CONTENTS

COLOUR PLATES

Front cover:

Kim Jungyn (Kisan), "[Band of musicians] playing", the drawing No. 24 from the album preserved in the collection of the St. Petersburg Branch of the Institute of Oriental Studies (call number B-35), China ink and water-colours, the second half of the 19th century, 14.0 × 21.5 cm.

Back cover:

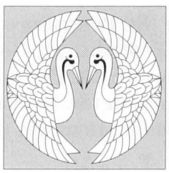
Plate 1. Kim Jungyn (Kisan), "That is how officials (= eunuchs?) in charge of security and palace's tidiness look like (?)", the drawing No. 37 from the same album, China ink and water-colours, 14.0 × 21.5 cm.

Plate 2. Kim Jungyn (Kisan), "This is how the officials clad in formal red garments and hats decorated with gold (for a morning audience) look like", the drawing No. 54 from the same album, China ink and water-colours, 14.0 × 21.5 cm.

Plate 3. Kim Jungyn (Kisan), "The uniform of the official in charge of the sovereign's safety and responsible for passing his orders", the drawing No. 28 from the same album, China ink and water-colours, 14.0 × 21.5 cm.

Plate 4. Kim Jungyn (Kisan), "Officials attached to the sovereign", the drawing No. 35 from the same album, China ink and water-colours, 14.0 × 21.5 cm.

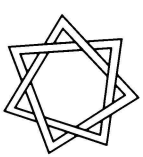
RUSSIAN ACADEMY OF SCIENCES
THE INSTITUTE OF ORIENTAL STUDIES
ST.PETERSBURG BRANCH



Manuscripta Orientalia

International Journal for Oriental Manuscript Research

Vol. 2 No. 3 September 1996



TESA
St.Petersburg-Helsinki

ORIENTAL MANUSCRIPTS AND NEW INFORMATION TECHNOLOGIES

E. A. Rezvan, N. S. Kondybaev

NEW TOOL FOR ANALYSIS OF HANDWRITTEN SCRIPT*

In the end of 1980th a group of scholars from St. Petersburg Branch of the Institute of Oriental Studies (Russian Academy of Sciences) started realization of the "Asiatic Museum" project [1]. The goal was to produce the data-base on the manuscripts preserved in the collection of the St. Petersburg Branch of the Institute as a step towards the creation of the hierarchy of expert systems on different kinds of Oriental manuscripts.

The data-base had to present the three level computer description of all manuscripts from the collection. The differences between the levels were connected with the depth of description and degree of access freedom to the information via network. The first level represented the basic information on a manuscript which is nearly common to all the national traditions (see *Table 1*). The second level had to be enriched by image files of incipit and some other elements of the manuscript, as well as by the block of codicological information. The questions that are posed by a scholar to the manuscript on the second level of description are common to all manuscripts within the national tradition (see *Table 2*). On the second and the third level national languages for description were expected to be used [2].

The third level of computer description (see *Table 3*) had to deal with the group of manuscripts within the national tradition (for example, Qur'ān or Bible manuscripts within the Arabic and Hebrew traditions correspondingly) or with certain elements of a manuscript (for example, paper, binding, script, etc.) [3].

It was decided to start the creation of "Asiatic Museum" data-base with ten thousand Arabic and ten thousand Tibetan manuscripts. The Qur'ānic manuscripts had been taken as the pilot group of MSS for the realization of the third level description [4].

One of the main problems in the modern Qur'ānic studies is connected with the estrangement of the analysis of Muslim tradition from the description and study of the Qur'ānic manuscripts [5]. In this connection the creation the data-base on Qur'ānic manuscripts written in the variety of angular scripts commonly described as Kufic and early cursive variant of the Arabic script like *hijāzī* or *mā'il* seems to be the only way to reconstruct the real picture of the early text history of the Qur'ān. Such a data-base would be not only a simple computer catalogue, but a highly important research tool as well. It could be viewed as the first step on the way to the realization of the idea of the expert system on Qur'ānic manuscripts (see *Table 4*).

Owing to unique information gathered by German and French scholars [6], we know that even when one deals with the scripts looked very much alike, it is only necessary to trace the shape of final *qāf* or *mīm* to distinguish the hands. It was proposed also to add to the script description the analysis of the shape of *alif* and several ligatures. In this connection new possibilities have been opened with the automatization of graphic recognition. Scanning may be used for the purpose of automatic or semi-automatic comparison of the letter shapes and various fragments of illustration.

The approach described below is connected with an attempt to create the software for this purpose. Since the software proved to have been fruitful tool for the shape analysis of nearly any script (for instance, Norwegian runes), we decided to present here the basic mathematical description of the approach. We hope that this would be of some use for our colleagues who cope, as far as we know, with close tasks in different research centers.

