Towards Building an Intelligent Call Routing System

Thien Khai Tran[1], Dung Minh Pham[2], Binh Van Huynh[3]
Faculty of Information Technology
Ho Chi Minh City University of Foreign Languages and Information Technology
Ho Chi Minh City, Vietnam

***Abstract*—This paper presents EduICR - an Intelligent Call Routing system. This system can route calls to the most appropriate agent using routing rules built by the text classifier. EduICR includes the following main components: telephone communication network; Vietnamese speech recognition; Text classifier/ Natural language processor and Vietnamese speech synthesis. To our best knowledge, this is one of the first systems in Vietnam to implement the integration mechanism of text processing and speech processing. This allows voice applications to be more intelligent, able to communicate with humans in natural language with high accuracy and reasonable speed. Having been built and tested in real environment, our system proves its accuracy attaining more than 95%.**

***Keywords*—EduICR; spoken dialog systems; intelligent call center; voice application**

I. Introduction

In recent years, Vietnamese have been seeing many remarkable publications which displayed by groups devoting to spoken Vietnamese recognition researches from Institute of Information Technology (Vietnamese Academy of Science and Technology) and University of Science, VNU-HCM. It is worth mentioning the works of Thang Vu and Mai Luong [7] as well as Quan Vu et al. [3,5]. These studies crucially concentrated on improving the efficiency of their voice recognition system, such as the Quan Vu et al. 's one which obtained the precision rate of over than 93% and this group successfully built many voice applications on this base. For example, in [5], Quan Vu et., al. successfully built VIS::DIR system, which caller can say the names of departments/offices in a university and the system will forward/redirect/route these calls to the associate agents without any help from receptionist. Nevertheless, all the applications have not been accompanied with an efficient text processing mechanism yet, which is the important mechanism in view of helping the system with understanding commands.

In this study we would like to propose some approaches to build EduICR - an Intelligent Call Routing system. This system can route calls to the most appropriate agent using routing rules built by the text classifier. In this work, we have approached two techniques: SVM (Support Vector Machine) machine learning method to classify the text commands, and DCG (Definite Clause Grammar) rule-based method [1] to deal with the syntax and semantic analysis of the text commands. Our system also has Automatic Speech Recognition module and Speech Synthesis module. Same to our approach in [8,9,10] we deal with a Vietnamese speech recognition task by using HTK (Hidden Markov Model Toolkit) [6] and speech synthesis operations by using Unit-Selection method [2].

II. System Architecture

Our system is designed with the following functions: identifying the commands via telephone; classifying the commands; routing and answering the user via telephone. An inquiring session between the system and the user is described in Figure 1.

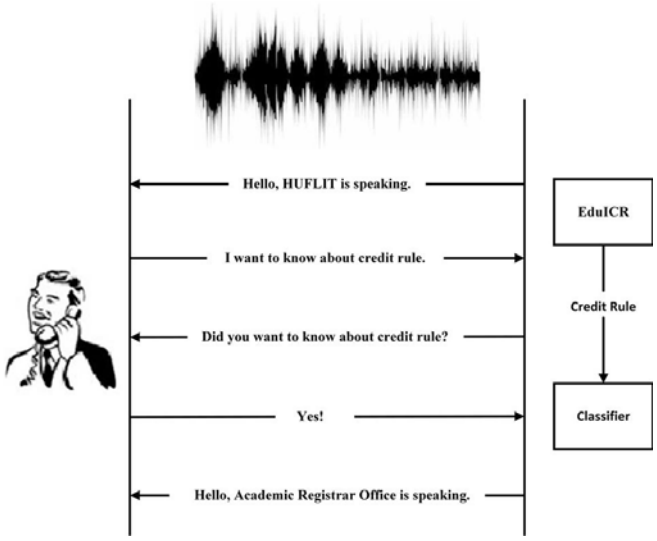


Fig. 1. An inquiring session

In order to realize the functions that given in the scenario in Figure 1, the system must consist of the following components:

- Automatic speech recognizer (ASR): to identify words that the user speaks, then convert them into text.
- Text processor: to classify the text commands to associate with agents.
- Synthesizer: to convert text to speech.

