Automatic Construction of a Knowledge System Using Text Data on the Internet

Junichi Takeno, Satoru Ikemasu, Yukihiro Kato
Meiji University, School of Science and Technology, Computer Science
Higashimita 1-1-1, Tama-ku, Kawasaki-shi, 214-8571 Kanagawa, JAPAN
takeno@cs.meiji.ac.jp (juntakeno@gmail.com)

Abstract

In order to be truly useful to humans, robots should be able to communicate smoothly with them. To gain this capability, robots must possess cognitive and speech functions similar to the consciousness function of humans. It is important for robots to be capable of a flow of consciousness and emotions, and for this purpose robots must first of all possess a so-called memory of knowledge.

This paper defines a network of association and *kansei* (a Japanese term relating to emotions and feelings) values that model the memory of knowledge. It also describes the construction of an association-*kansei* database comprising the association-*kansei* networks. This database searches through text data on the Internet, statistically processes the data, automatically calculates the strength of associations between collected words, extracts emotional elements incidental to respective words, and stores the results in the database. This paper further introduces techniques used to handle idioms, separate words of different meanings and classify concepts using superordinate-subordinate relationships. Lastly, the paper describes evaluation experiments that confirm the usefulness of the constructed database.

1. Introduction

Robot technology has been advancing rapidly in recent years. Robots are being manufactured to serve various purposes. In addition to ubiquitous industrial robots, many robots have been developed to work in human living spaces. Robots are advancing into the communities of humans and coexisting with them. For this reason, robots are increasingly being required to possess the capability of communicating with humans.

To be able to communicate with humans, robots must possess a so-called memory of knowledge, which would be the source of the flow of consciousness and an information source for emotion. The authors propose an association-*kansei* network as the model for memory of knowledge. *Kansei*, a Japanese word, is equivalent to emotion and feelings in English. In this model, associations between words are networked and correlated with mental images via certain factors to collectively handle the memory of knowledge and *kansei* of humans. In our present study, the association-*kansei* network has been incorporated into a database, which we call the association-*kansei* database.

Fairly similar research is being performed at Doshisha University's Graduate School of Engineering in a project called the Commonsense Consciousness Judgment System. The theme of the project is not exactly consciousness per se, but for a given input, a database is searched for a similar conception to generate an image, which makes this a kind of expert system [1]. The primary objective of that research is to have a computer possess a degree of "common sense," and it is not particularly concerned with developing communication capability and constructing consciousness. There is also a website called KwMap. This site is a search site on the Internet and the site's method and results of collecting related words (which are nearly synonymous to what we term association words) are very similar to our approach which will be described in this paper. However, the study in the KwMap website is not relevant to feelings. It can be said that this study is a natural language processing research. However, it is unique in that the grammar is rarely used in our study.

This paper describes the association-*kansei* network and a model of "flow of consciousness," which is represented by said network. Next, the paper describes the automatic computerized construction of the association-*kansei* database that contains the association-*kansei* networks, and reports on the results of evaluation experiments that confirm the usefulness of the database. The paper further discusses how to handle idioms, a technique to separate words of different meanings (polysemic words), and hierarchization of concepts, both of which are indispensable for robots to simulate the knowledge structure of humans more closely.

2. Knowledge and Consciousness Modeling

This section describes the nature of the association-*kansei* network that models the knowledge structure of humans, and introduces the consciousness network that models consciousness.

2.1 Association-*Kansei* Network

In the association-*kansei* network, each word is a node and is connected to its associable words by directed edges. Each edge is given a value that indicates the strength of association. This value is called the association value. Words associated from other words are called association words. This network is capable of calculating the mental images it holds in the face of a subject by using *kansei*-related adjectives that are possessed by the network itself as index. Adjectives used for *kansei* calculation are described shortly. The association-*kansei* network shown in Fig. 1 is an example presented only for illustrative purposes. [2][3][4][5]

2.2 Conscious Network and Flow of Consciousness

Modeling the human flow of consciousness is an effective means to develop a robot capable of thinking like humans. Consciousness in humans is developed by perceiving and recognizing things while integrating information derived from external (via the five senses) and internal stimuli (self-desire and intense association). The authors believe that the consciousness function of humans is a function for recognizing the information derived from external and internal stimuli and integrating it into a unified, single concept. The authors also believe that the flow of consciousness is a process wherein the conscious concept moves from the self to the other dynamically upon perceiving new internal and external stimuli.

