Keyword Spotting of Radio News based on Topic Identification for Speech Recognition

Yoshimi Suzuki and Fumiyo Fukumoto†and Yoshihiro Sekiguchi‡
Dept. of Electrical Engineering and Computer Science, Yamanashi University
E-mail: {ysuzuki@suwa, fukumoto@skye†, sekiguti@saiko‡}.esi.yamanashi.ac.jp
Mailing address: 4-3-11 Takeda, Kofu 400 Japan
Telephone: +81-552-20-8484 Fax: +81-552-20-8483

abstract

In this paper, we propose a method for identifying topic of radio news. In our approach, firstly, keywords which characterize each topic of newspaper are automatically extracted from newspaper articles. Then, feature vectors whose elements are χ^2 values between each keyword and each topic are calculated. Using feature vectors, a topic of each part of radio news is identified. We also propose a method of keyword spotting by using the result of topic identification of radio news. By using our keyword spotting method, the number of selected correct keywords is twice as many as the number without the method. The results of experiments show that the proposed method is robust and effective for the speech recognition system.

1 INTRODUCTION

Recently, speech recognition systems are designed for various kinds of tasks. However until now, many speech recognition systems are fixed for certain tasks, for example, a ticket vendor, a tourist information, a hamburger store, and so on. The task which consists of various kinds of topics seems to be in demand for speech recognition systems (e.g., a news dictation system, a minutes writing system for meetings, interactive information retrieval system for large area and so on).

In order to treat several kinds of topics, the system has to have large vocabulary. If a speech recognition system has to recognize words from a dictionary which has large vocabulary, the word accuracy becomes lower. In order to cope with this problem, N-gram models have been utilized for word selection from large vocabulary. N-gram models are based on statistical approach, and using N-gram models, appearance probability of words estimated automatically. However one of the problems using N-gram models is that very large corpus are necessary for recognizing discourse which consists of various topics.

Topic identification seems to be useful for keyword spotting, because if correct topic is identified, candidates of keyword can be narrowed. Some researchers are studying about topic identification of discourse [1],[2]. However there are few studies related to keyword spotting by using the results of topic identification.

In this paper, we propose a method for identifying topic of each part of radio news. In the method, feature vectors are utilized for topic identification. Each element of feature vectors is calculated based on frequency of each keyword in each topic. We use χ^2 values for elements of feature vectors. The feature vector of each topic is automatically calculated by using newspaper articles which are classified into each topic. Keywords for each topic are selected by using the feature vector of each topic. The topic which has the largest similarity between the unit of news and the feature vector of each topic is selected as topic of the unit. Our method is robust to partial phoneme misunderstanding, because whole phoneme sequence is considered for keyword spotting.

We also propose a method of keyword spotting by using the result of topic identification. Our keyword spotting method uses the most suitable keyword path which is produced in the procedure of topic identification. There are many correct keywords on the most suitable keyword path of correct topic. Therefore, the similarity between the unit and the feature vector of correct topic is larger than those of any other topics. Using our method, even if there are many words whose phoneme sequence are similar to correct keyword in the keyword dictionary, correct keywords are selected.

We explain the related work in Section 2. We describe the feature vectors that characterize each topic in Section 3. In Section 4, we show topic identification method using feature vectors. In Section 5 we report on the experiments of topic identification and keyword spotting, and finally, we discuss the possibility of utilization of extracted keywords in the most suitable keyword path for speech recognition.

2 RELATED WORK

In late years, there are many studies of topic identification which use statistical information of words in written language [3] and spoken language [1, 2, 4, 5, 6].

McDonough proposed a topic identification using switch board corpus. Also Yokoi proposed a method which uses the keywords based on statistical information. Keywords are determined by the value of mutual information between the dictionary of kana-to-kanji conversion system (Wnn) and head words of the encyclopedia of current terms (Chiezo) [7]. They reported the best number of keywords in keyword dictionary is about 800. However, when there is one keyword in a sentence,the system has to decide the topic by using only one keyword. If a phoneme recognizer can't recognize correct phonemes at a part of phoneme sequence which has a keyword, the keyword must not be extracted and correct topic must not be selected. Furthermore, the number of vocabulary must be increased for robust topic identification of short part of discourse.

Some studies for transcription of broadcast news are going to be carried out [8] [9] [10]. However there are few studies which apply topic identification method to keyword spotting.