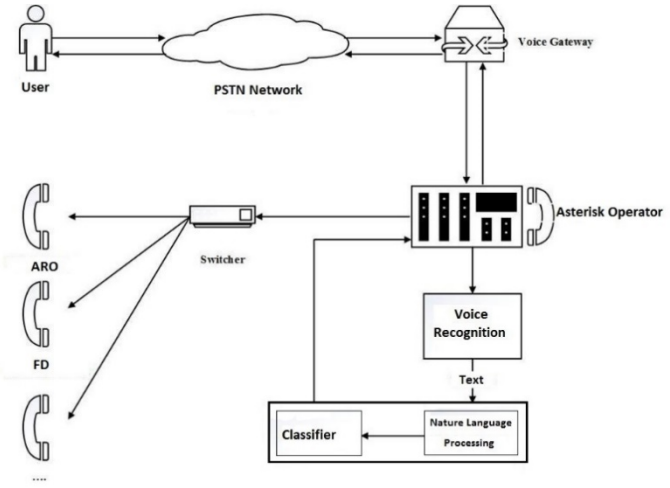


Fig. 2. Architecture of EduICR

III. The Speech Recognition

In EduICR system, we have used HTK [8] to build the Automatic Speech Recognition component. Employing the same approach as in [5,8,9,10] we have applied the context-dependent model based on tri-phone to recognize words. Besides, we have defined the tied rules for its grammar. Figure 3 shows some steps in order to create the Automatic Speech Recognizer.

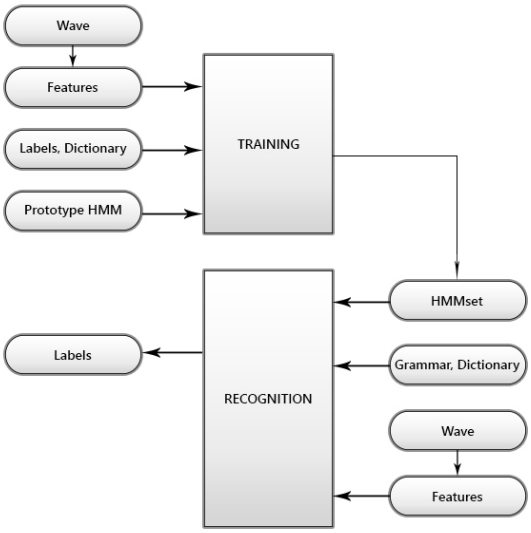


Fig. 3. Steps to build the Automatic Speech Recognizer [8]

A. Training data

There are 3,500 sentences in the speech corpus (3,500 sentences have been taken from users and they have not randomly). Total audio training covers 3 hours. All speech was sampled at 8,000Hz, 16bit by PCM format in a relatively quiet environment with 30 speakers. The lexicon comprises of 153 keywords in 3,500 sentences as being shown in Table I.

TABLE I. List Of Keywords

bao	đào	là	nhiêu	thông
bảng	đại	làm	như	thời
bảo	đăng	lạc	những	thu
bằng	điều	lại	nộp	thực
biểu	điểm	lệ	nối	thức
biết	định	liên	nợ	tiêu
bình	đối	lịch	nữa	tin
bộ	dục	lơ	ôn	tích
bổng	gặp	luận	ở	tín
cầu	giấy	lưu	phần	tính
cao	gian	luỹ	phí	tôi
các	giảm	luyện	phòng	tốt
cách	giáo	máy	phúc	trả
chất	giới	miễn	quốc	trễ
chế	hai	môn	quy	trình
chi	hạn	mở	ra	trung
chỉ	hết	mùa	rèn	trường
chính	hè	muốn	sách	tuyển
cho	hoàn	mức	sau	và
chuẩn	học	nào	sinh	vào
chương	hỏi	năm	sĩ	văn
chứng	huỷ	nay	tập	về
chuyên	khảo	ngành	tạo	việc
còn	khi	nghệ	thành	viên
công	khoa	nghiệp	thạc	vọng
có	khoản	nghiên	thế	với
của	không	ngoài	thế	xanh
cương	khoá	ngoại	theo	xét
cứu	kiện	ngữ	thẻ	yêu
đầu	kỳ	nguyện	thi	
đâu	ký	nhận	thiệu	

B. Grammar Rules / Constraints

A grammar is a set of constraints defining the phrases that a speech recognition engine can use to match speech input. Moreover, we also can provide the speech recognition engine with the predefined grammar that are included custom grammar that we create.

