Some Considerations on an Automatic Indexing System by Use of a Title of the Document

NORIHIRO ABE*, JUN-ICHI TOYODA* AND KOHKICHI TANAKA*

1. *Introduction*

This paper presents a method of automatic subject recognition in scientific papers. It is shown that our system with role-indicators is expected to have higher recall factor than usual ones. In this system, the concepts of role-indicators and the algorithm which is able to assign automatically the index to a given title of the document are quite different from usual ones. Therefore the ambiguity of rules which assign role-indicators to words in the documents and miss-assignment of role-indicators by the man may be avoided. On the design of the algorithm, not only the statistical estimation are considered in KWIC and so on, but also the general characteristics inherent to natural languages and the syntactic structure of the title are considered. By considering such a limit of availability of the present computer as encountered in the recognition of semantics and the trade-off between the manageability of the system used and the retrieval efficiencies, we must design a meta-language L^* which describe the subject of the document D written in a natural language L, and to construct an algorithm $f: D_{\text{over}L} \rightarrow L^*$ in order to divide the documents set into fairly large subsets, as in the case of a uniterm indexing system. The title of a scientific paper gives us only an outline of the articles discussed in its paper. Our system has a function that is able to transform the given title to the indexes.

2. *The Characteristics of the Title of Scientific Paper*

2.1. *Syntactic Characteristics*

Almost every title may seem to be constructed by such a phrase-structure as a noun phrase, and the role of each phrase and the relation between phrases are given by prepositions. The fact expressed by each phrase is given by the phrase itself and the function of a preposition associated with it.

2.2. *Lexical Characteristics*

Words and prepositions used in the title have a function of defining the fact expressed by the preceding and succeeding phrases as well as by the

This paper first appeared in Japanese in Joho Shori (the Journal of the Information Processing Society of Japan), Vol. 11, No. 12 (1971), pp. 699-710.

* Osaka University.

phrase containing those words. For instances,

An Introduction to..... A General Method for.....

A Note on.....

are the typical examples. These words underlined do not express significant subjects but are used as index terms to clarify really significant subjects. It is, therefore, possible to clearly describe the fact which the phrase represents by what kind of words are used in its phrase and by what kind of prepositions succeed them.

2.3. *The Subject of the Title*

We have designed the symbols and roles to extract the subject of the title. These symbols are called syntactic symbols.

M: the most significant and mainly discussed matter of the paper (abbreviated MSMM.).

P: the property, attribute of MSMM.

R: the associated matter with MSMM.

O: the object matter of MSMM.

D: the restrictive matter on MSMM.

C: the term on area disscussed in the paper and the term of indicating instruments used for the paper.

Next, we define a phrase contained in the title as the semantic symbol and it has a one-to-one correspondence to a syntactic symbol. Then the meta-language of indicating the subject is $\{n-NP\}$, where $n \in \{M, P, R, O, D, C\}$, NP is a noun phrase.

3. *The Design of the Algorithm of the Automatic Subject Analysis*

2.1. *Preliminary Considerations on the Algorithm.*

The algorithm of the automatic subject analysis may be designed under several assumptions based on the characteristic feature of the title and on the experimental studies.

(1) The Function of the Prepositions. We make the assumption that the preposition has the following function in the title.

"OF" links MSMM with its attribute or related item.

"FOR" links MSMM with the item of its object.

"ON" links MSMM with the item of its attribute or its condition.

"BY" links MSMM with the item of its condition.

"IN" links MSMM with the item of its condition.

"TO" links MSMM with the item of its object or its attribute.

"WITH" links MSMM with the item of its restrictive condition.

In actual sentences there are several exceptions to these assumptions. For example,

* A Design of Automata. (1), * A Principle of Automata. (2)

* A Covering of Automata. (3)

It would be clear that the MSMM in the title (1) and (2) is each of the second phrase "Automata" and the same one in (3) the first phrase "A Covering". Since each first phrase of (1) and (2) consists of "Design" and "Principle", respectively which are so called common words, these phrases are considered to be the explanatory parts. On the contrary "Covering" in the title (3) mentioned above is not a common word but forms an important class of automata theory. Hence, it seems to be an MSMM. In case of simple title, it is also possible to assign a role to each phrase consulting a dictionary in which, for example, a list of common words and so on are stored. In case of titles containing several prepositions what kind of role may be assigned to each phrase in the titles depends upon the kind of prepositions and their combinations. Consider the following sentences.