* * *

Selection of the separating parameters for automatic classification of images is a significant problem of recognizing the image. These parameters should provide an essential symbol information about object being displayed as an image. The set of parameters depends on selection of an image representation model. These models are built with

* The realization of the project became possible only because of the energy of Professor Joseph Bell and financial support of the Bergen University and the Norwegian Center for the Computing in Humanities (special thanks to Dr Espen S. Ore), as well as due to the grant of the Secretariat for Cultural Exchange Programmes (KAS) of the Research Council of Norway.

the help of the analysis of object's structure and are based on such parameters as contrast, brightness, shape and texture. The most important parameter for a symbolic image is its shape characteristics. Therefore it is these characteristics that are used in many algorithms of Optical Character Recognition (OCR) [7]. The analysis of object's structure is a matter of some difficulty and takes much of CPU time because of different styles of symbol writing.

One of the goals of the present paper is to define the object-classification parameters which describe an object as a whole. These parameters can be obtained using the maximum entropy approach. There are two factors contributing to this approach:

— "image is a set of s-invariant probability measures P defined on image algebra" [8];

— the principle of maximum entropy reads that for drawing the inference based on incomplete information it is necessary to use such a probability distribution whereat maximum entropy is reached under certain restrictions;

— the above principle enables us to introduce a limiting information of uncertainly, thus making it possible to construct a statistical model of symbolic images, as well as to develop an algorithm for symbolic pattern classification using parameters of this model.

Statistical model of symbolic image

A symbolic image is normally considered within a limited domain of plane referred to as symbol perimeter (SP). Assume that a symbol picture is scaled to the dimensions corresponding to SP and appropriately processed prior to recognition.

Now we will consider a symbol picture having distortion of structure's elements caused only by different style of writing.

SP dimensions are $l_x \times l_y$.

Let us define a function for SP:

$$C(x_i, y_i) = \mathrm{Abs}(x_0 - x_i) * \mathrm{Abs}(y_0 - y_i) \qquad (1)$$

where x_i, y_i are current coordinates; x_0, y_0 are coordinates of SP's center.

This function is a weighing one having the following properties:

1) $C(x_0, y_0) = C(x_0, y_i) = C(x_i, y_0) = 0$.

2) Points of one and the same weight ($C(x, y) = \mathrm{Const}$) belongs to a hyperbolic curve.

3) $$\int_0^{l_x/2} \int_0^{l_x/2} C(x, y)\, dx\, dy = \frac{l_x^2 \cdot l_y^2}{64}.$$

We divide symbol picture over n_s cells. The relative position and number of cells depends on contour of symbol.

Now each cell has coordinates (x_i, y_i) and weight of $C(x_i, y_i)$, the total weight of an image on the given picture being as follows:

$$\sum_{i=1}^{n_s} C(x_i, y_i)$$

where summation proceeds over cells having brightness differing from zero.

Function $C(x, y)$ can be interpreted as a value of deviation of cell coordinates from center of SP. By virtue of this function we can estimate the value of distortion (tension, compression, inclination) for the symbol image structure's elements.

Let us consider a symbolic image. It is located on n-cells of SP and has coordinates (x_i, y_i). For each cell we will assign its rate of appearance on symbol picture considered

$$p_i = \frac{\omega_i}{\sum_{i=1}^{n} \omega_i}, \qquad i = 1..n \qquad (2)$$

where ω_i is brightness (the number of pixels) of i-cell with coordinates (x_i, y_i).

Then entropy of the given picture is

$$H = -\sum_{i=1}^{n} p_i \cdot \ln(p_i).$$

Maximization of function $H(p_1,..p_n)$ produces smoothing effects [9], *i. e.* the probabilities p_i and p_j approach each other with the brightness between ω_i and ω_j respectively approaching each other as well.

For pictures of one and the same image we need obeying the constraint on possible probability distributions p_i:

$$\sum_{i=1}^{n} C_i \cdot p_i = a_s, \qquad \text{where } C_i = C(x_i, y_i).$$

It is value of possible image distortion on the given picture.

Finally we have a variation problem:

Define ω_i values, which maximize the function:

$$H = \sum_{i=1}^{n} p_i \cdot \ln(p_i) \qquad (3)$$

under constraints

$$\sum_{i=1}^{n} C_i \cdot p_i = a_s, \qquad \sum_{i=1}^{n} \omega_i = \omega^0. \qquad (4)$$

This is a typical problem of finding the conditional extremum which can be solved by virtue of the Lagrange method of uncertain multipliers [10]. In addition to restrictions (4) it is necessary to use a standard condition:

$$\sum_{i=1}^{n} p_i = 1.$$

Then we will find extremum of function:

$$J = -\sum_i p_i \cdot \ln(p_i) - \beta \sum_i C_i \cdot p_i - \gamma \cdot \sum_i p_i. \quad (5)$$

The following distribution makes maximum available for function (5):

$$\tilde{p}_i = \frac{e^{-\beta C_i}}{\sum_i e^{-\beta C_i}}, \quad \tilde{\omega}_i = e^{-\beta C_i}, \quad \beta = -\frac{1+\gamma}{a_s}.$$

This maximum is $-2 * (1 + \gamma)$.