In this paper, we propose a topic identification method of each part of radio news using feature vectors which are extracted by newspaper articles. Our method shows good performance by using a keyword dictionary which has large vocabulary. We also propose a keyword spotting method by using the result of topic identification. We have conducted the topic identification experiments and keyword spotting experiments by using the phoneme lattice which is the result of phoneme recognition. As a result of topic identification experiments, we have obtained 70% of correct ratio. The results of keyword spotting experiments demonstrate the effectiveness of our method for speech recognition.

3 FEATURE VECTOR FOR EACH TOPIC

3.1 Extraction of Feature Vector Using Newspaper Articles

In our method, each part of radio news story is classified into a topic using the feature vectors from newspaper articles which are classified into topics. Each topic is characterized by a feature vector. Each element of feature vectors was based on the frequency of each noun in newspaper articles which are classified into each topic. However the noun which appeared frequently in articles of every topic does not show the characteristics of a topic.

Term weighting method is an important issue for extracting keywords. There has been many weighting methods, such as TF (Term Frequency), IDF (Inverse Document Frequency), TF*IDF, WIDF (Weighted Inverse Document Frequency)[11],χ^2method. We used χ^2 method for weighting.

3.2 Usage of χ^2 values

In general, a topic in each discourse is characterized by words which are appeared frequently in the discourse. It is very useful to calculate frequency of each keyword for an automatic topic identification. However, all words which frequently appear do not always characterize the topic. If a word appears frequently in many topics, the word does not contribute to characterize the topic. In order to cope with this problem, term weighting by using χ^2 is used in our method.
$\boldsymbol{\chi}_k^2$ (χ^2 vector of word$_k$) is shown in formula (1).

$$\boldsymbol{\chi}_k^2 = (\chi_{k1}^2, \chi_{k2}^2, \cdots, \chi_{kn}^2) \tag{1}$$

where,

$$\chi_{kj}^2 = \begin{cases} \frac{(x_{kj}-m_{kj})^2}{m_{kj}} & \text{if } x_{kj} > m_{kj} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$m_{kj} = \frac{\sum_{j=1}^{n} x_{kj}}{\sum_{k=1}^{m}\sum_{j=1}^{n} x_{kj}} \times \sum_{k=1}^{m} x_{kj} \tag{3}$$

(m : the number of nouns, n : the number of topics, x_{kj} : frequency of word$_k$ in topic$_j$, m_{kj} : ideal frequency of word$_k$ in topic$_j$).

Ideal frequency means the frequency when the word appears in every topic with the same frequency.

Table 1 shows examples of χ^2 vector. POL, ECO, INT, SPR and ACC means politics, economy, international, sports and accident, respectively. Numerals show χ^2 values in Table 1. In Table 1, topic "politics" is characterized by "Prime Minister", and topic "international" is characterized by "President".

Table 1: Examples of χ^2 vectors

Noun	χ^2 vector				
	POL	ECO	INT	SPR	ACC
Prime Minister	<u>142.26</u>	0.00	0.87	0.00	0.00
President	0.00	0.00	<u>254.42</u>	0.00	0.00

The system identifies the topic of news story by using feature vectors. Each topic is characterized by a feature vector whose coordinate is an m-dimensional Euclidean space, where m is the number of nouns which are selected from newspaper articles. Each element of feature vectors is χ^2 value. Figure 1 shows χ^2 vector of word_k $(1 \le k \le m)$ and feature vector of each topic in χ^2 matrix. The dotted circle shows feature vector of INT and circle shows χ^2 vector of word_k. The number of element of a feature vector is the number of words of which χ^2 vector was calculated.

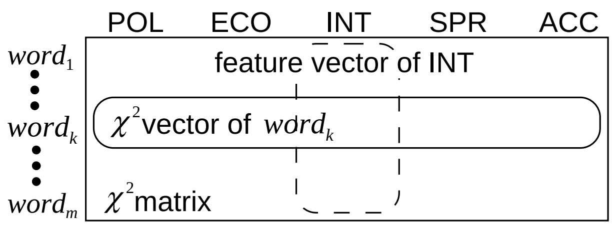

Figure 1: χ^2 vector and feature vector in χ^2 matrix

3.3 Estimation of χ^2 Vector Using Mutual Information

One of major problems of the method which is based on word frequency is data sparseness problem, i.e., the system can not identify topic, when there is no keywords in the unit. To cope with this problem, we estimate χ^2 vector of the word by using mutual information, and increase the number of keywords in feature vectors.

