Facilitating Understanding for Children by Translating Web Contents into a Storybook

Kaoru Sumi[1] and Katsumi Tanaka[1,2]

[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 created 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 an 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 primarily it is still primarily adults who experience the globalising 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 specialisation would be very helpful. For users who are children, an appropriate domain might be a storybook world. Our

L. Bolc et al. (Eds.): IMTCI 2004, LNAI 3490, pp. 175–184, 2005.

long-term goal is to help broaden children's intellectual curiosity [3] by broadening their world [1].

Attempts to transform natural language (NL) into animation began in the 1970s with SHRDLU [4], which represents a building -block world and shows the animations of adding or removing blocks. In the 1980s and 1990s, HOMER [5], and Put-that-there [6], AnimNL [7], and other applications, in which the 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.

In contrast, our research aims to make contents easier for users to understand, regardless of agent humanity. 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. 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.

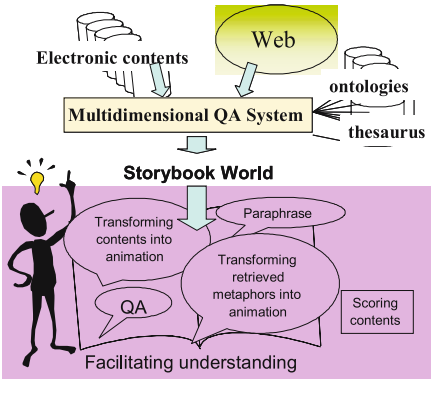

Fig. 1. Interactive e-Hon: This system transforms electronic contents into a storybook world by using animation and multidimensional QA (MQA), which includes ontologies and a thesaurus

Interactive e-Hon transforms electronic contents into a storybook world that uses dialogues to answer questions and explain concepts along with 3D animations derived from NL and based on conceptual metaphors. This system was originally designed for the Japanese language. It is based on a multidimensional question answering (MQA) system that includes ontologies and a thesaurus, and on text information with seman-

tic tags following the Global Document Annotation (GDA) [2] tagging standard, with other, additional semantic tags.

MQA create a question-answering system automatically by using digital documents that might appear anywhere. Users can receive not just the normal answer to a question but also generalised or concretised answers to a given problem; this is accomplished by using ontologies and a thesaurus. Simultaneously, users can access multiple content domains, which can be limited to a peculiar domain or not limited at all. Thus, in addition to a simple answer, a user can receive generalised or concretised answers to a given problem with the chosen domain configuration. MQA aims to transform a certain concept in a target field into a concrete example in another field.

Documents are tagged by using tags with several semantic meanings for every morpheme, such as "length", "weight", "organisation", etc. To tag electronic documents, we apply GDA, with which Internet authors can annotate their electronic documents with a common, standard tag set, allowing machines to automatically recognise the semantic and pragmatic structures of the documents[1]. To provide normal answers with MQA, the system searches for tags according to the meaning of a question. To provide generalised and concretised answers with MQA, after searching tags according to a question's meaning and obtaining one normal answer, the system then generalises/concretises the answer by using ontologies. Recently, the Semantic Web [2] and its associated activities have adopted tagged documentation. Tagging will be done 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.

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 et al. [13], a conversational setting for users and agents involves more cooperative interaction. This classification includes various style of conversation, e.g., non-interactive presentation, hyperpresentation/dialogue, presentation teams, and multi-party, multi-threaded conversation. 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

[1] http://i-content.org/GDA

e-Hon corresponds to between a multi-party, multi-threaded conversation and a presentation team. As for considering the grain size of Interaction between contents and agents, 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

For transforming contents into dialogues and animations, we first make a list which that includes subjects, objects, predicates, and modifiers from the text information of content. It also means to shorten and divide the long and complicated sentences.

