Journal of Computing and Information Technology - CIT 18, 2010, 4, 309–316
doi:10.2498/cit.1001916

A Croatian Weather Domain Spoken Dialog System Prototype

Ana Meštrović, Luka Bernić, Miran Pobar, Sanda Martinčić-Ipšić and Ivo Ipšić

Department of Informatics, University of Rijeka, Croatia

Speech technologies and language technologies have been already in use in IT for a certain time. Because of their great impact and fast growth, it is necessary to introduce these technologies for Croatian language. In this paper we propose a solution for developing a domain-oriented spoken dialog system for Croatian language. We have chosen a weather domain because it has limited vocabulary, it has easily accessible data and it is highly applicable. The Croatian weather dialog system provides information about weather in different regions of Croatia. The modules of the spoken dialog system perform automatic word recognition, semantic analysis, dialog management, response generation and text-to-speech synthesis. This is a first attempt to develop such a system for Croatian language and some new approaches are presented.

Keywords: spoken dialog system, automatic speech recognition, speech understanding, semantic analysis, dialog manager, text-to-speech

1. Introduction

The development of a spoken dialog system concerns solutions to speech recognition problems, as well as speech understanding and human machine interaction problems. The major problems in the development of a continuous speech understanding systems arise due to the nature of the spoken language: there are no clear boundaries between words, since the phonetic beginning and ending of words are influenced by neighboring words; additionally, variability in speech between different speakers can be noticed, and the speech signal may be affected by noise. There are also problems with ambiguity of words with different meanings and anaphoras appearing in the text. To avoid these difficulties, spoken dialog systems are usually limited by different constraints: the vocabulary size is about one thousand words, the communication domain is task-oriented, and the sentence structure is usually limited by a simple grammar. The communication domain restriction enables fully interpretable semantic description of the domain which provides usability with high accuracy.

Several weather domain conversational systems were developed for English [8], Japanese [9] and Slovene [6, 11]. In these systems the domain knowledge is captured in semantic frames [4, 5, 8] and the sentence understanding is performed using simple context-free grammars (CFG) with parsing [5], shallow parsing [8, 9], context language models or statistic modeling [7]. The speech generation is performed using CFG or response sentence templates [8, 10].

For Croatian language, there were no spoken dialog systems developed so far and speech technologies and language technologies were not employed enough. The reasons for that situation are well known: Croatian language is spoken by a limited number of people, it is not a widespread language and it has a specific and complex structure. Our motivation was to apply the existing technologies, adopt them for Croatian language and develop a spoken dialog system for the weather domain. In this paper, we present all specific problems that arise in Croatian language modeling in the process of designing a Croatian weather domain spoken dialog system.

The modules of a spoken dialog system perform word recognition, semantic analysis, dialog management, response generation and spe-

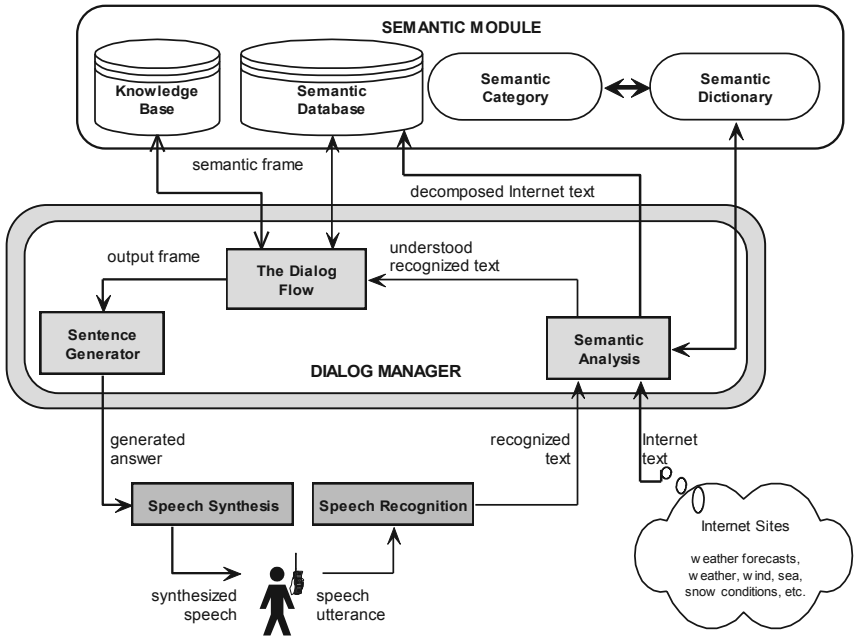

Figure 1. Spoken dialog system overview.