This process is explained using the association-*kansei* network as described below. Upon receiving certain stimuli, a conscious concept and its association word group are extracted as a sub-graph. The association word group is a concept associated from the conscious concept, or a sub-conscious concept. The extracted sub-graph is called a consciousness network. Upon receiving some other internal and external stimuli, the central concept moves to other concepts, generating new consciousness networks. Through the repetition of this process, consciousness networks are generated one after another, thereby generating the flow of consciousness.

3. Association-*Kansei* Database

The authors have constructed a database of association-*kansei* networks and named it the association-*kansei* (AK) database. This section describes the association words and values, *kansei* words and values, and the automatic database construction procedure using text data extracted from the Internet. It also discusses the handling of idioms (a process by which the system can approximate the knowledge structure of humans), a technique for separating polysemic words, and concept hierarchization. [3][4][5]

3.1 Association Words and Association Values

An association word is a word associated from other words. The association value is an index of the ease of association. In Fig. 1, for example, Toyota, Mercedes-Benz and Accident are the association words of the central word Car. The ease of association from the central word Car differs for each of these three association words. The ease of association is quantified by the association value.

3.2 *Kansei* Words and Values

Kansei refers to the mental images that a person has in the face of external stimuli. For example, dog-lovers have a good impression of dogs whereas those who do not like dogs can have a hatred for them. *Kansei* is considered an element that generates emotion in humans. In the present study, *kansei* is represented by adjectives related to five of six basic human emotions: Happiness, Anger, Fear, Disgust and Sadness. These adjectives are called *kansei* words. *Kansei* values refer to the quantified level of mental images humans have for various stimuli. In the above example, “refreshing” and “fearful” are the *kansei* words for Happiness and Fear, respectively (Fig. 1). [6]

The remaining sixth basic emotion of Surprise cannot be expressed by adjectives, and thus is not considered in the present study. A list of some adjectives related to respective emotions is given in the Appendix (Table 1).

3.3 Constructing a Knowledge Database Using the Internet

Construction of the knowledge database is briefly described below. Association words are extracted from text data collected from the Internet, and the association value and *kansei* value are calculated for each association word.

The authors decided to use the Internet as a data source to construct the knowledge database because of its three major merits: (1) a large volume of information may be collected with ease because the Internet is the world's largest database, (2) information is renewed daily and the latest information is accessible at all times, and (3) text data on the Internet are written by humans and thus reflect human sensibilities very well.

3.3.1 Text Extraction Process

The process of text extraction from the Internet is described below.

At first some websites were selected arbitrarily. The text data and links to other websites found on webpages were picked up and saved. The external links were followed to their linked pages, and then the text data and links to other websites found there were also picked up and saved. These operations were repeated to collect a huge amount of text data.

3.3.2 Extraction of Association Words and Values

To define association words for a given central word, we must check the correlation between the given word and other words. Any sentence written by a human was written with a presumed intention to convey certain information to other humans. Given this fact, we believe that there are certain correlations among the words that appear in any given single sentence. This may be said to be an extensive interpretation of Hebb's rule of simultaneity of stimuli. Based on this idea, we define the words appearing in the same sentence as association words and extract them from the collected text data.

The association value is calculated as the ratio of the simultaneous appearance of the central word and the association words in the same sentence to the total count of the appearance of the central word. Specifically, assume x is the central word, $c(x)$ the total count of appearance of the central word, a_i an arbitrary association word of x, and $c(x:a_i)$ the count of the simultaneous appearance of x and a_i in the same sentence, then the association value $p(x:a_i)$ from the central word x to the association word a_i is given by the following equation:

$$p(x:a_i) = \frac{c(x:a_i)}{c(x)} \quad \left(0 \le p(x:a_i) \le 1 \right) \tag{1}$$

3.3.3 Calculation of *Kansei* Values

The *kansei* value, like the association value, is calculated by how frequently a *kansei* word appears in the sentence containing the central word.