In addition, HTK provides the grammar definition format and an associated HParse tool that is used to build this network of words automatically. We store the grammar definition in a file called gram.txt. In our application, a part of its grammar is on the following:

```
$timetable = ( LIJCH | THOWFI GIAN | THOWFI KHOSA
BIEERU )
$graduate = ( SAU DDAJI HOJC | THAJC SIX )
$sen23 = $subject2 $target TUYEERN SINH $graduate NAWM
NAY LAF BAO NHIEEU;
$sen25 = $subject2 $timetable  OON THI CAO HOJC;
```

IV. The Speech Synthesizer

Speech synthesizer is a system that converts free text into the speech. Observably, this is a process that a computer reads out the text for people. The speech synthesis can be performed using Formant synthesis [2] or Unit-selection method [2]. With EduICR, we chose an integrative approach by Unit-selection methods, complying process as summarization in Figure 4.

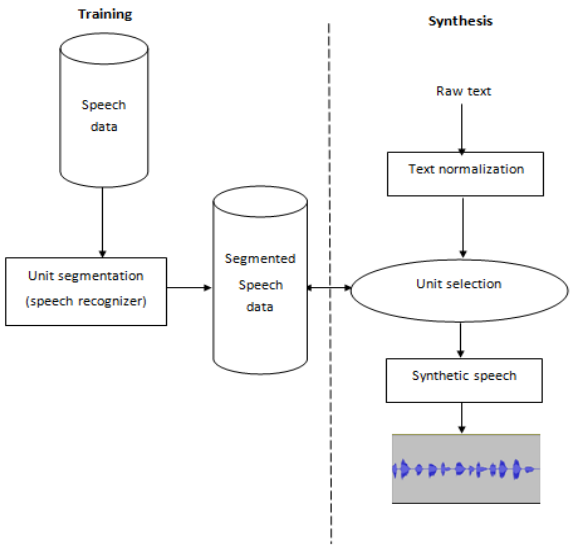


Fig. 4. Speech Synthesizer blueprint [5]

V. VIETNAMESE TEXT PROCESSING

In our system, there are 32 types of commands, being presented in Table II.

In this report, we approached two techniques: SVM (Support Vector Machine) machine learning method to classify the text commands, and DCG (Definite Clause Grammar) rule-based method to deal with the syntax and semantic analysis of the text commands.

With the Vietnam text processing, our system can easily classify/route a command/call as one of associate agents: Academic Registrar Office (ARO); Student Service Department (SSD); Finance Department (FD); Graduate Education Department (GED) and Faculty of Information Technology (FIT).

