Proposal for a Call for Participation
Handwriting Recognition and Document Analysis Competition

Since 2001, the International Conference on Document Analysis and Recognition (ICDAR) proposes competitions in the document field starting from page segmentation to Arabic handwriting recognition. In order to extend the latter to other languages, a new competition dedicated to French handwriting recognition is proposed in the framework of **ICDAR 2009, July 26-29, Barcelona, Spain**. Its goal is to evaluate automatic handwriting recognition systems on snippets of French handwritten words extracted from mails sent by individuals to companies or administrations. For this, the high-quality database created in the framework of the **RIMES**[1] campaign is used. It is composed of more than 12000 pages entirely annotated, 100000 snippets of characters and 250000 snippets of words. The proposed competition comes within the scope of the RIMES evaluation campaign which has gathered all French laboratories working on the subject since 2 years over various tasks linked to recognition of documents (segmentation, handwriting recognition and writer identification)

Evaluation Process:

The goal of the proposed competition is to evaluate automatic handwriting recognition systems on snippets of French words with the help of given dictionaries of different sizes (e.g. 100 and 2000 words). The participants will be given about 43000 snippets of words to train their system and a validation database of more than 7000 to test them. At the start of the test period, each participant will have access to the unknown test dataset composed of about 7000 snippets to run his own software on them in his own hardware environment. The result files in the expected format have to be sent back before the end of the test phase.

Participants commit themselves not to modify their system during the test phase. Multiple runs are accepted, but participants must identify one of them as their primary one.

After analysis of the result files, the ground-truths are sent back to all participants with results obtained by metric tools. The participants will have thus the opportunity to propose corrections to ground truths if necessary.

The chosen primary error rate measure consists in counting word error rate. As most of word recognition tools return not a single answer but a list of words with confidence score, a measure of the presence of correct answer in the N best recognition list (N equal to 10) is added.

Format of the output file:

The text file (Participant_System_SizeOfDictionnary.txt) sent back by each participant should have one line for each input image containing the name of the considered image file followed by the recognition results separated by one space as follows:

set_1/image1.tiff recognition1 recognition2 recognition3 recognition4 … recognition10.

Important Dates:
- System registration: as soon as possible
- Training and validation datasets available : 20th January
- Test period between March 31th and April 3rd
- Submission of results: 03 April 2009
- Results will be published at ICDAR 2009

---

1   RIMES : *Reconnaissance et Indexation de données Manuscrites et de fac similÉS/ Recognition and Indexing of handwritten documents and faxes (*http://rimes.it-sudparis.eu)

Contacts :

Emmanuèle Grosicki
  DGA/CEP Arcueil
  16 bis avenue prieur de la Côte d'Or
  94114 Arcueil (France)

  emmanuele.grosicki@etca.fr

Haikal El Abed
  Technical University Braunschweig
  Institut for Communication technologies (IfN)
  Schleinitzstrasse 22, 38106 Braunschweig
  Germany
  elabed@tu-bs.de