A Proposal of 3-dimensional Self-organizing Memory and its Application to Knowledge Extraction from Natural Language

Kai Sakakibara, Masafumi Hagiwara
Department of Science and Technology, Keio University
3–14–1 Hiyoshi, Kohoku, Yokohama, 223–8522, Japan
E-mail: {sakaki, hagiwara}@soft.ics.keio.ac.jp

***Abstract*—**
In this paper, we propose a 3-dimensional self-organizing memory and describe its application to knowledge extraction from natural language. First, the proposed system extracts a relation between words by JUMAN (morpheme analysis system) and KNP (syntax analysis system), and stores it in short-term memory. In the short-term memory, the relations are attenuated with the passage of processing. However, the relations with high frequency of appearance are stored in the long-term memory without attenuation. The relations in the long-term memory are placed to the proposed 3-dimensional self-organizing memory. We used a new learning algorithm called "Potential Firing" in the learning phase. In the recall phase, the proposed system recalls relational knowledge from the learned knowledge based on the input sentence. We used a new recall algorithm called "Waterfall Recall" in the recall phase. We added a function to respond to questions in natural language with "yes/no" in order to confirm the validity of proposed system by evaluating the quantity of correct answers.

***Index Terms*—knowledge representation, natural language processing, 3-dimensional self organizing memory**

I. Introduction

In recent years, the approach called ubiquitous computing becomes a hot topic in a daily life[1]. This approach proposes a computing system without awareness of user. In such society, the opportunity for communication between humans and computers will increase dramatically. Therefore, needs for natural interface has been increasing.

Various interfaces have been proposed such as command line, GUI, haptic sense, movement of eyeballs and so on[2],[3]. However, the method based on natural language is the most popular to human communications. In addition, various natural language-based interfaces are proposed[4],[5].

When it comes to dealing with natural language in a computer system, it is the most difficult to express the knowledge by language. The computer systems for languages and knowledge processing can be classified into the following three.

- Systems dealing with language grammatically
  The systems of this category aim to analyze natural language grammatically and to decompose it correctly[6]-[8]. This is a very important field because computer systems can't deal with natural language directly. However, meaning understandings are seldom treated.
- Systems with artificial top-down architecture
  The systems of this category deal with natural language with artificial architecture[9],[10]. One of the advantages of such systems is easiness of handling. However, it is questionable whether human brains really process natural language based on such an architecture. On the other hand, it is said that artificial architecture is difficult to respond to unknown input.
- Systems with network and self-organizing architecture
  The systems of this category aim to extract knowledge from natural language with self-organizing network[11]-[14]. These systems are able to extract knowledge without assumption of knowledge like dictionaries. However, it is difficult to deal with natural language directly because of transformation for the network architecture.

Based on the above points, we propose a 3-dimensional self-organizing memory and apply it to knowledge extraction from natural language. The proposed system performs knowledge extraction with imitation of human brains and learns it in self-organizing manner. Therefore, the proposed system can deal with natural language directly and extract knowledge without assumption of knowledge.

This paper is organized as follows. Section II describes details of the proposed system. Section III presents descriptions of experiment and its result. Finally, Section IV concludes the paper.

II. Proposed system

A. Overview

The proposed system consists of three parts: the natural language processing part which decomposes input natural languages into links between words; the knowledge extraction part which extracts important links from output of natural language processing part as knowledge; the knowledge construction part which allocates extracted knowledge on the 3-dimensional self-organizing memory appropriately.

First, the natural language sentences inputted into the system are decomposed into simple directed links between words (word links) by the natural language processing part. In this part, we used JUMAN (morpheme analysis system) [15] and KNP (syntax analysis system) [16] that is provided by Language Media Laboratory in Kyoto University.

Next knowledge extraction part performs choice of important word links. In the knowledge extraction part, word links are stored in the short-term memory with certain weights, and the weights are attenuated with the passage of process. And, only word links which appeared in the document repeatedly certain times and span are transferred to the long-term memory.

The knowledge at the long-term memory is copied into the knowledge construction part. In the knowledge construction part, knowledge is allocated with significant location. In the recall phase, input sentences are decomposed into word links same as in the learning phase. The neurons with completely same knowledge or knowledge with common portion are fired. These neurons influence other neurons based on their importance. In this way, fire of neurons is propagated. Finally, the knowledge learned at fired neurons becomes output.

