



NEMLAR Newsletter

News on Arabic language resources and Arabic language technologies

Issue 20

July 2010

Newsletter content:

1. News from the MEDAR project
 2. Upcoming events
 3. Other news
 4. How to contribute to the MEDAR project
-

1. News from the MEDAR project

◆ The second MEDAR evaluation campaign for machine translation started on the July 9th, 2010. The translation direction is from English to Arabic. Both monolingual and parallel training corpora have been released to participants. They have till the July 23rd, 2010 to train their systems. The corpora sizes are around 1,5 billion words for the monolingual part and more than 3 million words for the parallel part. Moreover, the Linguistic Data Consortium kindly provide to the participants a 2GB monolingual corpus and around 500,000 words of parallel data.

On the July 23rd, 2010, participants will receive the test data of around 40,000 words. 10,000 words will be used for evaluation. Participants have to send back their translations (with a maximum of five translations per system) before the July 30th, 2010. Two MEDAR baseline systems (based on Moses) are also evaluated.

◆ MEDAR had a stand at the EC Project Village at LREC 2010 where many of the MEDAR project partners were present to talk to LREC participants interested in Arabic, in MT, in cooperation or in EU projects. At the stand we handed out a [short version](#) of the **Cooperation Roadmap** proposed by the MEDAR project (a new version of the entire roadmap is coming up) and a new version of the [MEDAR flyer](#).

2. Upcoming Events

- [TSD 2010](#)
Thirteenth International Conference on TEXT, SPEECH and DIALOGUE
Brno, Czech Republic, 6-10 September 2010

- [IMCSIT 2010](#)
International Multiconferenc on Computer Science and Information Technology
Wisla, Poland, October 18-20, 2010
- Workshop:
[Computational Linguistics - Applications \(CLA'10\)](#)
Wisla, Poland, October 18-20, 2010
- [PaIR'10](#)
3rd International CIKM Workshop on Patent Information Retrieval
October 26, 2010
Toronto, Canada (co-located with CIKM 2010)
- Uses and Users for Parallel Corpora in the Translation Process (workshop)
Association for Machine Translation in the Americas (AMTA)
in conjunction with the American Translators Association Conference
November 4-5, 2010 Denver, Colorado
- [ICFHR2010 Arabic Handwriting Recognition Competition](#)
November 16-18, 2010
Kolkata, INDIA
- ASLIB - Translating and the Computer 32 Conference
18-19 November 2010 in London
- International Workshop on Web Mining and Information Retrieval ([WMIR2010](#))
Under the framework of the 10th International Conference on Intelligent Systems, Design and Applications, [ISDA'10](#)
November 29 - December 1, 2010, Cairo, Egypt
- [The fourth International Arabic Linguistics Symposium](#)
Alexandria University, Egypt December 11-12, 2010.
- [ACIT'2010](#)
The International Arab Conference on Information Technology
University of Garyounis, Benghazi, Libya
December 14-16, 2010
- Tenth Conference on Language Engineering ([ESOLEC'10](#))
16-17, December 2010,
Faculty of Engineering, Ain Shams University, Cairo, Egypt.
- Call for papers:
[ICMCS'11](#) - IEEE co-sponsored Conference
2nd International Conference on Multimedia Computing and Systems
Ouarzazate- Morocco, April 7-9, 2011.
Submission deadline: November 10, 2010

3. Other news

♦ [The Essex Arabic Summaries Corpus \(EASC\)](#)

The EASC contains 153 Arabic articles and 765 human-generated extractive summaries of those articles. These summaries were generated using Mechanical Turk. It can be downloaded free of charge for research purposes.

♦ [LDC Standard Arabic Morphological Analyzer \(SAMA\) Version 3.1](#) is now available from the LDC catalog (for members).

♦ [NIST 2004 Open Machine Translation \(OpenMT\) Evaluation](#), a package containing source data, reference translations, and scoring software used in the NIST 2004 OpenMT evaluation, can be obtained from LDC.

4. How to contribute to the MEDAR project

Please help by sending us information on coming events, new software, resources, books, papers and journals. We will distribute it via this newsletter and put it on the MEDAR homepage www.medar.info.

Join the new [NEMLAR Network](#) group in LinkedIn.

You can contribute by:

- giving a description of your fields of interest and expertise in Arabic language technology
- linking to your activities and languages resources
- starting and participating in discussions
- starting and participating in subgroups

To subscribe or unsubscribe to the newsletter, please send an email to: nemlar@hum.ku.dk

The newsletter is published by the MEDAR project www.medar.info and produced by Centre for Language Technology, University of Copenhagen, Denmark.

Contact: Bente Maegaard (Coordinator of MEDAR) or
Dorte Haltrup Hansen,
Centre for Language Technology,
University of Copenhagen,
Njalsgade 140,
Copenhagen S,
Denmark
Tel: +45 35 32 90 74, Fax: +45 35 32 90 89
nemlar@hum.ku.dk