ech synthesis as shown in Figure 1. User utterances are firstly digitized and transformed into a sequence of speech signal feature vectors. The sequence is passed to the word recognition module, which generates hypotheses of spoken word chains. The recognized words are handed over to the semantic module, which extracts a set of semantic concepts from the recognized words.

These concepts are passed on to the dialog manager. The dialog manager is responsible for all dialog actions. According to the dialog history and dialog strategy, the user is asked to confirm the parameters, or additional parameters are requested. If enough parameters are available, the requested data is queried from a semantic database. Database query results are transformed into sentences and using the speech synthesis module played back to the user over the telephone line.

The Croatian weather dialog system provides information about weather in different regions of Croatia and for different time periods, collecting the weather data from the available websites over the Internet [1]. The dialog system has to recognize and understand the spoken queries and it has to generate answers. The user can access information from a dialog system by telephone calls. Therefore the system has to be able to recognize, understand and process telephone speech as well as speech produced in good acoustical environments [2]. The response to the user query should be speech of high synthesized quality as well [12].

The second section of this paper describes the semantic component of the dialog system. In the third section, the dialog manager is described. Section four presents the frontal components of the system: automatic speech recognition and speech synthesis modules. In conclusion, some possible improvements are discussed and future work plans are presented.

2. The Semantic Analysis

A deductive object-oriented logic programming language, F-logic, is used for the semantic analysis of weather forecast data within the weather spoken dialog system. F-logic provides a natural way of defining a conceptual model of data semantics. It enables hierarchical representation and provides frames definition. Moreover, all rules for semantic analysis are defined in F-logic and thus implemented in the Flora-2 system.

The proposed semantic analysis for Croatian data is conducted through three main phases [1]. As displayed in Figure 2, the semantic context for the input text is determined in the first

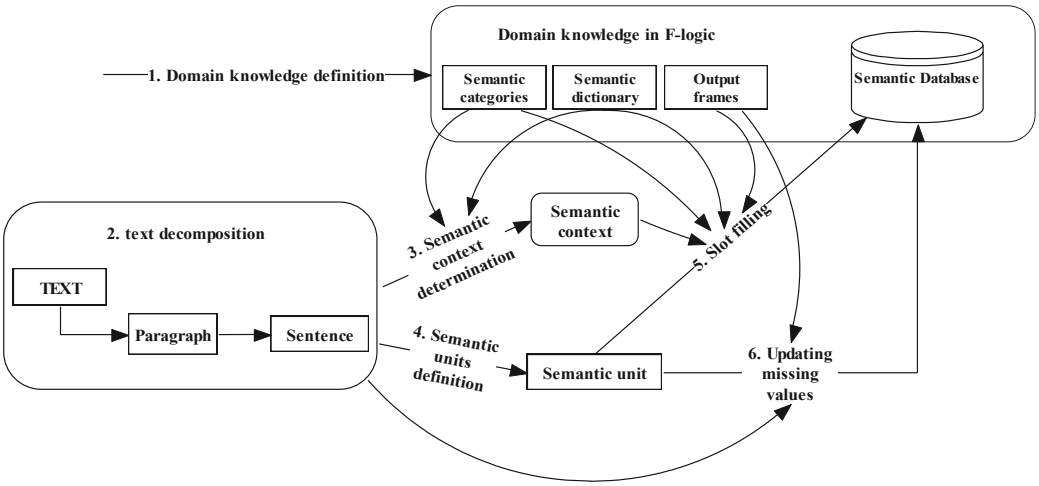

Figure 2. Semantic analysis in F-logic.

phase. In the second stage, the input is decomposed into semantic units. Semantic units are analyzed and semantic database slots are filled. Since some input data slots are missing in the third phase the incomplete data is updated with missing values. Semantic data analysis is based on a previously defined dictionary, phrases, semantic categories, and semantic database structure.

