Toward Fully Automatic Categorization for Commonsense Processing

Rafal RZEPKA, Toshihiko ITOH, Kenji ARAKI
Language Media Laboratory
Graduate School of Information Science and Technology
Hokkaido University
{kabura,t-itoh,araki}@media.eng.hokudai.ac.jp
Kita-ku, Kita 14-jo Nishi 9-chome 060-0814 Sapporo, Japan

Abstract

In this paper we propose a new approach to categorization where the category naming is fully automatic. It is performed by using occurrences of Japanese nouns, verbs and particles which have specific ability for combining grammatical and semantic information. We explain the importance of automatic categorization for our commonsense research and Artificial Intelligence. Then we will introduce the technical side of the idea and the results of the preliminary trials on randomly chosen sentences from the WWW corpus. In the last part we will propose an idea of using automatically labeled categories in our Schankian script retrieval project. By developing non-human, full automatic methods for machines we expect to achieve applications which could work in open domain with as little user's intervention as possible.

Keywords: *Categorization, Commonsense Processing, Machine Intelligence.*

1 INTRODUCTION

1.1 Computers without the Commonsense

Nowadays, the world of technology changes our lives as significantly as it did in XIX century but in a quite different manner. The first personal cars had only four functions - moving, stopping, turning left and turning right, but personal computers of XXI century can handle hundreds of much more complicated functions still having only one user. Furthermore, the meaning of "personal" in the expression "personal computer" will be changing rapidly from "owned by a person" into a nuance of "living with a person". People who does not like cars can use trains but probably there will be no other choice for people who does not like computers. When a computer become an indispensable item of controlling every electronic device around us, the quality level of our lives will depend heavily on how skillful we are in manipulating such "house control centers". Obviously the number of possibilities of interfering and combining multi-functions will be growing constantly, together with the development and miniaturizing the ubiquitous devices. Soon this number of possibilities will overgrow the programmers imaginations and users might demand things that they never thought of. As only a minority of society will be able to program such complicated systems, the perfect solution for this problem would be a computer which can understand our natural language and translate it to its own, artificial, machine language. Without this, a division into IT-literates and IT-illiterates will deepen, aging society will not be able to use the newest technology to help taking care of elders. Another problem is if the elders would like to trust the machines which do not understand basic things, human beings and their everyday life. Many of the elders would feel lonely and caring for a chit-chat with these soulless things and maybe get more frustrated by the boring, machine-like expressions. Achieving such naturalness is a goal for many researchers already, but we think they take a roundabout way. Current interfaces for human-computer communication are being developed to *behave* as naturally as possible, their eyes follow our faces, artificial hands try to gesticulate, faces expressions to react when hearing our laughter [1][2]. However, they still lack what is most important for natural communication — the naturalness of their language responses. We claim that without expressing itself naturally, the machine will be unnatural even if its body will be perfectly resembling human's. For over 40 years computer scientists have been trying to create a program behaving naturally enough to pass the famous Turing test but they always failed. The easiest way to distinguish a computer from a human is to talk about common things acquiring general knowledge which every one of us has from the childhood. Such knowledge is very difficult to input into a computer as it is so natural for us that we can hardly think about it. In my opinion it is impossible to input such amount of information by humans themselves and this thesis will propose the automatic approach for gathering such knowledge, usually called "commonsensical".

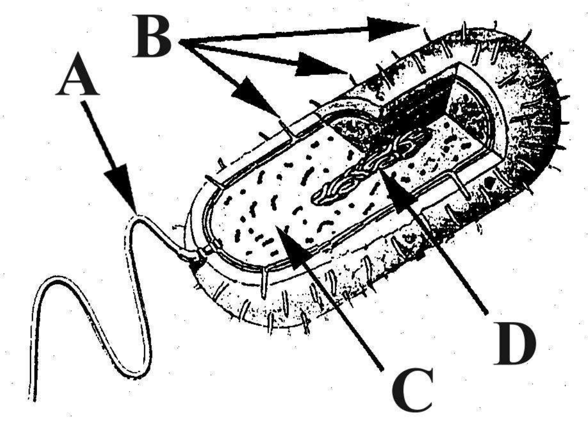

Figure 1. Bacterium Lingualis (A - Flagellum, B - Positiveness Receptors, C - Concrete and Abstract Knowledge Memory, D - GF Cell)