Detailed explanation on a knowledge extraction part and a knowledge construction part is followings.

B. Knowledge extraction part

The knowledge extraction part consists of short-term memory and long-term memory. Here, we define "knowledge" as a word link decomposed at the natural language processing part. For example, a sentence "Birds fly." is decomposed into the word link "bird - fly". In the following part, the details of short-term memory and long-term memory are described.

1) Short-term memory: The word links decomposed at natural language processing part are stored in the short-term memory with default weight value W_0. The weights W of short-term memory are attenuated by decay rate d with each new input. Therefore, the weight $W(t)$ at time t of word link inputted at time t_0 is

$$W(t) = W_0 d^{(t-t_0)}. \tag{1}$$

The word links are deleted when their weights are attenuated below the minimum threshold W_{min}. And, if the learned word links appear before deletion, default weight value W_0 is added to their weights. So, the relation between previous weight $W(t)$ and new weight $W(t+1)$ in this case is

$$W(t+1) = W(t) + W_0. \tag{2}$$

These weights of the word links that exceed the max threshold W_{max} are stored at the long-term memory as the knowledge.

2) Long-term memory: At the long-term memory, if a weight of certain word link at short-term memory W_s exceeds its weight at the long-term memory W_l, the weight of long-term memory is updated. And, weights at long-term memory are not attenuated.

On the other hand, frequency of appearance of knowledge with mth appearance is counted as parameter I_m. The parameter I_m is joined to its weight W_m and they become two dimensional vector $\boldsymbol{v}_m$. In the followings, we use this vector $\boldsymbol{v}_m$ as a scale of importance of knowledge.

C. Knowledge construction part

In the knowledge construction part, the knowledge stored at the long-term memory is constructed by the 3-dimensional self-organizing memory. The 3-dimensional self-organizing memory consists of cubically placed neurons. It can represent ups and downs of knowledge importance by height and relations of knowledge by distance. The neurons hold the following three parameters.

- The learned knowledge
- The vector of knowledge importance(Code vector)
- The influence from other neurons

The height of the 3-dimensional self-organizing memory represents the importance of knowledge. Therefore, the knowledge locates in the high place is considered as important knowledge.

As the algorithm of winner neuron selection, we used a new algorithm called "Potential Firing". For the algorithm of association, we used a new algorithm called "Waterfall Recall".

In the followings, the detail of "Potential Firing" and "Waterfall Recall" are described.

1) Potential Firing: Potential Firing is the algorithm to place knowledge appropriately on the 3-dimensional self-organizing memory. The purposes are the followings.

- Represent importance of knowledge by height
- Similar knowledge placed close

In Potential Firing, the potential values of every neuron are calculated by input knowledge. Based on this value, a winner neuron is determined. The way to calculate potential value is the following.

The input that the 3-dimensional self-organizing memory receives is the following two kinds.

- Knowledge such as "bird - fly"
- The importance vectors of knowldge $\boldsymbol{v}$

By these 2 inputs, the vertical potentials are calcurated by the importance vectors and the locational potentials are decided by the locations of knowledge. These potentials are an increasing function from the center of potential. So, a neuron with the least potential learns knowledge. Figs.1-4 show example of the processes. First, vertical potential is calcurated as shown in Fig.1. The knowledge with strong connection are searched from the 3-dimensional self-organizing memory as shown in Fig.2. By these knowledge, locational potentials are decided as shown in Fig.3. Finally, potentials are overlapped and the winner neuron is determined as shown in Fig.4.

The way to decide vertical potential is based on Kohonen's self-organizing map [17],[18]. In other words, Euclidean norm between inputted vectors and code vectors. In addition, the height of the 3-dimensional self-organizing memory is devised to represent the importance of knowledge. Spread process of vertical potential are shown in Figs.5-6.