* Applications of Boolean Algebra to Switching Circuit Design. (4)

* An Approach to the Syntax-directed Analysis of Graphic Data. (5)

In such titles, it is necessary to determine with what phrase each preposition is associated. For this purpose, some assumptions should be set upon the function of prepositions. And title sentences are segmented into the subsequences so as to match to one of the following patterns as "*NP* of *NP*", "*NP* of *NP* for *NP*" and so on. For example, (4) and (5) are segmented as below, respectively;

Application of Boolean Algebra,
to Switching Circuit Design, } (4)

An Approach,
to the Syntax-directed Analysis of Graphic Data, } (5)

The title (4) is considered to have the structure of "*NP* of *NP*'„ "~to *NP*" and processed similarly as in the former examples (1), (2) and (3). However, the title (5) is not segmented into "*NP* to *NP*", "~of *NP*", because it is found from the experimental studies that such a segmentation is unnecessary to our system. These preposition sequences are listed in the Table 1. These patterns and the number of them are determined from the investigation of prepositions sequences used in actual titles. Next, consider the following examples:

* The Construction of Minimum Redundancy. (6)

* General Recursive Function of Natural Number. (7)

In these titles, as no common words are used, they can not be analyzed successfully by the above algorithm. Then, in the present paper it is assumed that an MSMM needs more "significant explanations" than others and has to be modified with more number of words. We mean by the word "significant explanation" those modifiers as articles, adjective such as "some", and adverbs

Table 1. Patterns of Preposition Sequence in the Title.

SEQUENCE	%	SEQUENCE	%
—on—.	2.1	—of—by—of—of—.	0.1
—on—of—.	2.6	—of—on.	0.7
—on—of—of—.	0.6	—of—for—.	1.3
—on—of—in—.	0.8	—of—for—of—.	0.7
—on—of—by—.	0.2	—of—for—with—.	0.1
—on—of—to—.	0.1	—of—for—on—.	0.3
—on—by—.	0.5	—of—with—.	1.7
—on—for—.	0.8	—of—to—.	1.1
—on—in—.	0.5	—of—to—with—.	0.2
—. (OR—and—.)	27.6	—of—to—to—.	0.1
—for—.	8.5	—of—of—.	2.8
—for—of.	3.3	—of—of—in—.	0.1
—for—with—.	0.6	—of—of—of—.	0.6
—for—in—.	1.3	—of—of—for—.	0.2
—for—by—of—.	0.2	—of—of—to—.	0.3
—for—of—of—.	0.1	—of—of—by—.	0.1
—in—.	5.2	—to—.	1.4
—in—of—.	0.7	—to—of—.	1.6
—of—.	21.5	—with—.	1.5
—of—in—.	4.8	—with—of—.	0.3
—of—in—of—.	0.1	—with—on—.	0.6
—of—by—.	0.9	OTHER	1.4
—of—by—of—.	0.1		

which do not effectively modify any words. These meaningless words are registered in the dictionary and are referred when demanded. The part of them are listed in Table 2.

(2) Some further Processings. Such titles as containing "AND" appear frequently as shown in Table 1 and it is important to process these titles as well.

(a) Processing of "AND". As the patterns of "AND" that connects words, the following four ones are set up from the investigation of actual examples.

noun phrase and noun phrase, noun phrase and a noun,
an adjective and noun phrase, a noun and a noun,

The present authors have considered only next two cases regarding "AND" used between sentences, that is, (i) "AND" in the title sentences in which any prepositions is not used. (ii) "AND" which is immediately followed by "its", "their" etc., i.e. possesive pronoun or an article. Of course, there are a few exceptions, but most titles can be considered to have the above-mentioned patterns.

(b) Processing of the Word which ends in the "ING" form. Consider the next example:

* An Algorithm for Assigning Role-indicator to the Title. (8)

Table 2. Examples of Φ Word and U-Word.