1.2 Idea of an Average Personality Model

The most popular objection made by the adversaries of Artificial Intelligence and systems behaving like humans is that a machine will never be able to go beyond the borders created by a programmer, that computers will always be only the prisoners of somebody's algorithm and personally chosen data. Deeply interested in commonsense problem, we have proposed an idea to escape this quite convincing objection: to build self-creating personality of average human being basing on the biggest existing database, which is the Internet. The humans' knowledge certainly differs from individual to individual but the biggest part of knowledge is common especially among one culture. In our approach, we do not concentrate on retrieving any specific knowledge which should be correct, we try to simulate the physical behaviors and reactions of unknown somebody in specific situations, which should lead us to more natural linguistic behaviors. By grasping the knowledge of average behaviors, the system is able to discover automatically what is unnatural and rare, what is original and what is surprising. Such reactions are the key for the natural reactions, also these made by expressions during the talk. We have partially confirmed that automatic personality modeling is possible and we anticipate that if we succeeds with all our plans, many new possibilities will appear for numerous projects which results are still unsatisfying as, for example, automatic creation of Schankian plans, goals, scripts[3] or Minskian frames[4]. But to realize such ideas form the cognitive sciences field, we will first need an elastic categorization method. This paper is our first approach to solve this problem.

1.3 The commonsense - State of Art

The idea of automatic personality modeling based on Internet resources is novel and has not been researched yet, which causes problems with comparisons and evaluating the introduced methods. Apparently there are several topics of computer science fields that are constantly influencing us and contributing with their ideas and technical methods. However, in this paper we decided to use our interdisciplinary background and combined the latest achievements of leading computer researchers with our ideas built during the studies in fields of linguistics and cognitive science. Even if using methods similar to their originals, the comparisons would be aimless as the goal of our research is different than goals of particular researchers who indirectly contributed with their technical ideas (as Inui [5]). Commonsense itself was a topic of computer science researches carried out, but the scientists often gave up the experiments because a very simple reason - the lack of data and machine power. Now, when the computers are getting faster and more effective, when enormous sets of data can be stored, the computer scientists tend to work on specialized fields as expert systems, where commonsense is thought to be a minor factor. There are only a few big projects on achieving commonsensical knowledge but they concentrate on manual data input which, as mentioned above, we claim against. Mueller's "Thought Treasure" is one of first quite successful examples of collecting and categorizing big amounts of commonsensical data [7], where the author was a system and ontologies architect. Lenat [6] decided to hire a group of specialists, which were not only inputting and creating categories and ontologies between concepts, but also were gathering data from several digitalized resources. As the Lenat's CYC project does not give satisfying results for years of absorbing millions of US dollars, the scientists form Massachusetts Institute of Technology decided to cooperate with Thought Treasure's author and started its own project based on the idea where the hundreds of Internet users are inputting the sentences which are supposed to be categorized within the proposed categories[8] [9]. These three approaches methods differ from our ones fundamentally by not using manually inputed data and retrieving the needed knowledge from the raw text of Internet pages. Another difference is mutual interference of commonsense and emotions as we presume that first the feelings are building human's commonsense and at some point the commonsense is something that blocks our feelings in everyday situations. The problem of categories is touched by above mentioned researchers but their approach is to label and manage the categories by hand while we aim a fully automatic method.