Then, by collecting these words and connecting them in a friendly, colloquial style, conversational sentences are made. In addition, the system prepares repetition by the conversational partner by changing phrases to a thesaurus. It prepares explanations through abstraction and concretisation using based on ontologiesy, which that it adds explanations of background knowledge. For example, in the case of abstraction, "Antananarivo in Madagascar" can be changed into "the area of Antananarivo in the nation of Madagascar", which uses ontologiesy "Antananarivo is the an area" and "Madagascar is the a nation". Similarly, in the case of concretisation, "woodwind" can be changed into "woodwind; for example, a clarinet, saxophone, and or flute." these transformations make it is easy to understand the concepts.

In the case of abstraction, our semantic tag "person" adds the expression, "person whose name is"; "location" adds the expression "the area of" or "the nation of"; "organization" adds "the organization of". In the case of concretisation, if a target concept includes lower concepts, the system uses explanations of the lower concepts.

2.3 Transformation of Contents into Animations

Interactive e-Hon transforms contents into animations by using the list, which described in the previous sub-section. A subject is treated as a character, and a predicate is treated as the passive action in an animation. One animation and one dialogue are generated by each list, and these are then played at the same time.

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

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 there is not a word correspondence, in the case of having a semantic tag of "person", the system selects a general person character by following an ontology of characters. When there is no semantic tag of "person", the system selects a general object by following an ontology of characters.

2.4 Searching and Transformation of Metaphors into Animations

If a user does not know the meaning of the term, "president", it would be helpful to show a dialogue explanation 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. People understand

unfamiliar concepts by transforming the concepts according to their own mental models [1][2]. The above example follows this process.

The explanation is shown according to the results of searching world-view databases. The process of searching is used in the explaining of the dialogue.

The system's world-view databases described the real world, storybooks (which children have in common), insects, flowers, stars, etc. Which world should be used depends on a user's curiosity, which is acquired from the user's input in the main menu. For example, there are "a company president controls a company" and "a bear attacks a man" in the common world-view database, and "a king reigns over the a country" and "a wolf attacked on Little Red Riding-Hood" 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 the storybook world-view database by including the synonyms from a thesaurus. Then, the system searchesd for "king" and obtains the explanation, "A company president who governs a company is similar to a king who governs the a nation". In the same way, the explanation of the a bear, "A bear that attacks a man is similar to the wolf that attacked Little Red Riding-Hood", is accompanied by an animation of a wolf.

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

If there is a result of (1), it uses (1). If there is no (1), it uses (2). If there is no (1) and (2), it uses (3), and so on.

Commonsense computing [14] is related research on describing world-views by using NL. In that research, world-view are transformed it into networks with well-defined data, like semantic networks. A special feature of our research is that we use NL in conjunction with ontologiesy and a thesaurus.

3 Application to Web Contents

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

For example, we might try to transform the actual contents of "the origin of the *teddy bear's* name" from the web into an animation and dialogue (Figure. 2).

In this case, e-Hon is explaining the concept of "president" by showing a king's animation. The mother and child agents talk about the contents.

3.1 Transformation of Web Contents into Dialogues

As described above, the system first makes a list of subjects, objects, predicates, and modifiers from the content's text information; it then divides sentences. For example, it makes some lines of the list from a the long sentence as below:

Fig. 2. A sample view from Interactive e-Hon. (adapted from the original Japanese version)

(Original sentence 1)

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

(List 1) MS; a modifier of a subject, S; a subject, MO; a modifier of an object, O; an object, MP; a modifier of a predicate, P; a predicate

– S: confectioner, MS: who read it, P: make, O: stuffed bear,
– S: confectioner, P: find, O: nickname "Teddy", MO: his
– S: confectioner, P: name, MP: "Teddy bear", name
– 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 a dialogue lines one by one, putting them in the order (in Japanese) of a modifier of a subject, a subject, a modifier of an object, an object, a modifier of a predicate, and a predicate, according to the line units of line in the list. To 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 contents uses reverse conjunction, as with "but" or "however" in the following examples: "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 the case of an animation, the system combines 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 shown above, first,

– president (character) shoot (action)

– little bear (character; passive) is shot (action; passive)

are selected. After that,

– president (character) refused (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 action. 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 for 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 errors and unknown words from Web contents, such as a newly coined words, slang, locutions, and new manners of speaking. The text area in the system shows original sentences. Erroneousr words and unknown words are originally shown in the text area, but they are exempt from concept explanation by metaphor expression.