First, we calculated mutual information value between each noun pair in the all articles of Mainichi Shimbun '94 CD-ROM. Then, we collected pairs (α,β) which β is stored in χ^2 matrix and α is not. For each (α,β),$\boldsymbol{\chi}^2_\alpha$ is estimated by using the following formula.

$$\boldsymbol{\chi}^2_\alpha = \frac{\sum_{k=1}^{m} \boldsymbol{\chi}^2_{\beta_k} * f(\alpha, \beta_k)}{\sum_{k=1}^{m} f(\alpha, \beta_k)} \quad (4)$$

Here, m is the number of β. $f(\alpha, \beta_k)$ is co-occurrence between α and β_k in this order.

For example, the estimation of χ^2 vector of "Prime Minister" is the following three stages.

1. Calculating mutual information (Mu) between "Prime Minister" and every word which is the element of feature vectors.

Noun	POL	ECO	INT	SPR	ACC	Mu
cabinet	20	0	1	0	0	5
cabinet meeting	15	0	1	0	0	10

2. Extracting the words whose value of Mu is more than 3.

3. Calculating χ^2 vector of "Prime Minister".

$$\begin{aligned} &\chi^2 \text{ vector of "PrimeMinister''} \\ &= \frac{\sum_{k=1}^{t} m_k \times X_k^2}{\sum_{k=1}^{t} m_k} \\ &= \frac{5 \times (20,0,1,0,0) + 10 \times (15,0,1,0,0)}{5+10} \end{aligned}$$

Noun	POL	ECO	INT	SPR	ACC
Prime Minister	16.7	0	1	0	0

4 TOPIC IDENTIFICATION

In our method, topic of each unit of radio news story is identified by using feature vectors which were extracted in Section 3. Radio news stories which were used in our experiments were written in phonemes, and segmented by pauses which are longer than 0.5 second in recorded radio news. We call a part between pauses **unit**, and the system selects a topic to each unit.

4.1 Extraction of Word Candidates

Input news stories are shown by phoneme sequence without space and word boundary does not appear. The system selects maximum 20 word candidates at each phoneme as left of word candidates.

4.2 Similarity between Topic and Unit

Most words which appear frequently in newspaper articles about topic "POL" tend to appear in the unit about politics. If a word appears frequently in the topic_j, χ^2 value of the word in topic_j is large. Therefore, in a unit about POL, sum of $\chi^2_{w,\mathrm{POL}}$ tends to be large (w : a word in the unit). Then, the system selects a word sequence whose sum of $\chi^2_{k,j}$ is maximum among other word sequences at topic_j.

The similarity between the unit_i and topic_j is calculated using formula (5).

$$\begin{aligned} Sim(j,i) &= \max_{\textit{all paths}} Sim'(j,i) \\ &= \max_{\textit{all paths}} \sum_k np(\text{word}_k) \times \chi^2_{k,j} \quad (5) \end{aligned}$$

In formula (5), word_k is a word which is in word candidates obtained by Section 4.1, and each selected word doesn't share any phonemes with any other selected words. $np(\text{word}_k)$ is the number of phonemes of word_k. $\chi^2_{k,j}$ is χ^2value of word_k for topic_j. The system determines a keyword path whose $Sim'(j,i)$ is the largest among all keyword path for topic_j.

Figure 2 shows the method of calculating similarity between unit_i and $\text{topic}_{\mathrm{INT}}$. In Figure 2, there are many word paths from left to right. The system selects a path whose $Sim'(\mathrm{INT}, \mathrm{unit}_i)$ is larger than those of any other paths.

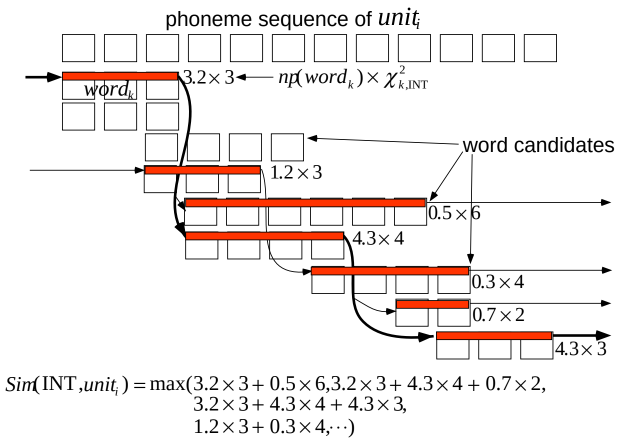

Figure 2: Calculating $Sim(\mathrm{INT}, \mathrm{unit}_i)$ similarity between unit_i and $\mathrm{topic}_{\mathrm{INT}}$