1.4 Merits of Fully Automatic Categorization

As it was said in the previous subsections, our goal is to achieve a system which uses the Web to make itself intelligent while current approaches concentrate on web-mining which purpose is to make us, humans, intelligent, to give us specific, not commonsensical, information. In the beginning of our categorization trials [10], we referred to already existing methods [11] and successfully used WWW to automatically assign object to basic categories made by hand using the occurrences in WWW corpus of Japanese language as we agree with Fillmore [12] who argues that Japanese and especially its particles are perfect for computational processing (for the particles list see the particles in Tab. 1):

- Animate: Object+ga+iru (ite/ita)[*is* for living beings]
- Inanimate: Object+ga+aru (att) [*is* for non-living objects]
- Place: Object+ni+iru (ite/ita) [*is* for living beings]
- Tool: Object+de+tsukuru (tsukutt) [*prepare, create*]
- Food: Object+wo+taberu (tabe) [*eat*]
- Drink: Object+wo+nomu (non) [*drink*]
- Vehicle: Object+ni+noru (nott) [*get on, ride*] etc.

But there were problem not only up to labor but difficulties on labeling the categories for abstractive objects, as *aidia* (idea) or *fun'iki* (atmosphere, mood). We understood that only the maximal limitation of human input can make the commonsense and artificial intelligence make a big leap, therefore we seek for methods letting the machine perform the categorization by itself with as little programmer's help as possible. The main bottleneck of making a computer to learn categories is that we need to prepare understandable category names and put objects under these categories in order to create the learning set. To evaluate such learning process, the category manual labeling is necessary, as the human must understand the category name to estimate if given objects were collected properly. We claim that it coerces us into preparing laborious data sets manually only for the sake of evaluation, while the evaluation could be performed later, for instance in the application level.

Particle	Role
WA	Topic-Indicating
GA	Linking-Indicating
NO	Possessive-Indicating
WO	Object-Indicating
NI	Direction-Indicating
DE	Place or Means of Action-Indicating
HE	Destination-Indicating
TO	Connective
MO	Addition-Indicating
YORI	Comparison-Indicating
NODE	Reason-Indicating
V-KARA	Reason Indicating
N-KARA	Lower Limit-Indicating
MADE	Upper Limit-Indicating
DEMO	Emphasis-Indicating

Figure 2. Main Japanese particles and their functions

2 TECHNICAL SOLUTIONS

2.1 Bacterium Lingualis

Full automatizing needs as big corpora as possible, since, as we demonstrated in previous works [10], not only the quality, but also the number of commonsensical inputs is crucial for learning the laws ruling our world. We "stepped back" in evolution and started creating an insect to start learning from the very bottom without forcing it to behave on Cartesian philosophy. By Latin "Bacterium Lingualis" we mean a web crawler exploiting only the textual level of WWW and treats it as its natural environment. We assume that cognition, by which we mean the process or result of recognizing, interpreting, judging, and reasoning, is possible without inputs other than word-level ones - as haptic or visual [13, 14]. Although such data could significantly support our method, a robot which is able to travel from one place to another in order to touch something, would cost enormous amount of money, not mention a fact that current sensors technology is not ready for such an undertaking. There are several goals we want to achieve with Bacterium Lingualis. The main one is to make it search for the learning examples and learn from them unsupervisedly. For that reason we decided to move back in evolution and initiate self-developing BL on the simplest level with as few human factors as possible. Other goal is to become a tool for knowledge acquisition which involves language acquisition as the living environment for our program is language itself. We imagine a language as a space where its components live together in a symbiosis. Its internal correlations are not understandable for BL and the learning task is to discover them. It also is to discover its own categories for further usage, for example in the talking interfaces. For exploring such an area we use simple web-mining methods inspired on Heylighen et al.'s work [15]. As we described other particulars about Bacterium Lingualis before [10] we mention only that part C — the Concrete and Abstract Knowledge Memory — are responsible for the categorization in our project.

3 DATABASE FOR CATEGORIES

3.1 The Algorithm for Labeling

The categories were created from 18673 most frequent nouns from ChaSen noun dictionary [16]. The frequency was calculated for 58796 nouns occurrences in 3.282.217 sentences corpus made randomly from WWW by Larbin robot [17] and spaced by Kakasi program [18]. Every of 18673 nouns was checked for its occurrence with Japanese particles in above mentioned corpus. The particles first letters in occurrence order created the category labels for given noun. If for instance, a noun *tekisuto*(text) occurred most frequently with particles Wo, Ga, De, Ni, To, Kara, Made, Yori — the label "WGDNTKMY" was automatically created. Any word which had the same sequence of particles was joining *tekisuto* in its category becoming a new element. This is how the upper level of Bacterium Memory, called an Abstract one, is created. As it is mostly not understandable for humans we also treat it as an indication of Minskian Alien Intelligence for current computers [19].