(WORDS OF NO SIGNIFICANCE)	(IDIOMATIC WORD)
A	ANALYSIS
ABNORMAL	APPLICATION
ABOUT	APPROACH
.....	APPROXIMATION
HIGH	ASPECT
ITS	.....
LARGE	CLASS
MANY	COMPARISON
MORE	DEVELOPMENT
NEW	EFFECT
.....	ESTIMATION
PRACTICAL	EXPRESSION
SOME	EXTENSION
THE	FOUNDATION
THEIR	.....
THREE	IMPROVEMENT
.....	INTRODUCTION
	INVESTIGATION
	.....
	METHOD
	NOTE
	PRINCIPLE
	PROBLEM
	PROCEDURE
	PROPERTY
	.....

the word "Assigning" should be separated from "Role-indicator" because it does not work as an adjective. For the purpose of this process, words such as "Switching" or "Covering" that is used as an adjective or a noun in the meta-language are registered into the dictionary, and referring to this dictionary, words which ends in the "ING" form are segmented from another words if necessary.

(c) Processing of the Word which ends in the "ED" form. Similar to the "ING" word, these words are registered into the dictionary, and to segment them from a preceding noun phrase, this dictionary is referred.

3.2. *List of Symbols and Definitions*

Definition 3.1.

(1) "♯" is the position of the title head.

(2) "NP_i" is the i-th noun phrase from the title head.

(3) "$W(NP_i)$" is the last noun of the NP_i. If the last word of NP_i is not a noun, then it means the noun just prior to the last word. And "$N(NP_i)$" is

the number of word of NP_i.

(4) If NP_i is a $(\varphi)^1N$, then it is denoted as $NP_i \Rightarrow N$, where φ is an element of Φ which is the subset of adjectives such as {a, an, the, its, their, ...}, and N is a noun and $(\varphi)^1$ implies φ or Λ (null word) and also if $\varphi \in \Phi$ then $N(\varphi)=0$.

(5) If $W(NP_i)$ is an element of the common word set U, it is denoted as $W(NP_i) \in U$, where U is the subset of idiomatic word such as {application, approach, ...}.

Definition 3.2.

(1) If each noun phrase NP_{i+n} of the sub-sequence of the title sentence $\sharp P_i\ NP_i\ \sharp P_{i+1}\ NP_{i+1}\sharp \cdots \sharp P_{i+n}NP_{i+n}$ is assigned for its role X_i under a condition α respectively, then it is denoted as. $n: \langle\!\langle \sim \sharp P_i\ NP_i\sharp\ \sharp P_{i+n}NP_{i+n} \rangle\!\rangle$; α, $[X_{i_1}, X_{i_2}, X_{i_n}]^k$ where "n" is a label of this assignment of roles, $\sharp P_j (i \leqslant j \leqslant i+n)$ is a preposition, $X_{ij} (i \leqslant j \leqslant n)$ is an element of the set of role-indicator $\{M, P, R, O, D, C\}$ and "k" is properly settled and registered within the dictionary of the patterns of preposition sequence.

(2) If $i=1$, the sub-sequence of (1) is denoted as $\langle\!\langle \Delta \sharp P_1\ NP_1\ \sharp P_2\ NP_2\sharp\ \sharp P_n\ NP_n \rangle\!\rangle$; α, $[X_1, X_2, \cdots, X_n]^k$. And if NP_{i+n} is a noun phrase of the end of the sentence, the sub-sequence of (1) is denoted as $\langle\!\langle \sim \sharp P_i\ NP_i\sharp\ \sharp P_{i+n}NP_{i+n} \rangle\!\rangle$; α, $[X_{i_1}, X_{i_2}, \cdots, X_{i_n}]^k$. Especially if $NP_1=\Lambda$, X_1 is Λ.

(3) If $X_{ik}=X_{i(k+1)}$, they are denoted as X_{ik}'. And if "n" is temporarily assingned to X_{ik}', what role-indicators are assigned to NP_{ik} and $NP_{i(k+1)}$, is determined by the n-th transformational rules of 3 (below).

(4) If $k=0$, then the further addition of role-indicators is not permitted whatever sequences succeed. If $k=1$, and the succeeding preposition belongs to {in, on, by}, then the role-indicator is assigned as the next sequence, else any of them are not assigned. If $k=2$, and the succeeding preposition belongs to {with, for, to, in, on, by}, then the role-indicator is assigned according as the next sequence, else any of them are not assigned.

3.3. *Transformational Rules*

Subject Recognition Algorithm. Lists of all rules are given in the Table 3.

4. *Computer Experiments and Conclusions*

This system has been described by COBOL and is now operating on FACOM 230-10. The experiment is being made on the title of papers in the IEEE Computer, the Information and Control, and so on. Several results are listed in the Table 4.