2.1. Semantic representation of the weather domain

In terms of a spoken dialog system, speech understanding mainly relies on the semantic and linguistic interpretation of the recognized utterance. The semantic interpretation of a weather domain in terms of information extraction is enabled with a semantic and knowledge database. The weather domain knowledge is captured in semantic categories, the semantic dictionary and output frames.

2.1.1. Semantic categories

Each word from the weather domain is associated with a semantic category. Semantic categories are represented in F-logic language as classes and subclasses. There are 10 main semantic categories in the dictionary: *the weather forecast, the biometeorology forecast, meteorology, the state of the river, the wind, the temperature, the place, the time, the description and an irrelevant category*. Each of these categories consists of semantic subcategories. Overall there are 36 basic semantic subcategories.

2.1.2. Semantic dictionary

According to the semantic categories of all words in a sentence it is possible to define the semantic context of a whole sentence. For example, the semantic context of the sentence can be wind, temperature, the level of the river water, etc.

The word dictionary was prepared from collected weather-related texts and consists of almost 2300 different words. As the Croatian language is highly flective, the dictionary comprises all word forms that occur in the data.

2.1.3. Semantic database

The semantic database captures weather knowledge and weather data used for answer generation through the dialog manager. The semantic database structure is a hierarchy of semantic frames with slots [1] as shown in Figure 3.

There are six generic types of frames: weather, temperature, wind, sea, visibility and meteorology. Besides the sea weather forecast and the weather forecast, the biometeorological forecast, the level of river water and other types of forecasts are included in the domain as well. Time and place slots are common for every

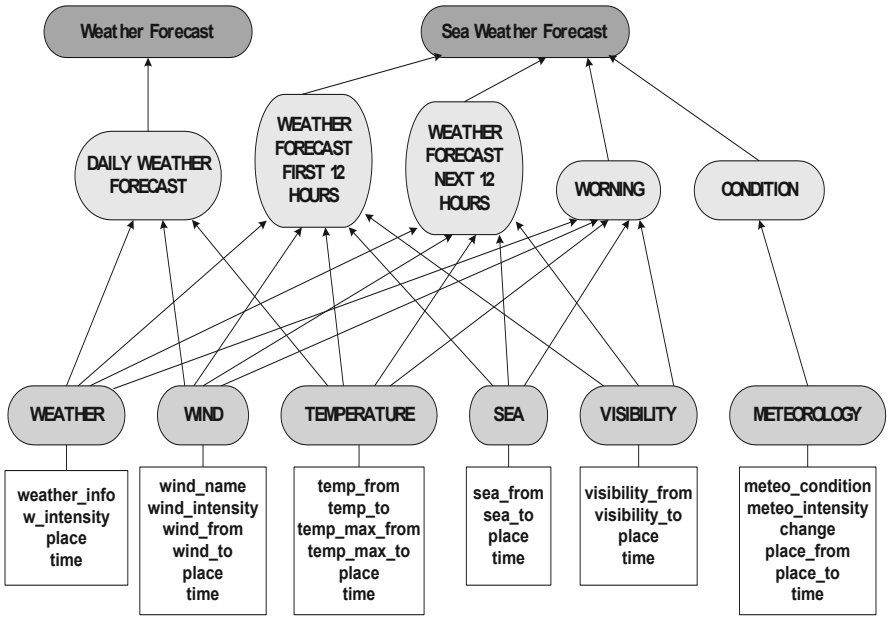

Figure 3. The semantic database structure.

frame. Other slots like wind name are dependent upon the frame type.

2.1.4. Knowledge base

The knowledge base is an extension of the semantic database with the knowledge of the selected domain in the form of rules. The rules represent expert knowledge about the domain, ranging from the highly general to the ones strictly relevant for meteorology. Geographical hierarchies and time relations are captured in the knowledge database, the transformations from different metric systems (for example: knots to kilometers per hour), Beaufort scale for wind and waviness of the sea etc.