4.3 Topic Identification Process

In the topic identification process, the system identifies topic of each small unit by using $Sim(\mathrm{topic}, \mathrm{unit}_i)$ of all topics. If a similarity between a unit and a topic is larger than similarities between a unit and any other topics, the topic seem to be the topic of the unit. Therefore, the system selects the topic which is the largest of all similarities in N of topics as the topic of the unit.

5 EXPERIMENT

We have conducted topic identification experiments and keyword spotting experiments with correct phonemes and phoneme lattices which are the results of phoneme recognition.

5.1 Data

The test data we have used are radio news which are selected from NHK 6 o'clock radio news in August and September of 1995. One day news consists of about 15 stories on average. There are some articles which are hard to be classified into one topic in news stories. Therefore we select news stories which two persons classified into the same topic are selected for the experiments. The **units** which are test data of the experiments are segmented by pauses which are longer than 0.5 second. We selected 50 units for the experiments. The 50 units consist of 10 units of each topic. We used two kinds of test data. One is described with phoneme sequence. The other is written in phoneme lattice which is the results of phoneme recognition [12]. Figure 3 shows the example of the results of phoneme recognition. They are spoken by a male speaker.

q	0.0				
K	2.9				
o	2.7	y	2.7		
o	2.4	w	3.1		
o	2.1				
m	2.1	n	2.3		
o	3.0	b	3.7	w	3.9
o	2.5	N	4.0	a	4.1
q	0.0				

Figure 3: Result of phoneme recognition /kyoono/

In Figure 3, "q K o o o m o o q" shows best one phoneme and numerals show the distances between each phoneme template and voice. In each segment, the number of phoneme candidates isn't exceed 3. The following equations show the results of phoneme recognition.

$$\frac{\text{correct phonemes in phoneme lattice}}{\text{uttered phonemes}} = 95.6\%$$
$$\frac{\text{correct phonemes in phoneme lattice}}{\text{phoneme segments}} = 81.2\%$$

243 articles of Mainichi Shimbun in 1994 from CD-ROM were used in order to calculate feature vectors. There are 427 characters par an article on average. We classified these articles into 5 topics. i.e., "politics", "economy", "international", "sports" and "accident" by using classification code of CD-Mainichi Shimbun 1994. There are about 20,000 characters in each topic of the corpus (in Table 2).

Table 2: Kinds of topics

Topic	the number of articles	total number of characters
politics(POL)	34	20,840
economy(ECO)	63	21,050
international(INT)	59	20,750
sports(SPR)	37	20,133
accident(ACC)	50	21,014

In order to calculate feature vectors of each topic, 243 articles are tagged by parts of speech using JUMAN [13]. As a result, there are 534,932 nouns in the articles. From these articles, we selected 796 nouns of which the frequency is larger than 5.

Because of data sparseness problem, the system can't calculate similarity value at some units which have no keywords in the newspaper articles. In order to cope with this problem, we estimate χ^2_α (word$_\alpha$ does not appear in the 243 newspaper articles). Using mutual information, we select the keywords which co-occur with the 796 keywords which have been selected for feature vectors from the articles of newspaper in CD-ROM, and increase keywords to 9,637 in Section 3.3.

5.2 Topic Identification Experiment

In the experiments of topic identification with phoneme lattice which is the results of phoneme recognition, word candidates are extracted by using DP matching between a part of the phoneme lattice and each phoneme sequence of word in the dictionary. The topic identification method is the same as the method with correct phoneme sequence. Figure 4 shows the results of the experiments of topic identification. It shows the results of topic identification using keywords of which the minimum value of mutual information is 10,11,12,13,14,15,16,17,18,19 and 20 respectively. Table 3 shows relation between value of mutual information and the number of keywords.

Table 3: Relation between the lower limit of mutual information and the number of keywords

MvMu	10	11	12	13	14	15
NoK	9,637	9,606	9,276	8,116	6,230	4,212
MvMu	16	17	18	19	20	—
NoK	2,586	1,628	1,153	949	796	—

MvMu : the minimum value of Mutual Information
NoK : the number of keywords

The best performance is when the system uses keywords of which mutual information is larger than 16 (2,586 keywords)(70%).