3.2 Evaluation of Category Labeling

As we expected, the results in this stage were hard to evaluate just by comparing the entries in the database as most of them do not seem to have any relationship between category elements. Although we used similarity algorithm based on [20] (this algorithm was independently discovered as described in [21]) to find examples of categories which elements are clearly related as above introduced category WGDNTKMY found by calculating similarities for noun *moji*.

OBJECT: noun *moji* (characters, letters) Upper Abstract Level Label: WGDNTK (W:wo G:ga D:de N:ni T:to K:kara M:made Y:yori)

- 0.86875 : *ji*(sign, character)
- 0.765578635014837 : *tekisuto* (text)
- 0.730650154798762 : *go* (language)
- 0.730538922155689 : *ryō* (material, fee)
- 0.72289156626506 : *printa* (printer)
- 0.722741433021807 : *ramu* (from *koramu* column)
- 0.716612377850163 : *han* (print, edition)

As we could not confirm the usability of such labeling on this level of abstraction, we decided to research a field where it would be easier to evaluate our way of thinking. We used above idea in our parallel project of automatic Schankian scripts creation which is for us an important part of commonsense processing.

4 CATEGORIES IN SCRIPTS RETRIEVAL

4.1 The Idea of Schankian Script

In his classic book [3] Schank has introduced inquiry into the nature of knowledge that is needed to understand the world and understand natural language. His main claim is that structured knowledge dominates understanding while the question is the content of these structures. He takes a pragmatic approach that does not separate form from content. He proposes conceptual primitives we decided to partially adopt although his proposal has not been tested psychologically. Schank concentrates on memory, and in particular memory organization but we use only our original Bacterium Lingualis Memory concept. He argues, and we agree with him, that understanding language involves causally connecting thoughts/sentences. Because causality is often implied or incompletely described, it is usually harder to understand connected text than individual sentences. Therefore, he describes a formal representation (causal syntax) of causal chains which is to the discourse level what Conceptual Dependency is to the sentence/thought level. Rules are provided in which every primitive action is associated with the set of states it can affect as well as those that enable it. Schank describes scripts as groups of causal chains that represent knowledge about frequently experienced events (the most famous example is "going to a restaurant"). In other words, a script is a stereotyped sequence of actions that defines a well-known situation and has associated with it: a number of roles for the actors (different points of view on the situation, e.g. customer vs waiter vs cook), different tracks (e.g. restaurant, fast-food), different scenes (e.g. enter, order, eat, pay); each scene has a MAINCON, i.e. a main conceptualization, which must have happened if the scene is instantiated, as well as props, entry conditions, results, branches and loops etc. Using scripts requires two mechanisms:

- **Script retrieval**: A script is retrieved if a state is mentioned that constitutes a precondition for the script (e.g. the customer is hungry and has money) and there is a direct reference to a MAINCON or a prop in one of the scenes (e.g. order a dish or step to the counter).
- **Script application**: An active script allows one to infer actions that were not stated (nor contradicted) as well as to instantiate roles etc. Hence the predictive power of scripts in conventional situations. The restaurant script is called a situational script (standard social situation in a specific locale etc.). Other types of scripts include personal scripts (e.g. hitting on the waitress) and instrumental scripts (e.g. lighting a cigarette).

Many interactions can arise in script-based understanding, because several scripts are active at the same time (interference e.g. train and restaurant scripts), or because an action has an unexpected outcome which prevents the script from continuing normally or invokes another script recursively (within an existing script) - script in abeyance. Of course, script-based understanding is only relevant when understanding stereotyped situations. Beyond these, it is necessary to have a model of the actors' goals and of the available plans to satisfy these goals. This kind of understanding is where we would like use WWW retrieval.