Now we consider several problems on the basis of experimental results.

(1) These are simple examples and of which results are obtained by the direct use of the transformational rule. In these examples, the algorithm works well.

Table 3. A List of Transformational Rules.

[1] $\langle\langle \# \Lambda\ NP_1\ \# \rangle\rangle$; $[M]^0$

[2] $\langle\langle \# \Lambda\ NP_1$ #OF $NP_2 \rangle\rangle$;

(i) $W(NP_1) \in U$, $[P, M]^2$

(ii) $\neg$(i) $\wedge (N(NP_1) > N(NP_2))$, $[M, R]^1$

(iii) $\neg$(i) $\wedge \neg$(ii) , $[P, M]^2$

[3] $\langle\langle \# \Lambda\ NP_1$ #OF NP_2 #OF $NP_3 \rangle\rangle$;

(i) $(W(NP_2) \in U \vee (W(NP_2) \notin U \wedge NP_2 \Rightarrow N)) \wedge NP_3 \nRightarrow N$, $[P, P, M]^2$

(ii) $(W(NP_2) \in U \vee (W(NP_2) \notin U \wedge NP_2 \Rightarrow N)) \wedge NP_3 \Rightarrow N$, $[P, M']^2$

(iii) $\neg$(i) $\wedge \neg$(ii) , $[P, M, R]^1$

[4] $\langle\langle \# \Lambda\ NP_1$ #OF NP_2 #FOR $NP_3 \rangle\rangle$;

(i) $NP_1 \Rightarrow N \wedge NP_2 \Rightarrow N \wedge (W(NP_1) \in U,\ W(NP_2) \in U)$, $[P, P, M]^2$

(ii) $NP_1 \Rightarrow N \wedge NP_2 \Rightarrow N$, $[M', O]^1$

(iii) $\neg$(i) $\wedge \neg$(ii) , $[2, O]^1$

[5] $\langle\langle \Lambda \#\ NP_1$ #ON $NP_2\ \# \rangle\rangle$;

(i) $W(NP_1) \in U$, $[P, M]^2$

(ii) $\neg$(i) , $[M, C]^0$

[6] $\langle\langle \# \Lambda\ NP_1$ #ON NP_2 #OF $NP_3 \rangle\rangle$;

(i) $NP_1 = \Lambda \wedge ((W(NP_2) \in U \wedge NP_2 \nRightarrow N) \vee (NP_2 \Rightarrow N \wedge NP_3 \nRightarrow N))$, $[\Lambda, P, M]^2$

(ii) $NP_1 = \Lambda \wedge (NP_2 \Rightarrow N \wedge NP_3 \Rightarrow N)$, $[\Lambda, M']^2$

(iii) $NP_1 = \Lambda \wedge (\neg$(ii) $\wedge \neg$(ii)) , $[\Lambda, M, R]^1$

(iv) $(W(NP_2) \in U \wedge NP_2 \nRightarrow N) \vee (NP_2 \Rightarrow N, NP_3 \nRightarrow N)$, $[P, P, M]^2$

(v) $NP_2 \Rightarrow N \wedge NP_3 \Rightarrow N$, $[P, M']^2$

(vi) $\neg$(iv) $\wedge \neg$(v) , $[P, M, R]^1$

[7] $\langle\langle \# \Lambda\ NP_1$ #ON NP_2 #OF NP_3 #OF $NP_4 \rangle\rangle$;

(i) $NP_1 = \Lambda \wedge (W(NP_2) \in U \wedge (W(NP_3) \in U)$, $[\Lambda, P, P, M]^2$

(ii) $NP_1 = \Lambda \wedge (NP_2 \Rightarrow N \wedge NP_3 \Rightarrow N)$, $[\Lambda, M', R]^1$

(iii) $NP_1 = \Lambda \wedge (\neg$(i) $\wedge \neg$(ii)) , $[\Lambda, P, M, R]^1$

(iv) $(W(NP_2) \in U \wedge W(NP_3) \in U)$, $[P, P, P, M]^2$

(v) $(NP_2 \Rightarrow N \wedge NP_3 \Rightarrow N)$, $[P, M', R]^1$

(vi) $\neg$(iv) $\wedge \neg$(v) , $[P, P, M, R]^1$

[8] $\langle\langle \# \Lambda\ NP_1$ #ON NP_2 #FOR $NP_3 \rangle\rangle$;