The problem of classification

Criterion of maximum entropy picks up from the p_i distributions, that one which matches the minimum structure information of the symbol image under certain conditions. By virtue of this distribution we can define ω_i values.

Now we can characterize the pictures of one and the same image by the following parameters:

1) $a_s = \sum C_i \cdot \frac{\omega_i}{\sum \omega_i}$ — extent of possible distortion.

2) $\beta = -\frac{1+\gamma}{a_s}$ — coefficient of average measure.

3) Expected value:

$$M\omega_i = \sum \tilde{\omega}_i \cdot \frac{e^{-\beta C_i}}{\sum e^{-\beta C_i}} = \frac{\sum (e^{-\beta C_i})^2}{\sum e^{-\beta C_i}}.$$

4) Variance:

$$D\omega_i = M(\tilde{\omega}_i - M\omega_i)^2 = \frac{\sum (e^{-\beta C_i})^3}{\sum e^{-\beta C_i}} - \left(\frac{\sum (e^{-\beta C_i})^2}{\sum e^{-\beta C_i}}\right)^2.$$

From the stated above we can conclude that these parameters for pictures of one and the same image are nearly equal, thus featuring any class.

By "class" we mean the probability distribution of brightness $\tilde{\omega}_i$ from SP center for primary standard symbol image. This is a distribution whereat maximum entropy is reached under restrictions (4) and with a certain weighing function (1).

For arbitrary symbol picture we compute a probability distribution whereat maximum entropy is reached. Then given picture has another probability description. We compute parameters 1) — 4). The problem of symbol classification is solved by virtue of computation of minimum distance between parameters of a given picture and parameters of primary standard image.

If we take into account the problem of symbol identification, then this approach permits the parameters to be introduced for different symbols inside one and the same class:

— ω_i square deviation of a given picture from $\tilde{\omega}_i$ values of primary standard image:

$$\Delta\omega^2 = \sum \left(\omega_i - e^{-\beta C_i}\right)^2$$

— deviation of entropy from maximum entropy H_{max}.

These parameters offer scope for separation of symbolic images corresponding to different handwriting styles and calligraphic writing.

Software development

The development of software implies 3 stages:

— program realization of algorithm for computation of symbol parameters;

— verification of validation and obtaining of experimental results for symbol parameters;

— creation of database for working with different handwritten manuscripts.

At present we have completed the first stage. The software was elaborated on 486DX2-80 computer by "*Borland Delphi for Windows*", so it could be used on Mac computers (under Windows for Mac) as well.

Main window of application is shown on *fig. 1*.

We can load and save graphic files (format *.bmp). With the files being loaded, we can correct fragments of a picture: draw, clear and drag (*fig. 2*).

The technique of operation is as follows:

— in one paper of handwritten manuscript we consider one and the same symbol;