4.2 ComAct Units

As the Schankian ideas are complex we decided to bring our own definitions sometimes very loosely related to the ideas introduced my this psychologist. First we had to define the basic semantic unit for processing which we called *ComAct* which is an abbreviation from Commonsensical Action. The nucleus of such unit is a verb with nine most frequently occurring nouns connected by three most frequent particles joining given verb with these noun (see Fig. 3). We use such self creating units for several processes but here we mention their role in the automatic script creation.

4.3 Automatic Script Creation

This is the youngest part of our large project for fully automatic processing for commonsense retrieval but categorization methods introduced above play an important role in making these methods more flexible and time for processing — shorter. So far, we concentrate on the simplest Schankian scripts modeling basing on two or three actions combined in one chain. The lowest level consists of two verbs bi-grams chosen by function of *rentaikei* - Japanese grammar form joining two actions following one after another (*V-te* enzyme in Bacteria Lingualis nomenclature, tekara and te-comma are also used).

$(eat) - - - [te] - - - (pay)$

or

$(eat) - -[te] - -(pay) - -[te] - -(leave)$

The second level of complexity is made by ComActs and statistically self created categories. Every verb in a script is equipped with its ComAct unit as in the following example:

([dinner,bread,apple,...]-wo-eat)

$${\scriptstyle te-enzyme}\Big\downarrow$$

([fee,money,tax,...]-wo-pay)

and other particles, as de, etc.:

([outside,everybody,oneself...]-de-eat)

$${\scriptstyle te-enzyme}\Big\downarrow$$

([banknotes,cash, window...]-de-pay)

The last, complete level is such a chain (script) which CoMArts consist only the elements semantically related to a given script. By now we eliminate not related nouns to a

$$\begin{tikzcd}
 & & \begin{bmatrix} N_{max} \\ N_{max-1} \\ N_{max-2} \\ \vdots \\ N_{max-n} \end{bmatrix} \arrow[d] \\
 & & P_{max-1} \arrow[d] \\
\begin{bmatrix} N_{max} \\ N_{max-1} \\ N_{max-2} \\ \vdots \\ N_{max-n} \end{bmatrix} \arrow[r] & P_{max} \arrow[r] & Verb_1 \\
 & & P_{max-2} \arrow[u] \\
 & & \begin{bmatrix} N_{max} \\ N_{max-1} \\ N_{max-2} \\ \vdots \\ N_{max-n} \end{bmatrix} \arrow[u]
\end{tikzcd}$$

$Verb_1\ :\ a\ nucleus\ of\ an\ unit$

$N_{max,\ max-1,\ max-2}\ :\ 3\ most\ frequent\ nouns$

$P_{max,\ max-1,\ max-2}\ :\ 3most\ frequent\ particles$

Figure 3. The structure of ComAct Unit

verbs chain by calculating the occurrences for given pairs in our corpus. For example when the restaurant script consists the verb “to pay” and three most frequent nouns for a place particle “de” include “a bank”, the system compares the occurrences of every ComArt noun and all verbs of a script, in this case “bank” and “eat” will have low occurrence and the other noun from “de group” can be used.

4.4 Preliminary Tests and its Results

Because of very long time for retrieval (one search takes 49.3 sec. in average) we managed to confirm only three place nouns which are classic examples in script research - *restaurant*, *hotel* and *department store*. Due to the long processing time the authors will complement the results and plan to present the most current state of experiments during the conference. Hereby we must underline remaining problems of statistical categorization as in two examples shown below. First group on nouns were picked up by high frequency of verbs “go” and “to be” (for non-animate objects):

ibu:suru.iku.aru: **X-mas Eve** (WNGDTK)
kissa:iku.aru.tsunageru: **café** (NGDW)
konpa:iku.aru.: **party** (NG)
konsāto:kiku.iku.aru: **concert** (WNGDHT)
madoguchi:uketsukeru.iku.aru: **office window** (DNG-WHM)
resutoran:iku.aru.taberu: **restaurant** (NGDWHT)
se:otozureru.iku.aru: **tide** (WNGTDM)
shika:iku.aru.fukumu: **dentist** (NGWDTH
toko:iku.aru.iku: **place** (NGDMWK)