The code vectors of the 3-dimensional self-organizing memory are initialized as the following. A initial code vector $\boldsymbol{k}_{xyz0}$ at coordinate (x, y, z) is

$$\boldsymbol{k}_{xyz0} = (z + R \mid z + R). \tag{3}$$

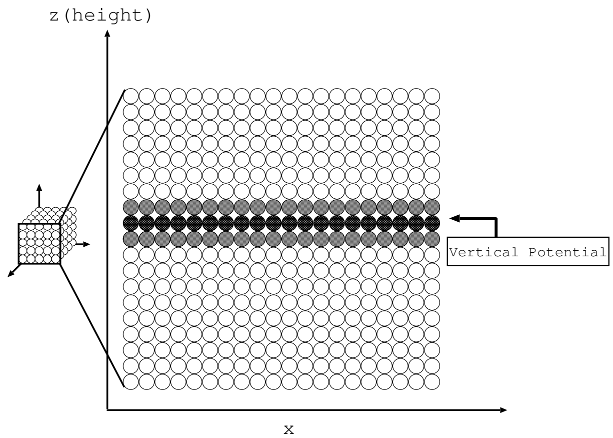

Fig. 1. An example of spread in potential field-A

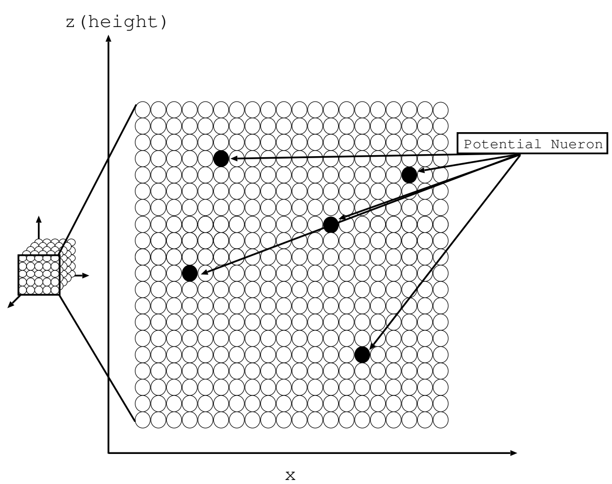

Fig. 2. An example of spread in potential field-B

Here, R is a small random number for distinguishing vectors in the same height.

According to this initialization, a vertical potential value $V(x, y, z)$ at coordinate (x, y, z) is expressed as

$$V(x, y, z) = ||\boldsymbol{v}_{in} - \boldsymbol{k}_{xyz}|| \tag{4}$$

where $\boldsymbol{v}_{in}$ is an inputted vector and $\boldsymbol{k}_{xyz}$ is a code vector at coordinate (x, y, z). By this equation, the code vectors at higher position are similar to the input vectors with large elements. As a result, the vertical potential values of higher position become lower. So, the knowledge with higher importance is placed at higher position of the 3-dimensional self-organizing memory.

In calculation of locational potential, the knowledge having common part to the input are searched on the memory. First, all knowledge having a common portion to an inputted knowledge is searched out of learned knowledge. The knowledge having a common portion is what has the same antecedent part or the same consequent part, for example “bird - fly” and “bird - flap”. The locational potential $L(x, y, z)$ at coordinate (x, y, z) is expressed as