The *kansei* value is calculated in two steps. First, assume x is the central word, $c(x)$ the total count of the appearance of the central word x, and $c(x:k)$ the number of the *kansei* words simultaneously appearing with the central word x, then we perform the following calculation. The derived value is tentatively called pre-*kansei* value $K_p(x)$, because our aim is to obtain the final *kansei* value.

$$K_p(x) = \frac{c(x:k)}{c(x)} \tag{2}$$

The pre-*kansei* value can exceed unity if there are many *kansei* words. Considering that the association value ranges from zero to unity, it is convenient if we express the *kansei* value in the same zero-to-unity range. After obtaining the pre-*kansei* values for all nodes, we therefore identify the maximum pre-*kansei* value $K_{\max}$, and define the *kansei* value $K(x)$ for the central word x as the ratio of the pre-*kansei* value $K_p(x)$ of a given central word x to $K_{\max}$. This is the second step in calculating the *kansei* value.

$$K(x) = \frac{K_p(x)}{K_{\max}} \quad \left(0 \le K(x) \le 1 \right) \tag{3}$$

As mentioned before, there are five kinds of *kansei* words, related to the five human emotions of Happiness, Anger, Fear, Disgust and Sadness. This means that each node has five kinds of *kansei* values. All these values are calculated in the same manner.

3.4 Processing to Approximate the Knowledge Structure of Humans

The current database is a simplified model of the human knowledge structure. The information contained in it is insufficient for a robot to communicate with humans. We added the process of hierarchizing concepts to improve the AK network ands make it closer to the human knowledge structure.

The current idiom-generating process uses an existing idiom dictionary. An automatic idiom generation algorithm is under development.

3.4.1 Semantic Classification

Concepts and words do not necessarily have a one-to-one correspondence in languages. Idioms are an example of this. There are also many polysemic words, that is, words that have several meanings.

Semantic classification refers to the handling several meanings of a polysemic word as separate nodes on the knowledge database. For example, the word Book can mean both reading material and to make a reservation. The word Will can mean determination, as well as a written statement specifying the distribution of a deceased person’s property. Depending on the context, the word Spring can mean a coil of metal, a season of the year, or water gushing up from underground. Take another example of the word Virus. The root meaning is the same but the image and the usage of the word Virus are quite different when referring to disease-causing microbes or a contagious computer program.

Semantic classification is a process for handling words of multiple different meanings separately according to respective concepts. The final objective is to construct a database of the principle of each node with a single concept.

Clustering is used as a means of semantic classification. As a kind of data-mining technique, clustering is based on the idea that similar data behave similarly. Similar data are picked up from a group of data and collected into respective clusters. Each cluster contains similar objects, and objects of different attributes are collected in different clusters to the extent possible.

Hierarchical clustering analysis is employed in our present study.

3.4.1.1 Hierarchical Clustering

Hierarchical clustering is an approach to grouping together objects that are "close" to one another sequentially. The hierarchical clustering approach is described below.

1) There are initially n objects $O_1, O_2 ... O_n$, each of which belongs to its own cluster, that is, there are n clusters in all.
2) We calculate the value d_{ij} representing the level of similarity between arbitrary objects O_i and O_j according to an arbitrary criterion. The pair of objects with the highest similarity index value is put into a new single cluster. There are n-1 clusters at this stage.

The above process is repeated until all data are finally collected in a single cluster (Fig. 2). In hierarchical cluster analysis, the sequence of agglomeration is shown graphically as a dendrogram (Fig. 3). The objective of hierarchical cluster analysis is to sort the data and generate the dendrogram. The dendrogram clearly shows how clusters are formed and which pairs of elements are closely related to other pairs.