The F-logic rule shows one time-related rule which derives the time of the day from the previous part of the sentence:

```
_X1[time->Time]:-
                _X1[time->not_defined],
                _X[time->Time].
```

3. The Dialog Manager

The programming language chosen to implement the dialog manager is the Common Lisp dialect of the Lisp programming language family. A side-benefit of using LISP is the unified syntax and unified representation of data. Frames are represented as LISP S-expressions (symbolic expressions) which are the fundamental syntax for representing data and code in LISP [14].

Well-known approaches to dialog management are grammar-based (frame), plan-based and, more recently, information state update approaches [3]. The grammar-based approach views the dialog as a series of exchanges or dialog acts that follow regular patterns such as question/answer, greeting/greeting etc. These patterns are represented with dialog grammars, which are used to parse the dialog structure. Plan-based systems model the dialog as an interaction between rational agents that strive to achieve certain goals. To reach these goals, agents plan actions (speech acts and real-world) to be taken by logical inference using models of other agent's belief, desires and intentions and recognized plans. The information state update approach uses a notion of information state, a model of dialog participants' knowledge, dialog history, conventions etc., a set of dialog moves that can be performed and a set of rules that specify how a dialog move affects the information state [7, 8, 9]. Selection and production of dialog moves is also specified by a set of rules.

The dialog manager is responsible for all dialog actions and the dialog strategy. There are three possible dialog strategies: system, mixed and user initiative. In our dialog prototype, the mixed initiative strategy is implemented. Following the initial greeting message from the system, the user is free to respond in any way (initially the user has initiative). The dialog manager then infers the semantic category from the user's utterance and prompts the user for the data that is missing to complete the query. In this stage the system holds the initiative and the user can only respond to the system prompts. For a more flexible dialog, more user actions such as restarting or changing some parts of information already given could be allowed in the future versions of the spoken dialog system.

Figure 4 displays the architecture of the dialog manager module. The dialog manager is realized as a three-layer structure [13]. At the topmost layer is the dialog flow control module which is responsible for the dialog system behavior through all interactions with a user.

The frame layer is responsible for semantic interpretation of the user utterance, gathering the data from the semantic layer consisting of the semantic dictionary, semantic database and expert knowledge base as presented in previous chapter. The frame composer is responsible for composing an output frame. Finally, the speech generation module transfers the output frame into a real world sentence that can be presented to the user.

3.1. The dialog flow layer

The dialog flow layer of the dialog manager is realized as a finite state machine. The user input processing is done at the level of complete utterances. It starts by issuing a greeting message and finishes when the user has stopped interacting with the system. The system exploits user input to gradually refine queries to the database layer, resulting in a natural language dialog between the user and the system.

In the current prototype version, only basic states are modeled: Opening Formalities, User Request Understanding, Information Gathering, Response Construction, Sentence Generation, and Closing Formalities.

3.2. The frame layer

The architecture of the dialog manager follows a three-tiered model. At the center of the model is the frame composing and decomposing engine, which is tasked with connecting the knowledge database and the user interface layer by composing and decomposing frames according to requests received from the user interface layer and notifications of change from the database layer. The frame composer/decomposer is realized as a set of purely functional operators based on elementary functional operations "compose" and "decompose". Inside the system, the output frames, as well as all the other data relevant to the frame, are represented as LISP S-expressions.

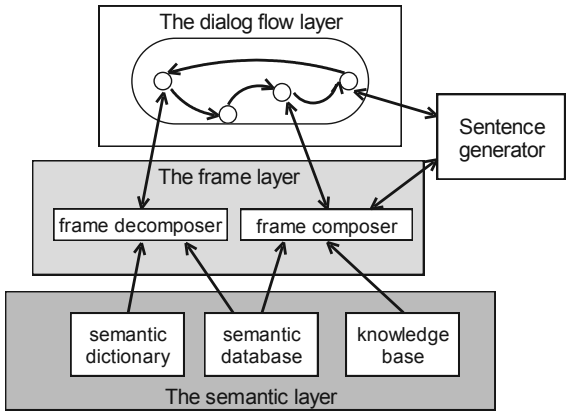

Figure 4. The dialog manager module.

