

ITISALAT OCR DISCUSSION

Announcing the Internet forum on Arabic computing, called: ITISALAT. ITISALAT, which made the pages of *Al-Hayat* (the world's most widely circulated Arabic language news daily) on 2 May 1995, promotes contact and stimulates the exchange of information in the field of Arabic computing. ITISALAT subscribers discuss the following topics and others:

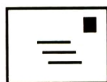
- Arabic optical character recognition (Arabic OCR);
- Arabic computational linguistics;
- Arabic machine translation;
- Arabic computer-based training/education;
- Arabic software development;
- Computing developments in the Arab World;
- Arabic corpora compilation;
- Arabic data storage & retrieval;
- Arabic hypertext;
- Arabic code standardization;
- Arabic email.

The name ITISALAT is actually a play on words. On the one hand it derives from the Arabic word "ittisalaat" which means "connections" or "communications". But it also forms the acronym for *IT IS Arabic Language And Technology*. More than half the 225+ ITISALAT subscribers live in North America. The rest are spread throughout Europe and the Middle East, including Egypt, Iran, Israel, Kuwait, Morocco, Saudi Arabia, Tunisia, and the UAE. And although they hail from widely diverging professions,

such as computer engineering, Arabic linguistics, and library science, they do share a common objective: to further the development of Arabic computing. ITISALAT began operating in May 1993.

The same wave of computer network communication that has captured the American imagination has begun to sweep over the Middle East. More and more Arab countries have connected to the Internet in recent years, and it won't be long before Arabic email becomes the medium of choice for communication with our friends and colleagues from Morocco in the West all the way to Iraq in the East. We hope that ITISALAT can play a small part in the Arabization of the Internet. Time will tell. ITISALAT frequently operates in cooperation with The Association for Arabic Computing in North America, ICEMCO, and The International Association for Arabic Computing. It has no official links with these or other organizations, but we all complement each other, and share a similar goal: to further the development of Arabic computing.

Arabic optical character recognition constitutes one of the most important topics of discussion on ITISALAT. On the surface, the discussion has concentrated on comparing and contrasting the various Arabic OCR software available on the market, along the dimensions of price, platform, ease of use, and quality of results. I hope that the readers of *Manuscripta Orientalia* will find the summary of ITISALAT discussion interesting*.



Even though I may seem an expert on Arabic OCR, I must stress I'm just an end-user with experience with two OCR-programs. I've reached acceptable results by using Al-Qari, but still sometimes (like yesterday night) I curse the product. I haven't published on the subject yet, but I'm now in the stage of preparing a paper about this subject for ICEMCO. This is also going to be a practical report on working with Al-Qari (and some remarks about the other program Iqra'). I will present some results during the presentation of the paper...

...I think I may say I am an experienced user of Al-Qari' al-Ali (Automatic Reader) of al-Alamiya (Sakhr), and also of Iqra' 4. Al-Qari is the newest and best performing of the two, so since about 6 months I stopped using Iqra'. I realized that in my enthusiasm when I reported on 99% success rate in Al Qari al-Ali, I maybe forgot to mention the spell-checking as being part of the process. So to repeat: I copy a page, remove all 'noise', scan it to make a .pcx file. Then I have Al-Qari recognize the .pcx-file. This is done at an average speed of 80—90 characters per second. When the text-file is on the screen, I run the spell check utility, that has a relation with the picture-file. When the spell

* To subscribe, send the command (A) to address (B): (A) — subscribe ITISALAT your-1st-name your-last-name (B) — listserv@listserv.georgetown.edu. For more information, contact: Dr. Paul Roochnik, Moderator, ITISALAT email: roochnik@ios.com. Paul Roochnik.