TABLE II. THE COMMAND TYPES

Type	Subject	Sentence Patterns
1	ARO	(Tôi muốn | Cho Tôi) (liên lạc | nối máy | gặp) phòng đào tạo. *(I want | Please) [to] (contact* | connect | *speak) to Academic Registrar Office.*
2	ARO	(Tôi muốn | Cho Tôi) (biết | hỏi) cách (đăng ký | huỷ | đổi) môn học [như thế nào?] *(I want | Please) [to] (know | ask | show me) how to (register | cancel | change) course.*
3	ARO	Tôi muốn nhận (chứng chỉ giáo dục thể chất | chứng chỉ giáo dục quốc phòng | bằng tốt nghiệp | bảng điểm*)? I want to get (Certificate of Physical Education | Certificate of National Defense Education | Graduate Certificate | Educational record).*
4	ARO	[Tôi muốn | Cho Tôi] [biết | hỏi] trường (tuyển sinh | xét tuyển) những ngành nào?[I want | *Please] [to] [know | ask | show me] which falcuties does HUFLIT have to be applied for.*
5	ARO	[Tôi muốn | Cho Tôi] [biết | hỏi] (điểm chuẩn | chỉ tiêu) [nguyện vọng 2] tuyển sinh đại học <năm><khoa> [là bao nhiêu?] *[I want | Please] [to] [know | ask | show me] <falcuty> 's second aspiration passing score in <year>.*
6	ARO	[Tôi muốn | Cho Tôi] [biết | hỏi] [cách tính] (điểm | điểm rèn luyện | điểm trung). *[I want | Please] [to] [know | ask | show me] [how to calculate] (Practise score | Average score).*
7	ARO	(Tôi muốn | Cho Tôi) (biết | hỏi) điều kiện xét tốt nghiệp. *(I want | Please) [to] (know | ask | show me) the Graduate requirements.*
8	SSD	(Tôi muốn | Cho Tôi) (liên lạc | nối máy | gặp) phòng tổ chức công tác sinh viên. (I want | Please) [to] (contact | connect | speak) to Student Service Department.
9	SSD	(Tôi muốn | Cho Tôi) nhận (thẻ sinh viên | giấy chứng nhận sinh viên | giấy giới thiệu thực tập) (I want | Please) [to] get (Student card | Student certificate | Practise referral form.)
10	SSD	(Tôi muốn | Cho Tôi) (biết | hỏi) [về các] [chương trình | chính sách] học bổng. (I want | Please) to (know | ask | show me) [about] scholarship [programs | policies].
11	SSD	(Tôi muốn | Cho Tôi) (biết | hỏi) [về các] (chương trình mùa tình nguyện | chương trình ngoại khoá). (I want | Please) [to] (know | ask | show me) [about] (Volunteer programs | Extracurricular programs).
12	SSD	(Tôi muốn | Cho Tôi) (biết | hỏi) điều kiện xét sinh viên 5 tốt. (I want | Please) [to] (know | ask | show me) Top Ranking Student requirements.
13	SSD	(Tôi muốn | Cho Tôi) (biết | hỏi) điều kiện xin miễn giảm học phí. (I want | Please) [to] (know | ask | show me Tuition fee (discount | free) requirements.
14	FD	(Tôi muốn | Cho Tôi) (liên lạc | nối máy | gặp) phòng tài vụ. (I want | Please) [to] (contact | connect | speak) to Finance Department.
15	FD	(Tôi muốn | Cho Tôi) (biết | hỏi) (cách | thời gian) nộp học phí. (I want | Please) [to] (know | ask | show me) (how to | when to) pay tuition fee.
16	FD	(Tôi muốn | Cho Tôi) (biết | hỏi) khi nào hết hạn nộp học phí. (I want | Please) [to] (know | ask | show me) Tuition fee payment's deadline.
17	FD	(Tôi muốn | Cho Tôi) (biết | hỏi) mức học phí của trường. (I want | Please) [to] (know | ask | show me) Tuition fee.
18	FD	(Tôi muốn | Cho Tôi) (biết | hỏi) về việc (đóng học phí trễ | hoàn trả nợ học phí | học phí học kì 3 | đóng học phí qua ngân hàng). (I want | Please) [to] (know | ask | show me) about (Late payment | Tuition fee refund | Third semester tuition fee | paying tuition fee through bank)

19	FD	(Tôi muốn | Cho Tôi) (biết | hỏi) ngoài học phí ra trường có thu các khoản phụ phí nào không. (I want | Please) [to] (know | ask | show me) extra fee of HUFLIT.
20	GED	(Tôi muốn | Cho Tôi) (liên lạc | nối máy | gặp) phòng sau đại học.(I want | Please) [to] (contact | connect | speak) to Graduate Education Department
21	GED	(Tôi muốn | Cho Tôi) (biết | hỏi) (điểm chuẩn | chỉ tiêu) tuyển sinh (sau đại học | thạc sĩ) năm nay là bao nhiêu? (I want | Please) [to] (know | ask | show me) passing score of Graduate Education admission this year?
22	GED	(Tôi muốn | Cho Tôi) (biết | hỏi) những quy định tuyển sinh thạc sĩ. (I want | Please) [to] (know | ask | show me) Master Degree Admission rules.
23	GED	(Tôi muốn | Cho Tôi) (biết | hỏi) (lịch | thời khoá biểu) ôn thi cao học. (I want | Please) [to] (know | ask | show me) Graduate Education review schedule.
24	GED	(Tôi muốn | Cho Tôi) (biết | hỏi) văn bằng hai có những ngành đào tạo nào. (I want | Please) [to] (know | ask | show me) which falcuties does Secondary Bachelor Degree have?
25	GED	(Tôi muốn | Cho Tôi) (biết | hỏi) yêu cầu ngoại ngữ đầu vào của chương trình thạc sĩ. (I want | Please) [to] (know | ask | show me) Foreign Language requirements of Master Degree program
26	FIT	(Tôi muốn | Cho Tôi) (liên lạc | nối máy | gặp) khoa CNTT. (I want | Please) [to] (contact | connect | speak) to Falcuty of Information Technology.
27	FIT	(Tôi muốn | Cho Tôi) (biết | hỏi) (điều kiện | lệ phí | thời gian) phúc khảo bài thi. (I want | Please) [to] (know | ask | show me) (conditions | fee | time) of Remarking test.
28	FIT	(Tôi muốn | Cho Tôi) (biết | hỏi) (điều kiện | lệ phí | thời gian) mở lớp học hè. (I want | Please) [to] (know | ask | show me) (conditions | fee | time) of summer courses.
29	FIT	(Tôi muốn | Cho Tôi) (biết | hỏi) (điều kiện | cách thức) bảo lưu chương trình [và theo học lại]. (I want | Please) [to] (know | ask | show me) (conditions | how to) [of] course reservation.
30	FIT	(Tôi muốn | Cho Tôi) (biết | hỏi) về (thực tập tốt nghiệp | khoá luận tốt nghiệp | nghiên cứu khoa học). (I want | Please) [to] (know | ask | show me) about (Graduate Practise | Graduate Thesis | Science Research.
31	FIT	(Tôi muốn | Cho Tôi) (biết | hỏi) khoa CNTT có những (môn | bộ môn | chuyên ngành) nào. (I want | Please) [to] (know | ask | show me) which courses does Falcuty of Information technology have?
32	FIT	(Tôi muốn | Cho Tôi) (biết | hỏi) (lịch | thời khoá biểu | thời gian học) môn tin học đại cương. (I want | Please) [to] (know | ask | show me) schedule of Informatic overview course.