3.3. The sentence generator

In the dialog prototype the sentence generator from the composed output frame constructs a Croatian sentence using a set of predefined sentence templates. Since there are a definite number of templates, correct word forms (case, gender, tense) are associated with the template. This way, programming of a Croatian morphological generator is avoided in this stage of the prototype development. The construction of a Croatian morphological generator is planned for the future.

An example template for Croatian sentence generation is shown below. The output frame slots are represented by their semantic category in line brackets notation

```
(‘‘Puhat će’’ ‘|jačina_vjetra|
    ‘‘do’’ ‘|jačina_vjetra|
‘|strana_vjetra| ‘|naziv_vjetra|).
```

4. The Frontal Dialog Modules

The frontal components of the system, automatic speech recognition and speech synthesis modules enable users spoken access to weather information. User utterances are first digitalized and transformed into a sequence of speech signal feature vectors. The sequence is passed to the word recognition module, which generates hypotheses of spoken word chains.

4.1. Speech recognition

The Croatian automatic speech recognition system is based on continuous hidden Markov models of monophones and triphones trained with the HTK Toolkit [2]. The monophone models with continuous Gaussian output probability functions were trained for the 30 standard Croatian phonemes and 4 additional models for silence, breathing (inspiration) sound, mispronounced words, hesitations and noise. Each monophone model consists of 5 states, where the first and last states have no output functions. The initial training of the Baum-Welch algorithm on HMM monophone models resulted in a monophone recognizer, which was used for the automatic segmentation of the speech signals. The automatically segmented speech database was used to model triphone models, where each phone model was extended by its left and right context.

The triphone models consist of 5 states with continuous density output functions (one to twenty mixture Gaussian density functions), described with diagonal covariance matrices. The state tying was performed, due to the lack of the acoustic material, using proposed Croatian phonetic rules [2]. The number of mixtures of output Gaussian probability density functions per state was increased up to 20 in the used triphone recognizer.

For speech recognition, the speech signal feature vectors consist of 12 mel-cepstrum coefficients and their derivatives and acceleration. The feature coefficients were computed every 10 ms for a speech signal frame length of 20 ms. The HMMs were trained using the VEPRAD speech corpus. Two recognition systems were trained separately: one with the radio speech about weather forecast spoken by professional speakers at the national radio (VEPRAD radio) and the other with the telephone speech about weather reports given by meteorologists (VEPRAD telephone) [2].

In all experiments a backoff bigram language model was used. Estimated perplexity of the VEPRAD radio bigram language model is 11.17 (1462 different words). Perplexity of the VEPRAD telephone bigram model is 18.09 (1788 different words). The achieved word error rate (WER) for the weather forecast task was 4.5%, while the word error rate increased to 10% for the telephone speech.

4.2. Speech synthesis

The text-to-speech system for Croatian is based on diphone concatenation using the TD-PSOLA algorithm. An utterance is synthesized by overlapping and adding the waveforms of diphones that correspond to the target phoneme string. A diphone is an acoustic unit spanning two phonemes from the middle of the first to the middle of the following, which captures the transition between phonemes [12].

A diphone database was built for the system, using parts of the same weather and news domain that radio corpus used in training the speech recognizer. In total, 2331 recordings or 2 h and 26 min of speech from one speaker were used in preparation of the database. The sound files were segmented using the previously described automated speech recognition system that generated phone labels and automatically determined time segments of each phone. In addition to the 30 standard phonemes in the Croatian language, accented variants of vowels (a:, e:, i:, o: and u:), sound r as a vowel and silence as a phoneme are distinguished, 37 phonemes in total. Out of 1396 possible diphones, 1057 were represented with at least one instance. Duplicate instances of diphones were discarded, keeping an instance extracted from the middle of the word where intonation was expected to be more neutral and of average duration.