Every example seems to belong to the same upper category of place but as the example with position changed verbs shows, ambiguous nouns might emerge:

aikyou:aru.iku.furimaku: **charm** GTWD
baiten:uru.aru.iku: **kiosk** DGNWMH
depāto:aru.iku.kau: **department store** GNDWHM
doukutsu:aru.iku.deru: **cave** GNWTDM
kimochi:tsutaeru.aru.iku: **feelings** WGDNKT
moushiire:suru.aru.iku: **proposal** WGN
rikyuu:aru.iku.: **imperial villa** GN
tariumu:yaru.aru.iku: **thallium** WGNH

As we can notice there are three or four items which are hard to be qualified as a place. We plan to avoid such entries again by statistical occurrences in WWW corpus — it is easy to discover that you cannot for example *work* at *feelings*. Although there are always possible errors above preliminary results show a light in the long tunnel to no-human-assistance programs for new non-logic based Artificial Intelligence described by nowadays thinkers as Penrose [27] or Devlin [28] which gave the authors their inspirations.

5 CONCLUSIONS

There are two main reasons for which the authors decided to introduce a research in the very beginning stage of its development. First one is to spark a discussion about the new, commonsensical, approach to the knowledge and language engineering among international researchers. Because of the size of our project we are seeking for help from different fields, cognitive science, linguistics and psychology while we argue that ideas from these fields should be revised in order to contribute to computer sciences’ progress. This is the second reason. So far we have confirmed our methods (except particle based labeling) on very small sets of examples but these trials seem logical and very promising which drove us to send this paper. We reached high level of certainty that automatic categorization based on statistics of verbs and nouns and Internet resources will have high accuracy.

6 FUTURE WORK

There are plenty of experiments still to be done in several fields of our project [22, 23, 24, 25, 26] in order to reach the final goal - the talking agent which would not need a user input to learn and confirmed retrieved commonsense. As we seriously consider classic cognitive science and psychology ideas as scripts, goals, plans or frames to be rethought, the first step of fully automatic script creation must be finished. To achieve this nearest goal we will concentrate on creating as precise scripts as possible and try to add actors, scenes and other elements of Schankian scripts. The next step will be an automatic script evaluation algorithm also based on WWW resources and statistics. We also believe that this permanently broadening corpus can also help with processing situations wandering away from the commonsense point of view which also will be important task of our project.

References

[1] R.W. Picard, J. Klein, Computers that Recognise and Respond to User Emotion: Theoretical and Practical Implications, **MIT Media Lab Tech Report 538**, to appear in Interacting with Computers, 2003

[2] M. Seif El-Nasr, T.R. Ioerger, J. Yen **A Web of Emotions**, Proceedings of Workshop on Emotion-Based Agent Architectures part of Autonomous Agents, 1999.

[3] Schank R.C. and Abelson R.P. **Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures (Chap. 1-3)**, L. Erlbaum, Hillsdale, NJ, 1977.

[4] Minsky, M. **A Framework for Representing Knowledge**, The Psychology of Computer Vision, P. H. Winston (ed.), McGraw-Hill, 1975.

[5] Inui T., Inui K. and Y. Matsumoto **What Kinds and Amounts of Casual Knowledge Can Be Acquired from Text by Using Connective Markers as Clues?**, 6th International Conference on Discovery Science (DS 2003),Sapporo, Japan. Lecture Notes in Artificial Intelligence, vol 2843, 180-193, 2003, Springer-Verlag, Berlin, Heidelberg, New York.

[6] D. Lenat et al.,**Common Sense Knowledge Database CYC**, 1995, http://www.opencyc.org/, http://www.cyc.com/

[7] E. Mueller, **Natural language processing with ThoughtTreasure**, Signiform, New York, 1998.

[8] P. Singh **The public acquisition of commonsense knowledge** Proceedings of AAAI Spring Symposium on Acquiring (and Using) Linguistic (and World) Knowledge for Information Access. Palo Alto, CA: AAAI, 2002.