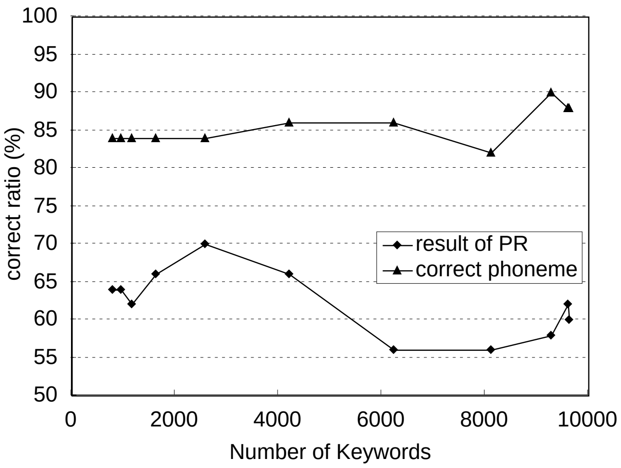

Figure 4: Identification Results

5.3 Word Spotting Experiment

We have conducted keyword spotting experiments. Figure 5 shows the result of the experiments. In Figure 5, "with TI" means the method by using topic identification. "without TI" means the method without topic identification. CP means correct phonemes and PR means the result of phoneme recognition. At each experiment, the number of keywords is slid from 796 to 9,637.

By using topic identification, the number of selected correct keywords is twice as many as the number without topic identification with 9,637 keywords in the experiments with the result of phoneme recognition.

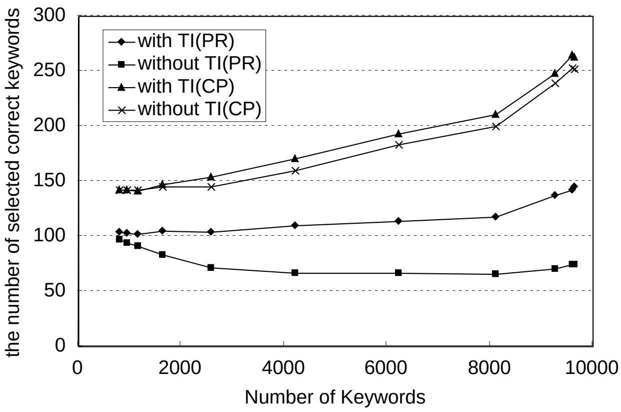

Figure 5: Performance comparison using two methods of keyword spotting: using topic identification (with TI(PR) and (with TI(CP)) and without topic identification (without TI(PR) and without TI(CP))

6 DISCUSSION

6.1 Topic Identification

Figure 4 shows the correct ratios of topic identification with the result of phoneme recognition by using various number of keywords. It shows that 70% of units are identified with correct topics by using 2,586 keywords.

For further improvement of topic identification, it is necessary to use larger corpus in order to precise feature vectors and have to improve phoneme recognition.

6.2 Word Spotting

Figure 5 shows the number of extracted keywords. It shows that the number of extracted keywords with large keyword dictionary is larger than that of small keyword dictionary. Gaps of the number of keywords between "with TI" and "without TI" with correct phonemes is small. However, the larger the number of keywords, the larger gaps of the number of keywords between "with TI" and "without TI" with the result of phoneme recognition is. The number of keywords by using "with TI" is twice as many as the number of keywords by using "without TI" with result of phoneme recognition, when the system uses 9,637 keywords.

Figure 6 shows percentage of uttered words in the most suitable keyword path which is used for selected topic in the most suitable keyword path which is used for selected topic (recall), and percentage of uttered words in the most suitable keyword path which is used for selected topic in uttered nouns (precision) at the 50 units. Using small dictionary (796 keywords), precision is about 28%, and recall is about 37%. Using large dictionary (9,637 keywords), precision is about 38%, and recall is about 22%. The result shows that the system extracted many incorrect keywords.