3.4.1.2 Semantic Classification of Association Words

Semantic classification is a process of dividing a node that has different images (a polysemic word node), which is treated as a single node or a word in the database, according to its different meanings. After being divided, association words are newly grouped for each of the divided nodes. Specifically, we calculate the similarity indices of the association words, and group them according to the individual meanings of the node. Each of the association word groups now contains association words of a similar meaning or words that are particularly strongly correlated (Fig. 4).

When clustering, we define the similarity index as the criterion to determine if any given pair of elements should be treated as the same group. The set of elements subject to clustering at this stage consists of the association words of individual polysemic nodes (primary association words). Two procedures may be used to calculate similarity indices as described below.

a. Using Secondary Association Words as Similarity Indices

The association words of a polysemic word (the data used for clustering) are called primary association words. The words associated from the primary association words are called secondary association words. In the first technique, similarity between the secondary association word groups is quantified to obtain the similarity indices.

The similarity between the secondary association word groups is determined by counting the number of common nodes. In Fig. 5, the association word groups for Library and Page have two words in common: Read and Web. The value 2 is directly used as the similarity index, and this is called the absolute similarity index. From another perspective, of the four words in the two secondary association word groups, the two words Read and Web are common, thus the ratio of the number of common nodes to the total number of nodes is 2/4. The similarity index in this representation is called a similarity index by ratio. In the remainder of this paper, when we speak of a similarity index, we are referring to a similarity index by ratio.

In hierarchical clustering, the elements are grouped together in the order of high similarity index. In semantic classification, there remains the problem of how to handle secondary association word groups when the elements are clustered. A sum-set or product-set approach can be used to determine the association words for the new cluster. For example, when the primary association words of Library and Page are clustered using the sum-set approach, the new association word group includes Read, Index, Number and Web. When using the product-set approach, the new association word group includes Read and Web. When the sum-set approach is used, the secondary association word groups increase rapidly as agglomeration proceeds, and the probability of matching other association word groups increases, which will eventually pick up even those elements that should not be classified into the same group. This will end up with space dilation. For this reason, we have selected the product-set approach.

The secondary association words are used to calculate the similarity indices because the association words represent the specific features of the central word. To improve the clustering accuracy, we remove in advance any nondistinctive words that are included in nearly all of the secondary association word groups. In our experiments, we define nondistinctive words to be words that exist in 80% of the secondary association word groups. We performed clustering after removing these nondistinctive words.

b. Using Connectivity for the Similarity Indices of Primary Association Words

The second available procedure involves using connectivity to define similarity indices. Connectivity, the strength of connection between two nodes, is expressed by the sum of the association values of the connected nodes. Association values indicate the ease of association from one given word to another and are shown by directed edges in the AK network. On the other hand, connectivity is shown by bi-directional segments that directly indicate the strength of the connection of the two words involved.

When clustering, the connectivity must be evaluated not only between single nodes but also between clusters. Connectivity between clusters is expressed as the average of the connectivity between individual nodes of one cluster and those in another. Specifically, the connectivity between cluster P (number of elements: n_p) and cluster Q (number of elements: n_q) is defined as follows: Assume p_i is the i-th node of cluster P and q_j the j-th node of cluster Q. Obtain all connectivity $b(p_i, q_j)$, and divide the sum by $p_i \times q_j$ (all combinations of p_i and q_j) to derive the average (equation 4 and Fig. 6).

$$b(P,Q) = \frac{\sum_{i=0}^{n_p} \sum_{j=0}^{n_q} b(p_i, q_j)}{n_p \times n_q} \tag{4}$$

$b(P,Q)$: Connectivity between clusters P and Q

$b(p_i,q_j)$: Connectivity between p_i and q_j

n_p : Number of elements in cluster P

n_q : Number of elements in cluster Q

c. Results of Experiments using the Two Procedures and a Comparison of Results

Experiments were conducted on the procedure of representing the common components of the secondary association word groups as similarity indices and on the procedure of using the similarity indices of the primary association words expressed by connectivity. The results are compared. It was found that, as a clustering technique, using the similarity indices of the primary association words expressed by connectivity yields better results than representing the common components of the secondary association word groups as similarity indices (Fig. 7). This is because of the following reason: in this knowledge database, the association index from one word to another is calculated for all possible combinations of words. As such, evaluating the association indices between words included in the primary association words (connectivity) is more direct than counting the number of common words included in the secondary association word groups. The authors therefore believe that the connectivity-based procedure more readily provides ample information for clustering.