The synthesizer and the supporting grapheme-to-phoneme translation were implemented in MATLAB. The grapheme-to-phoneme translation converts an ordinary text string into a string of phoneme labels from the phoneme set used in the diphone database which the synthesizer accepts. A phonetic dictionary with approximately 10000 words and a fallback set of replacement rules are used for this.

5. Conclusion

This paper presents the current status of a Croatian spoken dialog system prototype. The modules of the spoken dialog system perform automatic word recognition, semantic analysis, dialog management, response generation and text-to-speech synthesis. This prototype is weather domain related and users can retrieve information about weather in different regions of Croatia and for different time periods.

The Croatian automatic speech recognition system is based on continuous hidden Markov models of monophones and triphones. Word error rate (WER) of the telephone speech is 10%.

Domain knowledge representation and semantic analysis is implemented in F-logic using the Flora-2 system. The domain data semantics is captured in a semantic dictionary, semantic categories and output frames. The proposed semantic analysis of the Croatian language implements a slot filling technique in three stages. The knowledge database extends the semantic database with the expert rules.

The dialog manager is realized in three layers: the dialog flow control layer which is responsible for all interactions with the user, the frame layer which is responsible for semantic interpretation of the user utterance and for composing the output frame and the data layer. The speech generation module transfers the output frame into a real world sentence that can be presented to the user. The dialog manager is implemented as a finite state machine in Common Lisp.

The generated sentence is passed to the text-to-speech system, which is based on diphone concatenation using the TD-PSOLA algorithm. The synthesizer and the supporting grapheme-to-phoneme translation were implemented in MATLAB.

The main goal of the prototype development is connecting all dialog modules, so the user can interact with the weather dialog system. This prototype will be implemented for gathering dialogs and possible interaction strategies with users. According to the gathered data, the dialog manager will be improved: the dialog strategy will be shifted from system-driven to user-driven, the dialog manager will cover all listed scenarios, and the language generation module will be extended with Croatian morphological information, so the correct form of the generated words can be derived automatically by using a morphological generator. For a more flexible dialog, more user actions such as restarting or changing some parts of information retrieval should be considered. In the future, we will also consider the use of formal grammars in order to improve the updating of incomplete data with missing values. We henceforth plan to further expand the domain of interest.

References

[1] A. Meštrović, S. Martinčić-Ipšić, M. Čubrilo, Weather Forecast Data Semantic Analysis. *Journal of Information and Organization Sciences, JIOS*, Vol. 31(1), (2007) pp. 115–129.

[2] S. Martinčić–Ipšić, S. Ribarić, I. Ipšić, Acoustic Modelling for Croatian Speech Recognition and Synthesis. *Informatica*, Vol. 19(2), (2008) pp. 227–254.

[3] S. Larsson and D. R. Traum, Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural language engineering*, Vol. 6, (2001) pp. 323–340.

[4] H. Hardy, et al., The AMITIÉS system: Data-driven techniques for automated dialogue. *Speech Communication*, Vol. 48, (2006) pp. 354–373.

[5] B. Souvignier, et al., Strategies for Spoken Dialogue Systems, *IEEE Trans. Speech and Audio Processing*, Vol. 8, No. 1, (2000) pp. 51–62.

[6] M. Hajdinjak, F. Mihelić, A Dialogue-management Evaluation Study, *CIT*, Vol. 15(2), (2007) pp. 111–121.

[7] M. Rayner, et al., A Methodology for Comparing Grammar-based and Robust Approaches to Speech Understanding. *Interspeech 2005*, (2005) pp. 1877–1880.

[8] V. ZUE, ET AL., JUPITER: A Telephone-based Conversational Interface for Weather Information. *IEEE Transactions on Speech and Audio Processing*, Vol. 8, No. 1, (2000) pp. 85–96.

[9] M. NAKANO, ET AL., Mokusei: A Telephone-based Japanese Conversational System in the Weather Domain. *EUROSPEECH 2001*, (2003) pp. 1331–1334.

[10] O. GALIBERT, G. ILLOUZ, S. ROSSET, Ritel: An Open-domain, Human-computer Dialogue System. *Interspeech 2005*, (2005) pp. 909–912.

