Stagii de practică în cercetare – Laboratorul SpeeD

Laboratorul de cercetare SpeeD (speed.pub.ro), condus de prof. Corneliu Burileanu, propune studenților din anul III un stagiu de practică în cercetare pentru vara anului 2014. Stagiile de practică presupun lucrul la o temă de cercetare în tehnologia vorbirii, timp de 12 săptămâni (iunie-august 2014), în regim part-time (6 ore pe zi), în Laboratorul SpeeD. Proiectele de practică se vor realiza sub îndrumarea domnilor Horia Cucu și Andi Buzo și se vor finaliza prin realizarea unei aplicații practice și scrierea unui raport de cercetare (referat de practică).

Participarea la un stagiu de practică în cadrul Laboratorului SpeeD vă va oferi oportunitatea de a contribui la dezvoltarea unei aplicații în tehnologia limbajului vorbit sub îndrumarea unui grup de specialiști. De asemenea, proiectul de practică va putea fi utilizat (sau extins) pentru lucrarea de licență din vara anului 2015.

Miercuri, 5 martie, ora 18:00, în Sala de Calculatoare nr. 3, le propunem studenților interesați o întâlnire, cu scopul a discuta despre aceste stagii de practică și a le răspunde la eventualele întrebări.

Înscrierea la stagiile de practică se va face în perioada 3 – 14 martie, trimițând prin email la adresa horia.cucu@upb.ro un scurt Curriculum Vitae și specificând:

- numele și grupa din care faceți parte,
- motivul pentru care doriți să participați la stagiile de practică SpeeD,
- dacă ați participat sau nu la sesiunile de înregistrări audio realizate de Laboratorul SpeeD în toamna anului 2013,
- temele de care ați fi interesați.

Laboratorul SpeeD poate coordona un număr de 5 stagii de practică. Selecția studenților interesați va fi făcută în perioada 17 – 21 martie pe baza aplicațiilor primite și (eventual) în urma unor scurte interviuri.

Exemple de posibile teme de proiect

1. Multi-lingual speech recognition (2 students)

One of the new challenges in Automatic Speech Recognition (ASR) is dealing with multiple languages input. This condition has become extremely important for applications that target multi-lingual communities like in touristic centers, libraries, etc. One way of overcoming this issue is by implementing a Language IDentification (LID) component whose output is used to load previously trained acoustic and language models of the identified language. Another way is by building multi-lingual acoustic and language models by using heterogeneous speech data in training.

The objective of the project is to implement both solutions and compare their performance from the accuracy and implementation complexity point of view.

2. Unsupervised speaker adaptation for ASR (1 student)

Inter-speaker variability is an important issue in Automatic Speech Recognition. ASR systems trained (or adapted) to recognize a single speaker (speaker-dependent ASR) have significantly better performances than ASR systems which aim at recognizing any voice. However, speaker-independent ASR systems can be backed-up with an unsupervised, online speaker adaptation module to overcome the performance degrading caused by the inter-speaker variability.

This project requires an in-depth study of unsupervised speaker adaptation techniques and a Java implementation of such a module for an already existing speaker-independent ASR system for Romanian.

3. Unsupervised ASR domain (language model) adaptation (1 student)

Domain variability is an important issue in Automatic Speech Recognition. ASR systems designed to recognize speech from a single domain (sport news, travel dialogues, scientific speech, etc.) have significantly better performances than ASR systems which aim at recognizing speech from potentially any domain. However, general ASR systems can be backed-up with an unsupervised, online domain-adaptation module to overcome the performance degrading caused by domain variability.

This project requires an in-depth study of unsupervised domain adaptation techniques and a Java implementation of such a module for an already existing general ASR system for Romanian.

4. ASR output text restoration for increased readability (2 students)

A general ASR system has the purpose of converting the input speech signal into text. When the output text contains only a few words it is human intelligible, but when the ASR output has several pages of un-diacriticized, lowercase, punctuation-lacking text, its readability becomes a real issue.

The goal of this project is to develop an NLP (Natural Language Processing) module with the purpose of increasing the readability and eliminating the ambiguity of a long ASR output text. This NLP module should address: a) diacritics restoration, b) re-capitalization, c) text-to-digits conversion (for numbers) and d) punctuation restoration.

5. Speaker verification/identification system (1 student)

Speaker recognition has two main applications: speaker verification and speaker identification. Speaker verification means deciding whether a speech utterance belongs to a claimed speaker or not. Speaker identification means finding to which speaker belongs a given utterance. In each case the speakers must be modeled by using statistical models. The most common techniques are based on Gaussian Mixture Models (GMM). Common speech features like MFCC, PLP, etc., are used for modeling the speakers.

The objective of the project is to build a complete speaker identification system based on GMMs. It implies studying the GMMs and implementing the training and the classification algorithms.

6. ASR errors analysis and correction (1 student)

State-of-the-art general ASR systems are far from being perfect. The average error rate on a large vocabulary speech recognition task is about 15-20%. In this context ASR error analysis is an important field of study which can bring valuable information about specific systematical errors that can be automatically corrected. Moreover, an important question, regarding the potential benefit of using human corrected ASR output to create automatic correction modules, is still unanswered in the scientific literature.

The goal of this project is to analyze and study the errors performed by an already existing ASR system for Romanian and propose new automatic correction techniques.

7. Automatic phonetization tool for dynamic-vocabulary ASR (1 student)

The acoustic model inside an ASR system does not model the words of the language, but their composing acoustic units: the phonemes. Consequently, any ASR system needs a phonetic dictionary to specify the words pronunciation (the phoneme sequence composing the word). Manual phonetization of words is only feasible for small vocabulary ASR systems, while for large vocabulary ASR systems special automatic phonetization systems have to be developed. Moreover, in dynamic-vocabulary ASR systems the words need to be phonetized on the fly and therefore an automatic phonetization tool is indispensable.

The goal of this project is to develop a dynamic-vocabulary ASR system incorporating an automatic phonetization tool.

8. Text-to-Speech Synthesis for the Romanian language (2 students)

Speech Synthesis means producing a speech signal that carries the information given by a text. The main challenge is to make the speech as natural as possible, i.e. taking into account the speaking rhythm, intonation, syllable stressing, etc. There are several algorithms that are used in speech synthesis, such as PSOLA and HMM based ones. Other issues in Speech Synthesis are related to Natural Language Processing such as diacritics restoration, division into syllables, etc.

The objective of the project is to build a simple Text-to-Speech system based on HMM. This includes speech database collection and annotation, implementation of the HMM based algorithms and the evaluation of the system with variable text input.