[9] D.G. Stork, **"Open Mind Initiative"**, 1999, http://openmind.media.mit.edu/

[10] Rzepka R., Araki K. And Tochinai K. **Bacterium Lingualis - the Web-based Commonsensical Knowledge Discovery Method**, 6th International Conference on Discovery Science (DS 2003),Sapporo, Japan. Lecture Notes in Artificial Intelligence, vol 2843, 453-460, Springer-Verlag, Berlin, Heidelberg, New York, 2003.

[11] Fujita S., And Bond, F. **Extending the Coverage of a Valency Dictionary**, COLING-2002 workshop on Machine Translation in Asia, 67-73, Taipei, 2002.

[12] C.J. Fillmore, The Case for Case, in: E. Bach & R.T.Harms, eds., 1968: **Universals in Linguistic Theory**. New York: Holt, Rinehart & Winston, 1968, 1-88.

[13] C.F. Kielkopf, The Pictures in the Head of a Man Born Blind, **Philosophy and Phenomenological research**, Volume 28, Issue 4, June 1968, 501-513

[14] J.F. Fletcher, Spatial representation in blind children, Development compared to sighted children,**Journal of Visual Impairment and Blindness 74**, 1980, 381-385.

[15] F. Heylighen, Mining Associative Meanings from the Web: from word disambiguation to the global brain, **Proceedings of Trends in Special Language and Language Technology**, R. Temmerman (ed.), Standaard Publishers, Brussels, 2001.

[16] Y. Matsumoto, A. Kitauchi, T. Yamashita, Y. Hirano, H. Matsuda, and M. Asahara, **Japanese morphological analysis system ChaSen version 2.0 manual 2nd edition**, 1999.

[17] Larbin Web Robot, **Larbin Project**, 1999, http://larbin.sourceforge.net/

[18] H. Takahashi, **KAKASI - Simple Kana Kanji Converter**, 1992, http://kakasi.namazu.org/

[19] M. Minsky **Communication with Alien Intelligence**, Extraterrestrials: Science and Alien Intelligence. (Edward Regis, Ed.) Cambridge University Press, 1985. Reprinted in Byte Magazine, April 1985.

[20] E. Myers, **An O(ND) Difference Algorithm and its Variations**, Algorithmica Vol. 1 No. 2, 1986, pp. 251-266.

[21] E. Ukkonen, **Algorithms for Approximate String Matching**, Information and Control Vol. 64, 1985, pp. 100-118.

[22] R. Rzepka, K. Araki, K. Tochinai, Basic Idea of General Belief Retrieving Agent "GENTA" **2001 IEEE Hokkaido Chapters Joint Convention Record**, IEEE Sapporo Section, 2001,pp. 414-415.

[23] R. Rzepka, K. Araki, K. Tochinai, Prediction of the User's Reply Using Emotional Information Retrieved from Internet Resources **2002 IEEE Hokkaido Chapters Joint Convention Record**, IEEE Sapporo Section, 2002, pp. 229-230.

[24] R. Rzepka, K. Araki, K. Tochinai, Is It Out There? The Perspectives of Emotional Information Retrieval from the Internet Resources, **Proceedings of the IASTED Artificial Intelligence and Applications Conference**, ACTA Press, Malaga, 2002, pp. 22-27.

[25] Rzepka R., Araki K. and Tochinai, K., **Ideas for the Web-Based Affective Processing**, Proceedings of the Seventh Multi-Conference on Systemics, Cybernetics and Informatics; vol. XIV Computer Science, Engineering Applications, Orlando, Florida, USA, 376-381. 2003.

[26] R. Rzepka, K. Araki, K. Tochinai, Emotional Information Retrieval for a Dialog Agent, to appear in **Special Issue of Informatica - an International Journal of Computing and Informatics on Perception and Emotion Based Reasoning**, 2003.

[27] R. Penrose,**Shadows of the Mind. A Search for the Missing Science of Consciousness**,Oxford Univ. Press, 1994.

[28] K. Devlin, **Goodbye, Descartes. The End of Logic and the Search for a New Cosmology of the Mind**, John Wiley & Sons, Inc., 1997.