and Interactive Techniques archive, Proceedings of the 7th annual conference on Computer graphics and interactive techniques, ACM Press (1980).

7. N. Badler, C. Phillips, and B. Webber, Simulating Humans: Computer Graphics, Animation and Control. Oxford University Press (1993).

8. Justine Cassel, Hannes Hogni Vilhjalmsson and Timothy Bickmore: BEAT: the Behavior Expression Animation Toolkit, Life-Like Characters, Helmet Prendinger and Mitsuru Ishizuka Eds., pp. 163-187, Springer (2004).

9. Hozumi Tanaka et al: Animated Agents Capable of Understanding Natural Language and Performing Actions, Life-Like Characters, Helmet Prendinger and Mitsuru Ishizuka Eds., pp. 163-187, Springer, (2004).

10. Stacy Marsella, Jonathan Gratch and Jeff Rickel: Expressive Behaviors for Virtual World, Life-Like Characters, Helmet Prendinger and Mitsuru Ishizuka Eds., pp. 163-187, Springer (2004).

11. D. Fensel, J. Hendler, H. Liebermann, and W. Wahlster (Eds.) *Spinning the Semantic Web*, MIT Press (2002).

12. Toyoaki Nishida, Tetsuo Kinoshita, Yasuhiko Kitamura and Kenji Mase: *Agent Technology ,* Omu Sya (in Japanese) (2002).

13. Thomas Rist, Elizabeth Andre, Stephan Baldes, Patrick Gebhard, Martin Klesen, Michael Kipp, Peter Rist and Markus Schmitt: A Review of the Development of Embodied Presentation Agent and Their Application Fields, Life-Like Characters, Helmet Prendinger and Mitsuru Ishizuka Eds., pp. 377-404, Springer (2004).

14. Hugo Liu and Push Singh: Commonsense reasoning in and over natural language. Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES-2004) (2004).