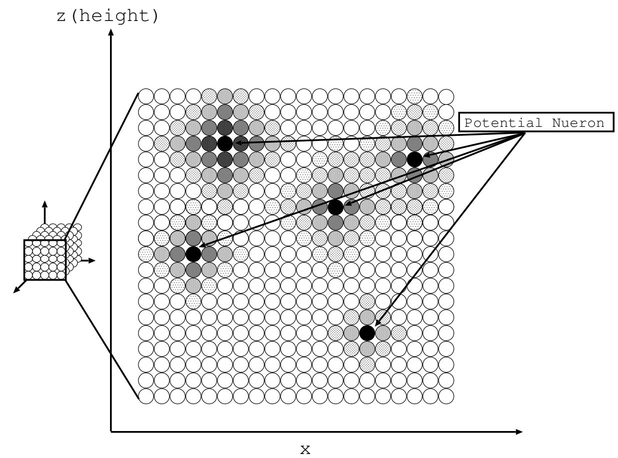

Fig. 3. An example of spread in potential field-C

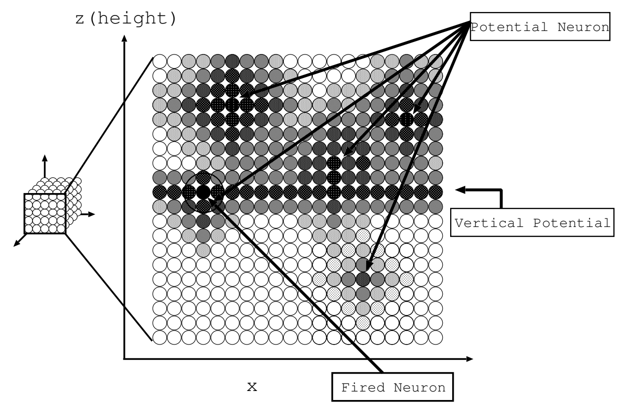

Fig. 4. An example of spread in potential field-D

$$L(x, y, z) = -\sum_{i=1}^{n}(||\boldsymbol{r_i} - \boldsymbol{r}_{xyz}|| \ \ ||\boldsymbol{k}_i||) \tag{5}$$

where n is the number of neurons learning the knowledge having a common portion to an inputted knowledge(potential neuron), $\boldsymbol{r}_i (i = 1, 2, \cdots, n)$ is its coordinate vector, $\boldsymbol{r_{xyz}}$ is the coordinate vector of a neuron at coordinate (x, y, z) and $\boldsymbol{k_i}$ is the code vector of ith potential neuron. By considering norm of $\boldsymbol{k_i}$, knowledge at higher position has strong locational potential. It is necessary to strengthen the influence of knowledge at higher position since they are more important. By the above processing, vertical potentials and locational potentials are determined.

However, with only this process, optimal placement cannot be obtained by one time learning, because placement result is influenced greatly by the order in learning.

In order to reduce the effect, a momentum term is added and iteration learning is performed. Finally, the potential value $P(x, y, z, t+1)$ of a neuron at coordinate (x, y, z) at tth learning is obtained as

$$\begin{aligned} P(x, y, z, t+1) \quad &= \quad a_1 V(x, y, z, t) + a_2 L(x, y, z, t) \\ &\quad + \alpha(t) M(\boldsymbol{r_{xyz}} - \boldsymbol{r}_{\ prev}, t) \end{aligned} \tag{6}$$