In generating dialogue expressions using such words, the resulting dialogues and animation may be generated strange because of misunderstood modification. In the case of a subject or predicate error, an animation cannot be generated. In this 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. In addition, new unknown words can be registered in the NL dictionary, animation library, and ontologies.

In fact, our example of "the origin of the *teddy bear*'s name" from the Web may have some errors in Japanese, such as "Teodore Roosevelt" or "Othedore Roosevelt". In such cases, since the original text is shown in the system's text area, and most rephrasing words which mean "Roosevelt" are "the president", this was not a big problem.

4 Experiment Using Subjects

We conducted an experiment using real subjects for 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 contents were presented to users and evaluated by them: the original content read by a voice synthesiser

(content 1), a dialogues generated by Interactive e-Hon and read by a voice synthesiser (content 2), and a dialogues with animation generated by Interactive e-Hon and read by a voice synthesiser (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 animations. T, however, liked content 1 the best, while S favoured 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 can 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 will would find it easiest to understand content 3 and would like it and be attracted 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, the contents outside a user's background knowledge, animations 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 understood through animation, regardless of the user's age.

5 Evaluation

Interactive e-Hon's expressions of dialogue and animation are generated from NL processing of the Web contents. For dialogue expression, it 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 selected animations for a subject, objects, and a predicate. The result is an expression of 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 of based on 3D contents, as opposed to an NL thesaurus.

Reference terms, (e.g., “it”, “that”, “this”), and verbal omission of an subject, which are open problems in natural language processing (NLP), still remained 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 on further progress in NLP.

As for the process of transforming dialogues, Interactive e-Hon adds all explanations of locations, people, and other concepts by using ontologiesy, but granular unification of the ontologiesy and user adaptations should be considered from the perspective of which is the best solution for a 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 shown that Interactive e-Hon can generated 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 who with different background knowledge, such as the elderly, people from different regions or cultures, or layman 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. Johnson-Laird, P. N.: Mental Models, Cambridge: Cambridge University Press. Cambridge, Mass.: Harvard University Press (1983)
2. Norman, D. A.: The Psychology of Everyday Things, Basic Books (1988)
3. Hatano, I.: Intellectual Curiosity, Cyuko Shinsho (in Japanese) (1973)
4. Winograd, T.: Understanding Natural Language, Academic Press (1972)
5. Vere, S., Bickmore, T.: A basic agent. Computational Intelligence, 6:41–60 (1990)
6. Bolt, R. A.: “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. Badler, N., Phillips, C., Webber, B.: Simulating Humans: Computer Graphics, Animation and Control. Oxford University Press (1993)
8. Cassel, J., Vilhjalmsson, H. H., Bickmore, T.: BEAT: the Behavior Expression Animation Toolkit, Life-Like Characters, Helmet Prendinger and Mitsuru Ishizuka Eds., pp. 163–187, Springer (2004)
9. Tanaka, H., 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. Marsella, S., Gratch, J., Rickel, J.: Expressive Behaviors for Virtual World, Life-Like Characters, Helmet Prendinger and Mitsuru Ishizuka Eds., pp. 163–187, Springer (2004)

11. Fensel, D., Hendler, J., Liebermann, H., Wahlster, W. (Eds.): Spinning the Semantic Web, MIT Press (2002)
12. Nishida, T., Kinoshita, T., Kitamura, Y., Mase, K.: Agent Technology, Omu Sya (in Japanese) (2002)
13. Rist, T., Andre, E., Baldes, S., Gebhard, P., Klesen, M., Kipp, M., Rist P., Schmitt, M.: 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. Liu, H., Singh, P.: Commonsense reasoning in and over natural language. Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES-2004) (2004)