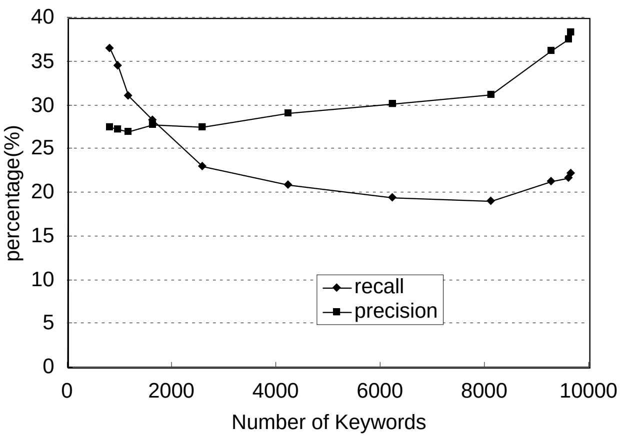

Figure 6: Recall and Precision

The system tries to find keywords for all parts of the units and extracts incorrect keywords. In order to extract only correct keywords, the system has to use co-occurrent frequency between keywords in the most suitable keyword path.

7 CONCLUSIONS

In this paper, we have proposed a topic identification method and keyword spotting method for radio news using the results of topic identification. As the results of topic identification experiments, 70% of spoken units can be identified as correct topics with 2,586 keywords dictionary (Figure 4). Using the results of topic identification, the number of selected correct keywords is larger than the number without the results (Figure 5). Using the result of topic identification, about 30% of uttered nouns can be extracted (Figure 6). We are now conducting an experiment of topic identification and keyword spotting with other news stories. In future, we will investigate how to remove incorrect nouns from extracted keywords in order to use our method for speech recognition.

8 ACKNOWLEDGMENT

The authors thank for permission to use newspaper articles Mainichi Shimbun and to use radio news Japan Broadcasting Corporation (NHK).

References

[1] J.McDonough, K.Ng, P.Jeanrenaud, H.Gish and J.R.Rohlicek, 1994. Approaches to Topic Identification on the Switchboard Corpus. *Proc. IEEE ICASSP'94, volume = 1, pp.385-388.*

[2] Kentaro Yokoi, Tatsuya Kawahara and Shuji Doshita, 1995. Topic Identification of News Speech based on Keyword Spotting. *Technical Report of IEICE,95-SLP-6-3, pp.15-20.*

[3] Kazuhide Yamamoto, Shigeru Masuyama and Naito Shozo, 1995. An Automatic Classification Method for Japanese Texts using Mutual Category Relations. *SIG-IPS Japan 106-2, pp.7-12.*

[4] Yoshiaki Itoh, Jiro Kiyama and Ryuichi Oka, Speech understanding and Speech retrieval for TV program based on spotting algorithms, *Proc. of ASJ Spring Meeting, 3-P-22.*

[5] Go Kato and Akira Kurematsu, 1995. A Topic Extraction with Keyword Spotting. *Proc. of ASJ Fall Meeting, 1-P-23.*

[6] Yoshimi Suzuki, Fumiyo Fukumoto and Yoshihiro Sekiguchi, 1996. Discourse segmentation for radio news. *ASA and ASJ Third Joint Meeting, PROCEEDINGS of the papers submitted to the ASJ, pp.1009-1014.*

[7] Shin Yamamoto, 1998. Chiezo, *Asahi Shimbun.*

[8] Baimo Bakis, Scott Chen, Ponani Gopalakrishnan, Ramesh Gopinath, Stephane Maes and Lazaros Pllymenakos, 1997. Transcription of Broadcast News - System Robustness Issues and Adaptation Techniques. *Proc. ICASSP'97, pp.711-714.*

[9] J.L.Gauvain, G. Adda, L. Lamel and M. Adda-Decker, 1997. Transcribing Broadcast News Shows. *Proc. ICASSP'97, pp.715-718.*

[10] P.C. Woodland, M.J.F. Gales, D. Pye and S.J. Young, 1997. Broadcast News Transcription Using HTK. *Proc. ICASSP'97, pp.719-722.*

[11] Takenobu Tokunaga and Makoto Iwayama, 1994. Text Categorization based on Weighted Inverse Document Frequency. *SIG-IPS Japan 100-5, pp.33-40.*

[12] Yoshimi Suzuki, Chieko Furuichi and Satoshi Imai, 1993. Spoken Japanese Sentence Recognition Using Dependency Relationship with Systematical Semantic Category. *Trans. of IEICE J76 D-II, volume11, pp.2264-2273.*

[13] Yuji Matsumoto, Sadao Kurohashi, Takehito Utsuro, Hiroshi Taeki and Makoto Nagao, Japanese Morphological Analysis System JUMAN Manual. *Nagao Lab. Kyoto University.*