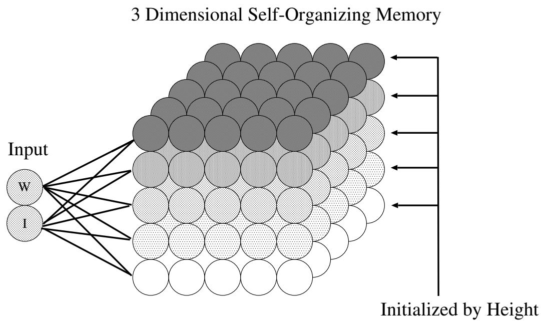

Fig. 5. An example of spread in vertical potential-A

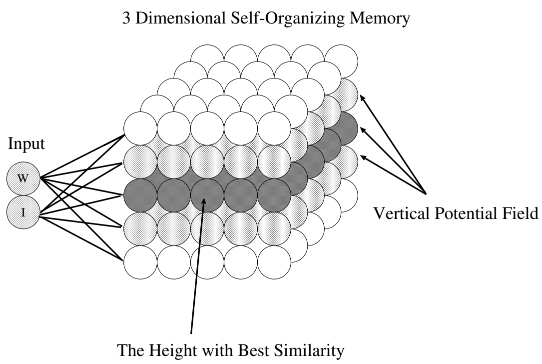

Fig. 6. An example of spread in vertical potential-B

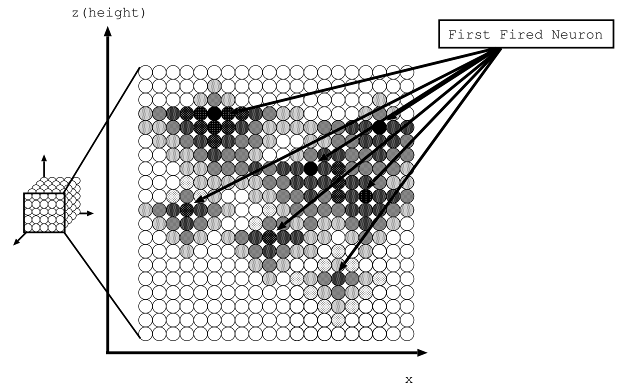

Fig. 7. An example of Waterfall Recall A

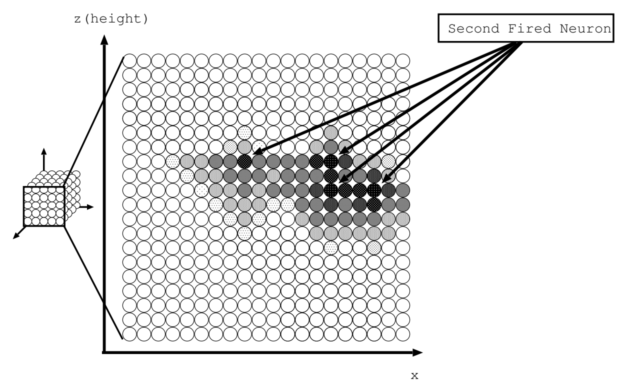

Fig. 8. An example of Waterfall Recall B

where a_1 and a_2 are the weight constants, $\alpha(t)$ is momentum factor increasing with learning times, $M(\boldsymbol{r}_{xyz} - \boldsymbol{r}_{prev}, t)$ is neighborhood function, $\boldsymbol{r}_{xyz}$ is the coordinate vector of a neuron at coordinate (x, y, z) and $\boldsymbol{r}_{prev}$ is previous coordinate vector of the knowledge. Owing to the momentum term, knowledge becomes less movable as learning goes on.

2) Waterfall Recall: Recall is performed to the inputted natural language. The inputted natural language sentences are decomposed into the word links same as in the learning phase. First, if knowledge which is completely same as the inputted knowledge is searched on the 3-dimensional self-organization memory, the neuron corresponding to the knowledge fires strongly. Second, if the knowledge with common portion to the input is found, the neurons corresponding to the fire weakly.

This attenuation function is constructed so that neurons at higher position fire strongly and neurons at lower position fire weakly. Moreover, firing toward a higher location in the 3-dimensional self-organization memory is weaker compared with firing toward lower location. The influence $E(n, f)$ of nth neuron is expressed as

$$E(n,f) = \begin{cases} \frac{z_n}{4}\exp\left(-\frac{||\boldsymbol{r}_f - \boldsymbol{r}_n||^2}{X^2+Y^2+Z^2}\right) & z_n > z_f \\ \frac{z_n}{4}\exp\left(-\frac{||\boldsymbol{r}_f - \boldsymbol{r}_n||^2}{X^2+Y^2+Z^2}\right) - d(z_n - z_f) & \\ & z_n \leq z_f \end{cases} \tag{7}$$

where $\boldsymbol{r}_f$ is the coordinate vector of fth fired neuron, X, Y, Z are the width, the depth, and the height of the 3-dimensional self-organizing memory, respectively, z_n is the height of nth neuron and z_f is the height of fth fired neuron. In short, if $z_n \leq z_f$, the influence is attenuated by attenuation constant d. This process is performed recursively, and the neurons influenced exceeding the threshold are fired.

The figures of these firing are shown in Fig.7 and Fig.8. In Fig.7, the influence of the first fired neurons is overlapped, and in Fig.8, the neurons which exceed the threshold are fired. This algorithm is called ”Waterfall Recall” because firing spreads from a top to the bottom. The sum of influence $F(n)$ of nth neuron is expressed with the following equation using $E(n, f)$.

$$F(n) = \sum_f (D(c)E(n,f)) - \theta \tag{8}$$

where $D(c)$ is the attenuate function and parameter c is the times of firing. That is, the influence of firing decreases according to the propagation.

Finally, N neurons whose sum of influence $F(n)$ is positive are fired. The parameter N is determined by user considering the size of document and range of recall.

III. Evaluation experiment

In order to confirm effectiveness of the proposed network, we implemented a postprocessor that can answer to question sentences by Yes/No. In the followings, we describe detail of the postprocessor and experimental result.

