ALBAIZYN 2008 EVALUATION CAMPAIGN

http://jth2008.ehu.es/en/albayzin.html

In order to promote Speech Technology research among new young researches, the Spanish Speech Technologies Thematic Network (*Red Temática en Tecnologías del Habla*) organizes a system evaluation proposal, on the next areas:

- ***Language recognition***   http://jth2008.ehu.es/albayzin_VL_en.pdf

In this area a language verification task, similar to the one organized by NIST is proposed. There are 4 target languages (Spanish, Catalan, Basque and Galician) and the signals have more bandwidth and SNR than in NIST Competition. Training, development and test materials are extracted from TV programs and most of them are clean studio recordings.

Verification results will be evaluated using the NIST 2007 Language Recognition Evaluation script, adapted to the proposed task. The training corpus will comprise about 8 hours by language. Together with this corpus a development corpus similar to the evaluation corpus will be provided. The evaluation corpus will contain no more than 2000 files in the 4 target languages and in other unknown languages and its duration will be less than 8 hours.

- ***Speech synthesis***   http://jth2008.ehu.es/albayzin_CTV_en.pdf

The main goal of this evaluation is to favour the collaboration between research teams that focus their research on TTS and to compare different techniques employed. With this purpose in mind, we propose a contest among speech synthesis systems, similar to the Blizzard Challenge. The voice development would be carried out on a common database for Spanish.

The participating teams will have up to seven weeks to build their system from the development material that will be supplied. Next, a set of sentences will be released and the teams will have one week to synthesize them and send the audio files back.

The corpus for development will consist of approximately an hour of recording by a female voice talent. Together with the voice waveforms, automatic segmentation labels in SAMPA format and pitch marks will be supplied. The participating teams will be able to use the data provided only for research purposes.

- ***Automatic text translation***   http://jth2008.ehu.es/albayzin_TA_en.pdf

The purpose of this evaluation is to promote the research on automatic translation systems between Spanish and Basque and to encourage the collaboration among research teams. It is interesting to note that there are few bilingual and monolingual resources related to Basque. In this first evaluation campaign the translation from Spanish to Basque will be developed. Two different systems are proposed: one using limited resources and another with no resources limitation.

For the systems with limited resources 58.000 sentences for each language will be provided for training, about 1.500 sentences for the tuning of the system and another 1.500 for the test. The final test will consist in the translation of some texts from the same domain as the training material but not included in the training corpus. For the evaluation the TC-STAR tools will be used (www.tc-star.org).

General conditions for participation

The participants undertake to present the evaluation results in a special session during the V Workshop on Speech Technology (*V Jornadas en Tecnología del Habla)* that will be held in Bilbao from 12th to 14th November 2008. Participants can take part individually or as a team.

Prizes

The team with the best result in each category will be awarded a diploma, a grant for one of the members of the team, and a prize in cash.

More information at http://jth2008.ehu.es/en/albayzin.html