3.4.2 Hierarchical Processing of Concepts

Humans not only share knowledge among themselves but also use hierarchization of language concepts to facilitate mutual understanding when communicating. For example, Hawk and Crow are subordinate concepts of Bird. Bird, on the other hand, is a part of a larger classification of Animal, or put differently, Animal is the superordinate concept of Bird. The concept gets more abstract as we go up in the hierarchy and gets more concrete as we go down the hierarchy. All concepts represented by languages have this type of hierarchical structure. This would indicate that humans are born to accumulate knowledge through hierarchization unconsciously (Fig. 8). [7] This structure helps to extend the span of communication. Humans can explain what they want to say to others by describing things in concrete or in the abstract as required. Assume one of your friends is wondering what kind of pet he/she should have. You want to recommend a dog, and you may simply, and generally, say, “Dogs are obedient to their masters and they’re easy to care for,” without taking the trouble of mentioning each and every individual breed of dog, say for instance, “A dachshund is obedient to its master and it’s easy to care for,” “A bulldog is…,” and so on. Then you can focus more concretely, saying, for example, “Of all dog breeds, a dachshund is docile and easy to keep.” How can we create such a hierarchical structure in our AK network? We explained that superordinate concepts are more abstract than subordinate concepts. Because of this fact, superordinate concepts are easier to find in text data than subordinate concepts. In other words, it is generally said that more superordinate concepts appear in text than subordinate concepts, and that the superordinate concepts have more association words than subordinate concepts. Accordingly, the values of the associations from the subordinate to the superordinate concepts are higher than the values of associations from the superordinate to the subordinate concepts.

We believe that this difference in the association values is an important factor in differentiating between the superordinate and subordinate concepts.

To verify this hypothesis, we conducted the experiments described below.

3.4.2.1 Experiments on Concept Hierarchical Processing Based on Differences in Association Values

Associations from subordinate to superordinate concepts are generally stronger than associations from superordinate to subordinate concepts as mentioned before. In our experiments, we included one word generally considered a superordinate concept and five words generally considered subordinate words to that superordinate word. These six words were presented to human subjects to determine whether or not the superordinate concept word was correctly extracted. Of the six words, one word with higher association values than the association values of all the other five words was extracted as the superordinate concept.

The superordinate concept was correctly extracted in many cases (Table 2). In one case, however, the hypothesis was not valid as shown by the failure in Table 3. This was because a polysemic word (Apple) was included in the word group. This indicates that proper processing of polysemic words is indispensable for enhancing the accuracy of concept hierarchization.

Our experiments have shown that this hypothesis is valid for a large number of concepts. It is difficult to achieve hierarchization of all concepts using this approach alone, but we consider this approach to be an important element of concept hierarchization.

4. Experiments and Considerations about the AK database

Experiments were conducted on the AK database to confirm the effectiveness of the approach used in our present study. Vocabulary matching was tested with the Standard Vocabulary List SVL12000 and the *Eijiro* electronic English-Japanese dictionary to determine to what extent standard words were included in the AK database. We conducted a questionnaire survey using general public subjects to verify if the association and *kansei* that were artificially calculated by the database approximated human sensibilities. The results of this survey and the information in the database were compared.

4.1 Comparing the AK Database with the Standard Vocabulary List and an E-J Dictionary

We visited about 1.2 million websites, collected text data and created a database containing 500,000 English words.

The purpose of our present study was to construct a knowledge database as mentioned before. As such, it was necessary that words actually used by humans be included in the database. We compared our database with the standard word list and a dictionary to find out to what extent our database contained words used by humans.