TABLE I

EXAMPLES OF QUESTIONS AND ANSWERS

Questions	Extracted knowledge	Recalled knowledge	Followed links	Ans.
Does bird fly?	bird - fly	bird - wing wing - fly bird - flap	bird - wing - fly	Yes
Is bird sparrow?	bird - sparrow	fly - airplane bird - fly flap - wing	bird - fly - airplane	No

TABLE II

LEARNING DOCUMENTS

	Words		
	celestial	earth	typhoon
number of sentences	815	2636	450
number of questions	21	16	14

TABLE III

PERCENTAGE OF CORRECT ANSWER BY EACH CRITERIA

	Words			Average
	celestial	earth	typhoon	
Criterion1	66.7	75.0	85.7	74.5
Criterion2	80.0	90.0	-	85.0

TABLE IV

EXAMPLES OF CORRECT AND INCORRECT ANSWER TO THE QUESTIONS INCLUDING THE WORD "CELESTIAL" WITH CRITERION1

Questions	Answer	Judgement
Does an astronomical observatory observe a celestial body?	Yes	○
Does a celestial body observe an astronomical observatory?	No	○
Do infrared rays emit a celestial body?	Yes	×

A. Postprocessing

The recalled knowledge are expressed by the directed word link as mentioned above. Therefore, by following the links, association becomes possible. "Following the links" means, for example, search knowledge whose antecedent part is "eagle" to the knowledge "bird - eagle".

Since the 3-dimensional self-organizing memory holds the importance of knowledge, it can limit the direction of association. Then, in this experiment, if the link could be followed in the knowledge recalled by the question sentence, the question presupposed that to be right. An simple example is shown Table I.

In the proposed system, height of the 3-dimensional self-organizing memory expresses the importance of knowledge. In view of construction of knowledge, it is appropriate to follow links from more important knowledge to less important knowledge. Therefore, only when the importance of knowledge is lower than the previous knowledge, following links is permitted. For example, if the importance of knowledge "eagle - fly" is higher than the importance of knowledge "fly - bird", link "eagle - fly - bird" is followed. But, if it is lower, the link isn't followed.

Therefore, if the question "Is eagle bird?" is asked to the system, system answers "Yes" because the links from "eagle" to "bird" are established. However, if the question "Is bird eagle?" is asked, system answers "No" because the link has directions and the links from "bird" to "eagle" are not established.

B. Experimental condition

In this experiment, we used an encyclopedia MYPEDIA'98[19] as a learning document. MYPEDIA'98 is an electronic encyclopedia that has about a million lines of text data. We extracted sentences including certain word from MYPEDIA'98. The words used in extraction and number of questions are shown in Table II.

When judging Yes/No, consequent part can be followed from the antecedent part of knowledge as mentioned above, it judges "Yes". However, if an input text is complicated, plural paths might exist. In such a case, it can be said that only a part of input knowledge can be followed correctly. So, the experiment was performed using the following two criteria, respectively.

- Criterion1
  When more than half of inputted knowledge can be followed correctly, the answer is "Yes", otherwise "No" is answered.
- Criterion2
  When all of the inputted knowledge can be followed correctly, the answer is "Yes" and when no inputted knowledge can be followed, "No" is answered.

However, when learning document does not contain concerning knowledge in both cases, the answer is "Judgment is impossible". Moreover, when two or more nouns existed in a question sentence like "Is eagle bird?", the sentence which surely replaced them was also used as an input. So in this example, the question "Is bird eagle?" is asked also. In addition, ambiguous question that answer is subjective is avoided.

C. Experimental result

The percentage of correct answer to the questions shown in Table II is summarized in Table III. The examples of correct and incorrect answer to the questions including the word "celestial" with each criteria are shown in Tables IV-V.

By the above experiment, it was confirmed that the 3-dimensional self-organizing memory can learn knowledge autonomously and can recall them appropriately.

IV. CONCLUSIONS

In this paper, we proposed the 3-dimensional self-organizing memory with a new learning algorithm and a new recall algorithm, and applied them to knowledge extraction from

TABLE V

EXAMPLES OF CORRECT AND INCORRECT ANSWER TO THE QUESTIONS INCLUDING THE WORD "CELESTIAL" WITH CRITERION2

Questions	Answer	Judgement
Is a star bright?	Yes	◯
Is the universe bright?	No	◯
Does a planet revolve around the sun?	No	×

natural language. From the result of evaluation experiment, we confirmed that the proposed system can extract and recall knowledge appropriately from natural language in documents. As a future work, application of recalled knowledge, improvement in accuracy, etc. are considered.