(i) $NP_1 = \Lambda \wedge (W(NP_2) \in U \wedge NP_2 \Rightarrow N)$, $[\Lambda, P, M]^2$

(ii) $NP_2 = \Lambda \wedge (\neg$(i)) , $[\Lambda, M, O]^1$

(iii) $W(NP_2) \in U \vee NP_2 \Rightarrow N$, $[P, P, M]^2$

(iv) $\neg$(iii) , $[P, M, O]^1$

[9] $\langle\langle \# \Lambda\ NP_1$ #FOR $NP_2 \rangle\rangle$;

(i) $W(NP_1) \in U \wedge NP_1 \Rightarrow N$, $[P, M]^2$

(ii) $\neg$(i) , $[M, O]^1$

[10] $\langle\langle \# \Lambda\ NP_1$ #FOR NP_2 #OF $NP_3 \rangle\rangle$;

(i) $W(NP_1) \in U \wedge NP_1 \Rightarrow N$, $[P, 2]$

(ii) $\neg$(i) , $[M, O']^1$

[11] $\langle\langle \# \Lambda\ NP_1$ #TO $NP_2 \rangle\rangle$;

(i) $W(NP_1) \in U$, $[P, M]^2$

(ii) $\neg$(i) , $[M, O]^1$

[12] $\langle\langle \# \Lambda\ NP_1$ #TO NP_2 #OF $NP_3 \rangle\rangle$;

(i) $W(NP_1) \in U$, , $[P, 2]$

(ii) $\neg$(i) , $[M, O]^1$

[13] $\langle\langle \# \Lambda\ NP_1$ #IN $NP_2 \rangle\rangle$; , $[M, C]^0$

[14] $\langle\langle \# \Lambda\ NP_1$ #WITH $NP_2 \rangle\rangle$; , $[M, D]^1$

[15] $\langle\langle \sim$#TO $NP \rangle\rangle$; , $[O]^1$

[16] $\langle\langle \sim$#IN $NP \rangle\rangle$; , $[C]^0$

[17] $\langle\langle \sim$#BY $NP \rangle\rangle$; , $[C]^0$

[18] $\langle\langle \sim$#ON $NP \rangle\rangle$; , $[C]^0$

[19] $\langle\langle \sim$WITH $NP \rangle\rangle$; , $[D]^1$

[20] $\langle\langle \sim$#FOR $NP \rangle\rangle$; , $[O]^1$

(2) These are examples of which structure is comparatively simple, and results are obtained by the use of combination of the transformational rules. These results are good as in (1).

(3) As the structure of sentences become complex, it is found that role-indicators can not be attached to each noun phrase by the transformational rules, and that role-indicators must be attached to a compound phrase containing preposition sequence. For examples, (i) When there exists a noun phrase of which expression is too simple in the title, it is meaningless to attach a role-indicator to such a noun phrase because of vagueness of the meaning expressed by its noun phrase. (ii) A given title is so long that a preposition sequence

Table 4. An Example of Automatic Subject Recognition on a Digital Computer.

(1) *AN EXAMPLE (P) OF SELF-ORGANIZING SYSTEM. (M)
*AN INTERNAL SORTING METHOD (M) FOR DIGITAL COMPUTERS. (O)
*GENERAL RECURSIVE FUNCTIONS (M) OF NATURAL NUMBER. (R)
*AN INTRODUCTION (P) TO STATISTICAL COMMUNICATION THEORY. (M)
*TWO DIMENSIONAL ARRAYS (M) WITH DESIRABLE CORRELATION. (D)

(2) *PROPERTIES (P) OF A NEURON (M) WITH MANY INPUTS. (D)
*UNIFORM ASYMPTOTIC THEORY (M) OF DIFFRACTION (R) BY A PLANE SCREEN. (C)
*STATISTICAL ESTIMATION (P) OF THE INTRINSIC DIMENSIONALITY (M) OF DATA COLLECTION. (R)
*A PROCEDURE (P) FOR THE DIAGONALIZATION (M) OF NORMAL MATRIX. (R)
*STRATEGIC APPROACHES (P) TO THE STUDY (P) OF BRAIN MODELS. (M)
*A NOTE (P) ON THE TWO ARMED BANDIT PROBLEM (M) WITH FINITE MEMORY. (D)
*THREE LEVELS (P) OF LINGUISTIC ANALYSIS (M) IN MACHINE TRANSMISSION. (C)