Vocabulary matching was performed using two lists. One was SVL12000, a list prepared by ALC who provide various support programs for English learning. The SVL12000 contains 12,000 English words that are generally used in conversation and written sentences, with the exception of proper nouns. The other was an English-Japanese electronic dictionary called *Eijiro* that contains more than 300,000 English words. The result of the comparison of our database with these two reference materials is described below (Table 4).

The coincidence ratio with *Eijiro* was 51.52%, which was an inadequate result. The coincidence ratio with SVL12000 was 100%, meaning that our database would be highly useful, with almost all generally used words included.

The reason why the coincidence ratio with *Eijiro* was low could be explained by the scarcity of text data used as the information source and the poor coverage of various genres in our visits to websites. For our database to be able to pick up more words found in *Eijiro*, we would need to collect much more text data and visit many more websites in different categories on the Internet.

4.2 Comparing Our Derived Association-*Kansei* with Human Association-*Kansei*

We conducted two questionnaire surveys using 73 Japanese subjects to learn how much the association and *kansei* of our database approximated human values.

4.2.1 Verifying Association Words by Questionnaire Survey

The objective of the first questionnaire survey was to evaluate the reliability of the association words in our AK database. Twelve basic words assumed to be known by everybody were selected from the database and shown to the subjects together with their respective association word groups. The subjects were asked to evaluate the association word groups in three grades: Natural, Not Sure and Unnatural. The result of the survey is shown at the top of Table 5, and the basic words and their association word groups at the bottom of the Table.

In Table 5, on average Natural was selected by about 66% of the respondents, overwhelmingly exceeding the Unnatural response selected by about 9%. Looking at the basic words individually, the number of respondents who thought the relevant association word group was Natural was predominantly larger than the number of respondents who thought it Unnatural, with the exception of the word Mushroom. For the word Mushroom, the respondents were nearly equally divided among Natural, Not Sure and Unnatural responses. This may have been because of dispersion attributable to different associations by people. None of the basic words was rated excessively Unnatural, which would indicate that as far as these 12 well-known words were concerned the association words in our AK database were effective and reliable.

4.2.2 Verifying *Kansei* Values by Questionnaire Survey

In our second questionnaire survey, nine basic words to be appraised were listed in the questionnaire. For each of the words, the emotion words of Happiness, Anger, Fear, Disgust and Sadness were shown opposite the basic word. The subjects were requested to circle each of the emotion words that he/she thought were applicable to the given basic word. The number of circles were tallied for each emotion word and divided by the number of respondents to derive the average *kansei* value of the respondents. The result of the experiment compares the values of human subjects with the *kansei* values of our AK database (Table 6).

To facilitate comparison, each of the five of the nine basic words selected has a significantly high value for a certain *kansei* in the AK database, while others have mixed kinds of *kansei*.

Of the nine basic words tested, four characteristic words are described below.

As seen in our results, for eight of the nine basic words, human subjects and our database shared the same emotion word as the word of the highest *kansei* value. Furthermore, for four of these eight words, human subjects and our database shared another same *kansei* word as the word of the second highest *kansei* value. For the basic word Complaint, which had the largest discrepancy between the human and database *kansei* of all nine basic words, the top two emotion words, put together, were identical (Disgust and Anger). It is true that humans can have emotions which are a mixture of three or more categories of *kansei*, and the agreement of the top two emotion words indicates that our database contains the features of human *kansei*.

5. Problems and Prospects

Our experiments have broadly verified the effectiveness of our AK database in terms of the number of nodes, properties of association words and *kansei*. The problem is that matching with the dictionary was low. Thus, it is necessary to extract more text data.

It is also essential, as mentioned before, to divide polysemic words by their meanings. This division and the accompanying classification would enable us to sort not only nodes but also the frequency of appearance and association words, and then assumed relationships would be applicable to many more concept hierarchies.