A. DCG rule-base method

In our system, so as to perform semantic commands, we utilized DCG with 14 structures. All means of performance are presented in Table III.

TABLE III. THE SEMANTIC PRESENTATION OF QUESTIONS

ID	Semantic presentations
1	query(callto(Dept))
2	query(course(Task))
3	query(get(Cert))
4	query(have(Faculty))
5	query(admissions(Score,Faculty,Year))
6	query(calculate(Score))
7	query(graduate(Requirement))
8	query(policy(Scholarship))
9	query(program(Voluntter))
10	query(fee(WhQues))
11	query(schedule(Revise))
12	query(require(Subject))
13	query(fit(Course))
14	query(time(Subject))

Example 1: Cho tôi hỏi cách huỷ môn học như thế nào? (*I would like to know how to drop a course*)

The syntactic & semantic rules in DCG are defined as below:

```
query(query(Course))-->w_ques, np_course(Course),w_tail.
w_ques-->[cho, tôi, hỏi].
w_tail-->[như, thế, nào].
np_course(course(hury))-->[cách, hủy, môn, học].
```

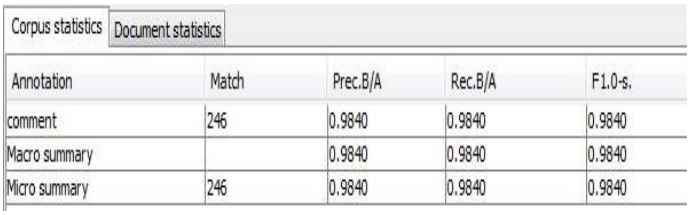
These syntactic and semantic rules determine the semantic structure of the command as: query(course(huỷ)). The semantic structure is the Structure 2 in Table III. With above semantic structures, the system will automatically convert these to the associated SQL queries.

We have done manual tests including 100 sentences for evaluating the performance of the Vietnamese processing component. They are sentences, which are found in 14 pattern rules, built with a lot of respect to the system. The latter is capable of handling all the pattern sentences. For sentences not covered by the syntactic structure, the system will return the false parsing result. These suggest that the DCG syntax rules which topics have built and dictionaries still cannot cover all the cases. If additional dictionaries from perfect DCG syntax rules, the coverage of the system will be increased to much higher level.

B. SVM Machine learning method

We collected 2,500 demo calls corresponding to 50 persons (by survey). All data was preprocessed and manually labeled as ARO, SSD, FD, GED, FIT. Next, word segmentation and POS tagger were performed, we chose vnTokenzier [4]. Then, we removed stop-words. We also removed all features whose number of occurrences doesn't meet a threshold. In this study, the threshold was set 3. Finally, the training set was vectorized and an SVM is used to compute a separating hyperplane.

250 calls were collected randomly for testing. We evaluated how well the system can identify ARO, SSD, FD, GED, FIT from the test data using the standard Precision, Recall and F-score measures. Figure 5 shows the results of the system running on test data with threshold = 3.



Corpus statistics | Document statistics

Annotation	Match	Prec.B/A	Rec.B/A	F1.0-s.
comment	246	0.9840	0.9840	0.9840
Macro summary		0.9840	0.9840	0.9840
Micro summary	246	0.9840	0.9840	0.9840

Fig. 5. The results of the system running on test data with threshold =3

VI. Experiments and Evaluation

The first test is performed on Speech Recognition. Next, we perform experiments on the system, as well as implement the perception survey/ assessment of users on the system, including speech synthesis component.