(3) *A COMPARISON (P) OF METHODS (P) FOR GENERATING NORMAL DEVIATES (M) ON DIGITAL COMPUTERS. (C)
*ON THE P-RANK OF THE DESIGN MATRIX (M) OF A DIFFERENT SET. (R)
*A NOTE (P) ON PRESERVATION OF LANGUAGE (M) BY TRANSDUCERS. (C)
*ON AN APPLICATION (P) OF DYNAMIC PROGRAMMING (M) TO THE SYNTHESIS OF LOGICAL SYSTEM. (O)
*ON THE FORMATION (P) OF A CONVERGING SHOCK WAVE (M) IN A GAS OF VARIABLE DENSITY. (C)

(4) *SOLUTION (P) OF ALGEBRAIC AND TRANSCENDENTAL EQUATION ON AN AUTOMATIC DIGITAL COMPUTER. (C)
{*M*—ALGEBRAIC EQUATION, TRANSCENDENTAL EQUATION}
*NEW FORMULAS (P) FOR COMPUTING INCOMPLETE INTEGRALS (M) OF FIRST AND SECOND KIND.
{*R*—FIRST KIND, SECOND KIND}

*SELECTION (P) AND ORDERING (P) OF FEATURE OBSERVATION (M) IN A PATTERN RECOGNITION SYSTEMS. (C)

*A GENERALIZATION (P) OF ALGOL (M) AND ITS FORMAL DEFINITIONS. (P)

(5) *MODERN COMMUNICATION PRINCIPLES (M) WITH APPLICATION TO DIGITAL (D) SIGNALING.

*SCENE ANALYSIS USING (M) BY CONCEPT OF A MODEL. (C)

*AN ALGORITHM (M) FOR THE DETERMINATION OF THE POLYNOMINAL OF (O) BEST MINIMAX APPROXIMATION TO A FUNCTION DEFINED ON FINITE (O) POINT SET.

---

contained in the title cannot be covered by any combinations of the transformational rules. In our consideration, the function of the prepositions such as "OF", "FOR", and so on is not set up, when these prepositions follow ones having a function of describing the condition such as "IN", "BY", and etc. in the preposition sequence of the title. The role-indicator, therefore, must be attached to the sub-sequence which follows the preposition such as "IN", "BY", and etc. by the transformational rules. Such a case where the role-indicator can't be assigned successfully as in (i) is caused from the expression of the title itself, then this problem does not stem from a defect of our system. On the contrary, for the improvement of case (ii) there may be designed a suitable transformational rule to be successfully applicable to segmentation of any title sentence by the more detailed analysis of data.

(4) These are some examples in which "AND" is used. As the structure of these examples in which "AND" is used is identical to the structures considered previously, the present method gives fairly sufficient result. As mentioned before, however, there exists a few examples which give unsatisfactory results because of incompletness of the analysis of "AND".

(5) These are examples of which results are unsatisfactory because of unexpected peculiar words or structures of a given title. For this reason, it is considered that the transformational rule prepared in our system are determined on the basis of only syntactic analysis. Explaining in more detailed, (i) When a given title has a deep structure, it may happen that any transformational rule cannot be applied to it. The last example in the Table 4 is a remarkable one. For such an example, it is considered that new additional transformational rules should be set up. But in such a case, there is the possibility that the role-indicators which are established in this paper fails in expressing the meaning of each noun phrase in the title. Consequently, there remains various problems

unsolved on both specification of role-indicators and construction of the transformational rules. (ii) Apart from the deepness in structure, it can't be helped that the obtained results are not satisfactory when an unexpected expression appears in a given title. The first and second example of (5) give such examples as corresponding to the above cases. For this defect similar to "ING" or "ED" words, new lists of idiom such as "with application to", "by (the) use of", etc. are registered into the dictionary. By these operations, improvements have been made on this problems.

If an algorithm which conjecture the relations between semantics of words and the role-indicators may be given, then it is easily expected that the quality of subject recognition becomes far better than the quality in our system. The informative retrieval system utilizing a natural language, therefore, must be designed by taking the semantics of a language as much as possible into considerations. Our system has many points to be improved regarding the semantics of the title sentence.