Our experiments have demonstrated that differences in association values between the superordinate and subordinate concepts play an important role in the construction of the hierarchical structure. But such relationships alone are not sufficient. To form a hierarchical structure for all concepts, the sorting of concepts into respective categories is necessary, in addition to the above relationships. To hierarchize concepts in practice, we specify central words, and extract superordinate and subordinate concepts based on the association words of respective central words. In reality, association words are a mixture of words of different categories, although the interrelationships of the words are strong. For example, the association words of the central word Dog may include Collar, Dog Food and other words that have no hierarchical relationship among them. These types of words must be eliminated during hierarchization. A method to evaluate association values among the association words is currently being studied. The underlying idea of this method is that the association value of association words belonging to the same category would be higher than the association values of words belonging to different categories.

6. Conclusion

This paper described an association-*kansei* (AK) network that is the information source for the artificial flow of consciousness and the creation of artificial emotions. It discussed the classification of polysemic words and concept hierarchization. The authors' AK database, incorporating these features, is able to associate words similar to human functions and has emotions that approximate those of humans as shown by the experiments. The Internet was used as an information source, so the system can flexibly reflect the trends of the times.

We believe that a robot capable of communicating with humans can be created by making the AK network closer to the knowledge structure of humans and achieving a more human-like flow of consciousness.

7. References

A. Kometani, H. Watabe, T. Kataoka, Constructing a Commonsense Consciousness Judgment System, *The 17th Conference of Japanese Society for Artificial Intelligence*, 2003.

T. Kaneko, J. Takeno : Information Scientific Research of Funniness and Sadness, *IEEE Int. Work. On Robot and Human Interactive Communication*, 0-7803-7222-0/01IEEE, pp.450-455, 2001.9

T. Kanamori, Y. Imai, J. Takeno, Extraction of 430,000 association words and phrases from internet web sites, *International Conference on Machine Automation*, 2005, 1929-1934.

Y. Imai, M. Tamada, J. Takeno : Method for Extracting Associated Words, Association Value and Kansei Value from the Internet, The 8th World Multiconference on Systemics, Cybernetics, and Systemics(IIIS), *Proceeding Vol.III*, ISBN 980-6560-13-2, pp.160-165, 2004-7.

Y. Kato, Y. Imai, T. Kanamori, K. Furuike, J. Takeno, Extraction of Association and Phrases Information from the Internet and Creation of a Knowledge Database, *3rd International Conference on Autonomous Robots and Agents (ICARA)*, 2006.

A. Ogiso, S. Kurokawa, M. Yamanaka, Y. Imai, J. Takeno, Expression of emotion in robots using flow of artificial consciousness, *IEEE Int. CIRA2005*, 2005, 421-426.

S. Kawakami, *An introduction to cognitive linguistics* (Tokyo, Kenkyusha, 1996).

N. Goldblum, *The brain-shaped Mind* (Cambridge University Press, Cambridge, 2001).

K. Araki, *An introduction to natural language processing* – a computer which can talk and learn a language - (Tokyo, Morikita Pulishing, 2004).

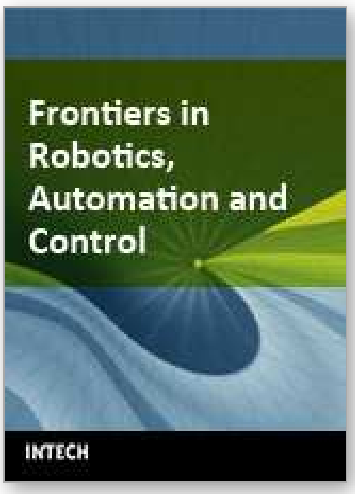
R.F. Simmons, *Computations from the english* (Englewood Cliffs, NJ: Prentice-Hall, 1984).

M. Nagao, *Natural language processing* (Tokyo, Iwanami Shoten, 1996).

J. Takeno, T. Kaneko: "*The Stream of the Consciousness Network of Machine Mind*", International Institute of Informatics and Systematic, 2002.6

Naomi Goldblum: "*The brain-shaped Mind*", Cambridge University Press, 2001

Manfred Spitzer: "*The Mind within the Net*", M.I.T. Press, 1989.



Frontiers in Robotics, Automation and Control
Edited by Alexander Zemliak

ISBN 978-953-7619-17-6
Hard cover, 450 pages
Publisher InTech
Published online 01, October, 2008
Published in print edition October, 2008

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821