[11] J. ŽIBERT, ET AL., Development of a Bilingual Spoken Dialogue System for Weather Information Retrieval, *EUROSPEECH 2003*, Vol. 1, (2003) pp. 1917–1920.

[12] M. POBAR, S. MARTINČIĆ–IPŠIĆ, I. IPŠIĆ, Text-To-Speech Synthesis: A Prototype System for the Croatian Language. *Engineering Review*, Vol. 28(2), (2008) pp. 31–44.

[13] L. BERNIĆ, S. MARTINČIĆ-IPŠIĆ, I. IPŠIĆ, A Croatian Spoken Dialogue Manager Proto-type. *MIPRO 2008*, (2008) Proc. 199–203.

[14] J. MCCHARTHY, Recursive Functions of Symbolic Expressions and Their Computation, Part I, *Communications of ACM*, Vol. 3(4) (1960) pp. 184–195.

Received: June, 2010
Accepted: November, 2010

Contact addresses:

Ana Meštrović, Ph.D., Assistant
University of Rijeka, Department of Informatics
Omladinska 14, 51000 Rijeka, Croatia
e-mail: amestrovic@uniri.hr

Luka Bernić
University of Rijeka, Department of Informatics
Omladinska 14, 51000 Rijeka, Croatia
e-mail: lbernic@ffri.hr

Miran Pobar, B.Sc., Assistant
University of Rijeka, Department of Informatics
Omladinska 14, 51000 Rijeka, Croatia
e-mail: mpobar@uniri.hr

Sanda Martinčić–Ipšić, Ph.D., Assistant Professor
University of Rijeka, Department of Informatics
Omladinska 14, 51000 Rijeka, Croatia
e-mail: smarti@uniri.hr

Ivo Ipšić, Ph.D., Full Professor
University of Rijeka, Faculty of Engineering
Vukovarska 58, 51000 Rijeka, Croatia
e-mail: ipsic@riteh.hr

ANA MEŠTROVIĆ was born in Rijeka, Croatia. In 2001 she graduated in mathematics and computer science from the University of Rijeka, Faculty of Arts and Sciences. She received M.Sc. and Ph.D. degree in computer science from the University of Zagreb, Faculty of Organization and Informatics Varaždin, in 2005 and 2009 respectively. She works as an Assistant Professor at the University of Rijeka, Department of Informatics. Her research interests include knowledge representation, semantic Web, semantic analysis and spoken dialog systems.

LUKA BERNIĆ was born in Rijeka. He graduated from high school in the USA. Currently he is a student of informatics and philosophy at the University of Rijeka, Department of Informatics.

MIRAN POBAR was born in 1983 in Rijeka, Croatia. In 2007 he obtained his B.Sc. degree in electrical engineering from the University of Rijeka, Faculty of Engineering. He currently works at the University of Rijeka, Department of Informatics, as an Assistant. In 2008 he commenced a Ph.D. postgraduate study programme at the University of Zagreb, Faculty of Electrical Engineering and Computing. His research interests include speech synthesis, speech recognition and speech technologies for Croatian language.

SANDA MARTINČIĆ-IPŠIĆ obtained her B.Sc. degree in computer science in 1994, from the University of Ljubljana, Faculty of Computer Science and Informatics and her M.Sc. degree in informatics from the University of Ljubljana, Faculty of Economy in 1999. In 2007 she obtained the Ph.D. degree in computer science from the University of Zagreb, Faculty of Electrical Engineering and Computing. She currently works as an Assistant Professor at the University of Rijeka, Department of Informatics. Her research interests include automatic speech recognition, speech synthesis, speech corpora development and spoken dialog systems, with a special focus on the Croatian language.

IVO IPŠIĆ obtained his B.Sc., M.Sc. and Ph.D. degrees in electrical engineering from the University of Ljubljana, Faculty of Electrical Engineering, in 1988, 1991 and 1996, respectively. From 1988–1998 he was a staff member of the Laboratory for Artificial Perception, at the University of Ljubljana, Faculty of Electrical Engineering. Since 1998 Ivo Ipšić has been a professor of computer science at the University of Rijeka, teaching computer science courses. His current research interests lie within the field of speech and language technologies.