

## CONTENTS

<i>TEXTS AND MANUSCRIPTS: DESCRIPTION AND RESEARCH</i>	3
<b>L. Menshikov.</b> A Fragment of an Unknown <i>Leishu</i> from Tunhuang . . . . .	3
<b>T. Sultanov.</b> The Structure of Islamic History Book (The Method of Analysis) .	16
<i>TO THE HISTORY OF ORIENTAL TEXTOLOGY.</i>	22
<b>K. Kepping.</b> The Official Name of the Tangut Empire as Reflected in the Native Tangut Texts .	22
<i>PRESENTING THE COLLECTIONS.</i>	33
<b>T. Pang.</b> Rare Manchu Manuscripts from the Collection of the St. Petersburg Branch of the Institute of Oriental Studies, Russian Academy of Sciences	33
<i>ORIENTAL MANUSCRIPTS AND NEW INFORMATION TECHNOLOGIES</i> <i>Correspondence Round table: Arabic/Farsi OCR .</i>	47
<b>A. Matveev.</b> Sakhr Bilingual OCR (Al-Qari' al-Ali). A User's Initial Impressions .	48
<b>J. Bell &amp; P. Zemanek.</b> Test of Two Arabic OCR Programs	55
<b>P. Roochnik.</b> Itisalat OCR Discussion	58
<i>PRESENTING THE MANUSCRIPT</i>	63
<b>O. Akimushkin.</b> <i>Muraqqa'</i> . Album of the Indian and Persian Miniatures of the 16—18th Centuries and the Models of the Persian Calligraphy of the Same Period	63
<i>BOOK AND SOFTWARE REVIEW</i>	68

**Color plates: *Muraqqa'*. Album of the Indian and Persian Miniatures of the 16—18th Centuries and the Models of the Persian Calligraphy of the Same Period (see p. 63—67).**

**Front cover:**

**Fol. 17a.** Portrait of a Man by Riḏā-yi 'Abbāsi, 11.8×8.2 cm.

**Back cover:**

**Plate 1.** Fol. 16a. Portrait of Timūr Khān Turkmān by Ṣādiqī beg Afshār, 19.3×11.6 cm.

**Plate 2.** Fol. 36a. The Darvishes Picnic in the Mountains. Probably Isfahan school, 25.5×14.5 cm.

**Plate 3.** Fol. 6a. The Shaykh and the Harlot by Muḥammad Yūsuf Muṣavvir, 18.2×11.3 cm.

**Plate 4.** Fol. 1a. Portrait of Mirzā Jalālā by 'Alī Qulī beg Jabbādār, 16.0×9.1 cm.

RUSSIAN ACADEMY OF SCIENCES  
THE INSTITUTE OF ORIENTAL STUDIES  
ST. PETERSBURG BRANCH



# **Manuscripta Orientalia**

*International Journal for Oriental Manuscript Research*

Vol. 1 No. 3 December 1995

THESA  
ST. PETERSBURG—HELSINKI

---

---

# ORIENTAL MANUSCRIPTS AND NEW INFORMATION TECHNOLOGIES

## Correspondence Round Table: Arabic/Farsi OCR

The Arabic OCR was and still is one of the most beloved babies of the group of St. Petersburg scholars. The transformation of the pure academic project into the software product with good market possibilities took place in a very short period of time and under the heavy pressure of "post-perestrojka" realities in the ex-USSR, sometimes sad, sometimes tragic... Ex-Soviet academic science and scholarship, Cinderellas at the ball of "new Russians", had to find their own way for survival in the situation of sharp budget cuttings — in 1995 the budget of several academic institutions with the world-wide fame was 200 times (!) less than that of 1985. The case of Arabic OCR, conceived and produced in St. Petersburg, is just one example of the new approach to the existing problems. To my mind, the final success of the project was based on the "three whales": the scholarly background of the specialists from the St. Petersburg Branch of the Institute of Oriental Studies, Russian Academy of Sciences, the talent and diligence of the group of young Russian programmers previously connected with ex-Soviet high-tech military industry, and last but not least, the broadmindedness of young and talented General Manager of al-Alamiah, Cairo, Dr. Ashraf Zaki. The OCR project is a good example of possible cooperation between different Russian and Arabic institutions, cooperation which is replacing now the years of ideologically based relations between the USSR and the Arab world.