A. Speech Recognizer

The speech recognition performance is typically evaluated in terms of Word Error Rate (WER), which can then be computed as: WER= (S + D + I) / N x 100% [6], where N is the total number of words in the testing data, S denotes the total number of substitution errors, D is the total number of deletion errors and I is the total number of insertion errors. We make use of Word Accuracy (WA) [6] instead, which is computed as WA = (1 – (S + D + I)/ N) x 100%, to report performance of the speech recognizer. The accuracy of the system is reported in Table IV.

TABLE IV. Test Result by Capacity of Corpus

Model	**Descriptions**	**Result (accuracy)**	
		Trained users	***Untrained users***
VNSS _C10	Train corpus of 10 speaker	99%	93%
VNSS _C20	Train corpus of 20 speakers	98%	94%
VNSS _C30	**Train corpus of 30 speakers**	**97%**	**95%**

B. Investigation

We have also performed the survey to 50 users using the system with the question: "The system is easy to use or not?" with 4 levels of evaluation, and the results in Table V.

TABLE V. Comfort Results

Very comfort	**Fairly comfort**	**A bit comfort**	**Not comfort**
28%	24%	26%	22%

C. System Experiments

Text processing: DCG rule-based method is capable of handling all the pattern sentences but for sentences not covered by the syntactic structure, the system will return the false parsing result. So we chose SVM machine learning for text classifier.

TABLE VI. Experimental Environments

Number of Questions	100
Environment	in-door
Sampling rate	8 kHz
Quantization	16 bits
Format	PCM

The system correctly analyzes and executes 95/100 of the spoken questions in Vietnamese language. The fault cases must be remained at the speech recognition step. So, our system demonstrates its accuracy attaining 95%. About 3.4 seconds for a command is spent as the average feedback time of the system.

VII. Conclusion

This paper has presented the architectural model of EduICR system as well as our approach to build it. In our opinion, this is one of the first systems in Vietnam equipped with a mechanism for text processing efficiency in voice applications. This study also opens up a new direction for the construction and development of systems inquiry that can understand and communicate with Vietnamese speaking users. Our upcoming improvement is to enrich the routing rules and widen application-oriented vocabulary.

Acknowledgment

This work is supported by Ho Chi Minh City University of Foreign Languages and Information Technology.

References

[1] Fernando C. N. Pereira and Stuart M. Shieber, Prolog and Natural-Language Analysis. Microtome Publishing, pp. 1 – 284, Massachusetts, 2005.

[2] Hunt, A. Black and W. Alan, “Unit selection in a concatenative speech synthesis system using a large speech database," Pro c. ICASSP-96, 1, pp. 373, 1996.

[3] Nhut Pham and Quan Vu, “A Spoken Dialog System for Stock Information Inquiry,” in Proc. IT@EDU, Ho Chi Minh City, Viet Nam, 2012.

[4] Phuong Le-Hong, Minh-Huyen Thi Nguyen, Azim Rouss-analy, and Tuong-Vinh Ho (2008), A Hybrid Approach to Word Segmentation of Vietnamese Texts. Language and Automata Theory and Applications, page 240, 2008.

[5] Quan Vu and Cuong Le, (2012). “Voice Server and its applications”. Technical report, Research project, HCM City Department of Science and Technology, Viet Nam.

[6] Steve Young et al, The HTK Book (version 3.4). [Online]. Available: www.htk.eng.cam.ac.uk/docs/docs.shtml, 2006.

[7] Thang Vu, Mai Luong (2012), “The Development of Vietnamese Corpora Toward Speech Translation System”, RIVF-VLSP 2012, Ho Chi Minh City, Viet Nam, 2012.

[8] Thien Khai Tran, Dang Tuan Nguyen (2013). “Semantic Processing Mechanism for Listening and Comprehension in VNSCalendar System”. International Journal on Natural Language Computing (IJNLC) Vol. 2, No.2, April 2013.

[9] Thien Khai Tran, Tien Cat Khai Tran, Tho Anh Mai, Nhat Minh H. Nguyen and Hien Thanh Vu (2014), “EDUVoice - a system for querying academic information via PSTN”, The Third Asian Conference on Information Systems (ACIS 2014). Nha Trang, 2014.

[10] Thien Khai Tran (2015), “SentiVoice - a system for querying hotel service reviews via PSTN”, IEEE-RIVF 2015, Can Tho, Viet Nam, 2015.