ACKNOWLEDGMENT

This research is partly supported by Keio University Special Grant-in-Aid for Innovative Collaborative Research Projects.

REFERENCES

[1] M. Weiser: The Computer for the Twenty-First Century, *Scientific American*, Vol.265, No.3, pp.94-104, Sep.1991.

[2] K. Tanaka and K. Fukui: "A User Interface via Agent-Based Interaction Using Gaze Input", *Trans. IEICE* Vol.J85-D-II, No.9, pp.1464-1472, Sep.2002.

[3] K. Kakusho, S. Hagiwara and M. Minoh: "Adaptive Registration through Manipulation of Virtual Objects by Augmented Reality", *Trans. IEICE*, Vol.J85-D-II, No.11, pp.1701-1713, Nov.2002.

[4] T. Hirata, H. Murakami and T.Nishida: "Support for Community Knowledge Sharing with Associative Representation and Talking-Virtualized-Egos Metaphor", *Trans. JSAI* Vol.16, No.2 I, pp.225-233, 2001.

[5] K. Matsuda, H. Ueno and T. Miyake: "Evaluation of Personal Agent-Oriented Virtual Society - PAW", *Trans. IEICE*, Vol.J82-D-II, No.10, pp.1675-1683, Oct.1999.

[6] Y. Ogawa, M. Mahsut, K. Totama and Y. Inagaki : "Derivational Grammar Approach to Morphological Analysis of Japanese Sentences", *IPSJ Journal*, Vol.40, No.3, pp.1080-1090, Mar.1999.

[7] K. Watabe, T. Uehara and K. Ishikawa: "A Sentence-Structure Collation System for EDR Japanese Corpus", emphTrans. IEICE, Vol.J84-D-II, No.1, pp.139-149, Jan.2001.

[8] H. Shinnou: "Conversion of Japanese Morphological Analysis into Classification Problem and Its Solving", *IPSJ Journal*, Vol.42, No.9, pp.2221-2228, Sep.2001.

[9] M. Harada and T. Mizuno: "Japanese Semantic Analysis System SAGE using EDR", emphTrans. JSAI, Vol.16, No.1 J, pp.85-93, 2001.

[10] K. Kosaki, Y. Kitamura, M. Ikeda and R. Mizoguchi: "Development at an Enviroment for Building Ontologies Based on a Fundamental Consideration of "Role" and "Relationship"", *Trans. JSAI*, Vol.17, No.3 B, pp.196-208, 2002.

[11] T. Tsuzuki, T. Kawahara and T. Kusumi: "Connectionist modeling of higher-level cognitive processes", Japanese Psychological Association *The Japanese Psychological Research*, Vol.72, No.6, pp.541-555, 2002.

[12] T. Omori and N. Yamagishi: "PATON: A Model of Concept Representation and Operation in Brain" , *Proc. of International Conference on Neural Network 94*, pp2227-2232, 1994

[13] N. Ikeda and M. Hagiwara: "A Proposal of Novel Knowledge Representation (Area Representation) and the Implementation by Neural Network", *Trans. IEICE*, Vol.J81-D-II, No.6, pp.1328-1335, Jun.1999.

[14] H. Tanaka, J. Huang and M. Hagiwara: "Concepts of Sentences Associative Neural Network", *Trans. IEICE*, Vol.J85-D-II, No.10, pp.1591-1601, Oct.2002.

[15] Morpheme analysis system JUMAN http://www.kc.t.u-tokyo.ac.jp/nl-resource/juman.html

[16] Syntax analysis system KNP http://www.kc.t.u-tokyo.ac.jp/nl-resource/knp.html

[17] M. Sakawa and M. Tanaka, *"The introduction of neuro-computing"*, 1st ed, Tokyo, Japan: Morikita, 1999.

[18] M. Hagiwara, *"Neural network, Fazzy, Genetic Algorithm"*, 1st ed, Tokyo, Japan: Sangyo-Tosho, 1994.

[19] Hitachi Digital Heibon-sha, *MYPEDIA'98*, Hitachi Digital Heibon-sha, 1998.