After the appearance of Al-Qari al-Ali on the market I received several letters from my friends and colleagues who are working in different research centres all over the world. The problem of Arabic/Farsi OCR was also intensively discussed by the ITISALAT subscribers. The Editorial Board of *Manuscripta Orientalia* proposed to publish some of the reviews. This was the birth hour of the idea of "Correspondence Round Table". Among its participants I have to mention:

*Joseph Norment Bell*, Professor of Arabic, Head of Section for Middle Eastern Languages and Cultures, University of Bergen, Bergen, Norway

*Tim Buckwalter*, Arabic Multimedia Developer and Project Manager Alpine Media (ICUE), Orem, Utah, USA

*Khosrow M. Hassibi*, Senior Staff Scientist, PI of Arabic OCR Project, Automatic Document Recognition Group, Mittek Systems, Inc., San Diego, CA, USA

*Jan Hoogland*, Assistant Professor of Arabic Language at Nijmegen University, the Netherlands.

*Alexander Matveev*, post graduate student of the Oriental Faculty of St. Petersburg State University and special student of Prof. Joseph Bell from University of Bergen

*Paul Roochnik*, Moderator, ITISALAT, Co-Founder and Secretary of the Association for Arabic Computing in North America; Arabic Computational Linguist at Applications Technology (AppTek) Inc., McLean, Virginia, USA

*Petr Zemanek*, Deputy Head, Institute of Ancient Near Eastern Studies, Charles University Prague, Czech Republic

as well as Dr. *Badr al-Badr*, Dr. *Sabri Kawash*, Dr. *Abdel-Hadi* and several others.

The programs mentioned and/or implied in the discussion are as follows (first three phone numbers are in Egypt, check out also the web page at <http://www.rdrop.com/~abdu/arabic.html>):

Co.: **Asset**

☎ 247-3916/248-3187

Prog.: Iqra' (OCR)

Price: L.E 4600.

Co.: **Sakhr**

☎ 274-9929

Prog.: Al-Qari' al-Ali

Price: L.E. 3900

Bilingual (Can OCR Arabic and OCR English).

Co.: **IBM**

☎ 349-2533

Prog.: Qari' Al-Nosoos

Price: \$1100

Runs under OS/2.

Co.: **CTA, S.A.**

☎ 34-3-4850410

Prog. TextPert 3.7 Arabic

Requirements: Apple Macintosh; all more recent models.

I bear all the responsibility for combining the messages of one and the same author once sent to ITISALAT. The information of personal character was omitted as well as the passages not connected with the OCR. The style and language of electronic messages was preserved without nearly any changes.

The St. Petersburg OCR project still has its academic continuation. Thanks to financial support of the Norwegian Centre for Computing in Humanities we are close to the realization of the idea to produce the specialized software for the automatic shape analysis and classification of handwritten symbols in early Qur'anic manuscripts. The preliminary name of the program is ENTRAP and those who are interested in the basic peculiarities of the approach can consult the paper of Mr. Nourlan Kondybaev *Maximum Entropy Approach in Automatic Classification of Symbolic Images* in the Proceedings of the 4th International Conference and Exhibition on Multi-Lingual Computing (Arabic and Roman Script. L., 1994, 2.4—2.5). Of course, the results of the work will find their reflection on the pages of our magazine in the new year.

I hope that the subscribers of *Manuscripta Orientalia* will receive this issue at the end of December, so let me take the opportunity to wish all our readers and, especially, all Arabic OCR users the 100% advance recognition of all possible problems to appear in the New Year and successful spell checking of all the mistakes we have to do in our life.

*Efim Rezvan*,  
Convener of the  
"Correspondence Round Table"