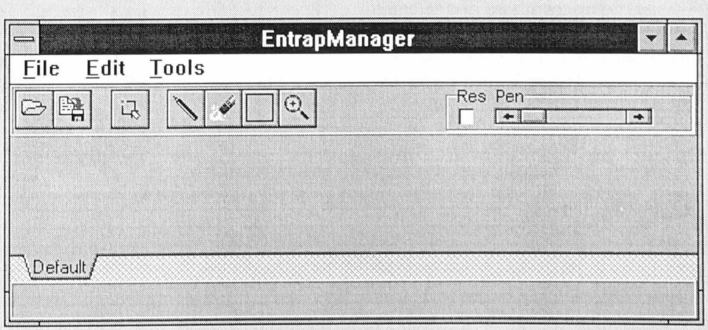


Fig. 1

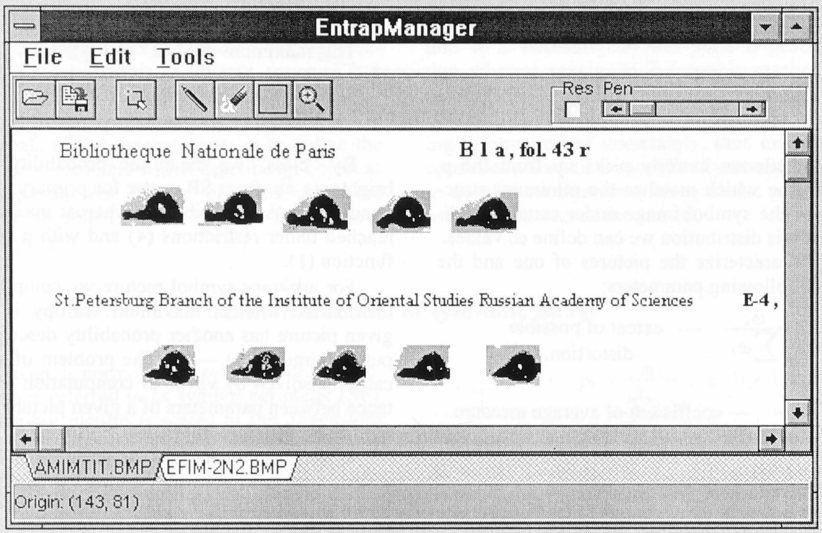

EntrapManager
File Edit Tools
Res Pen
Bibliotheque Nationale de Paris
B 1 a, fol. 43 r
St.Petersburg Branch of the Institute of Oriental Studies Russian Academy of Sciences
E-4,
AMIMTIT.BMP
EFIM-2N2.BMP
Origin: (143, 81)

Fig. 2

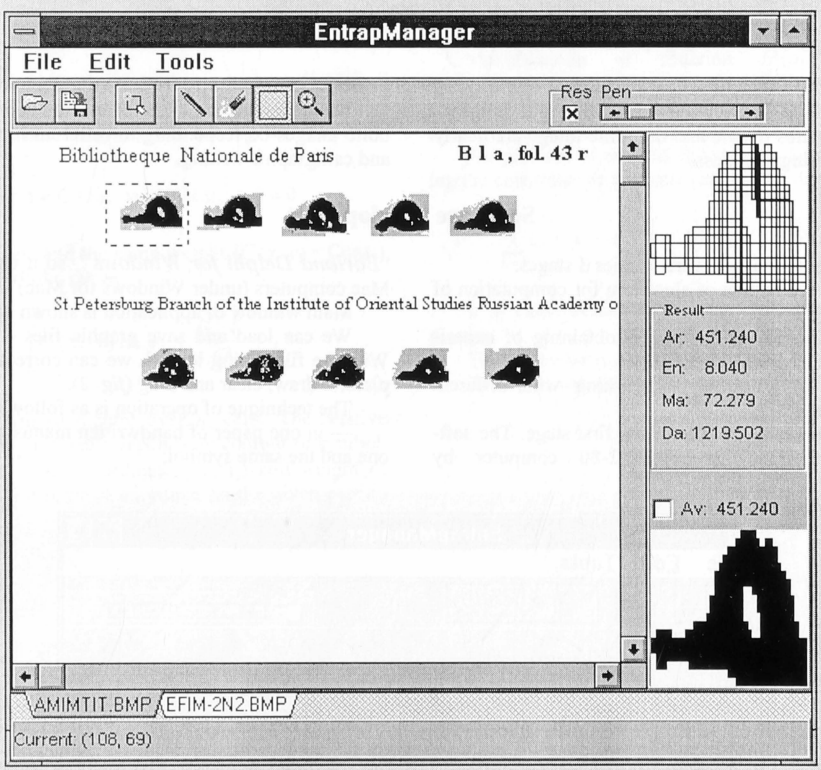

EntrapManager
File Edit Tools
Res Pen
Bibliotheque Nationale de Paris
B 1 a, fol. 43 r
St.Petersburg Branch of the Institute of Oriental Studies Russian Academy o
Result
Ar: 451.240
En: 8.040
Ma: 72.279
Da: 1219.502
Av: 451.240
AMIMTIT.BMP
EFIM-2N2.BMP
Current: (108, 69)

Fig. 3

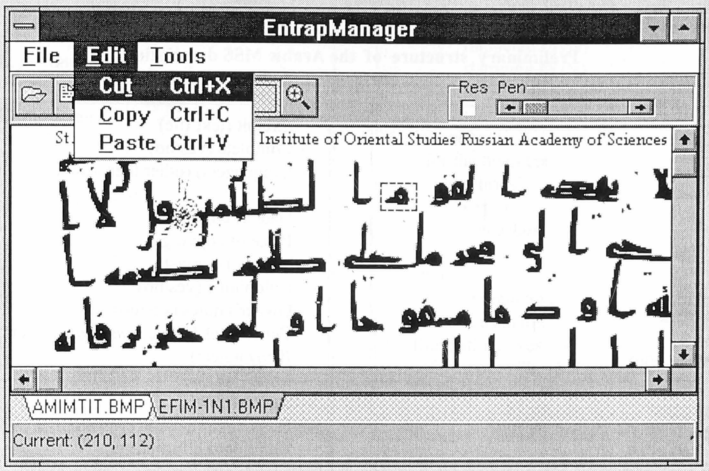


Fig. 4

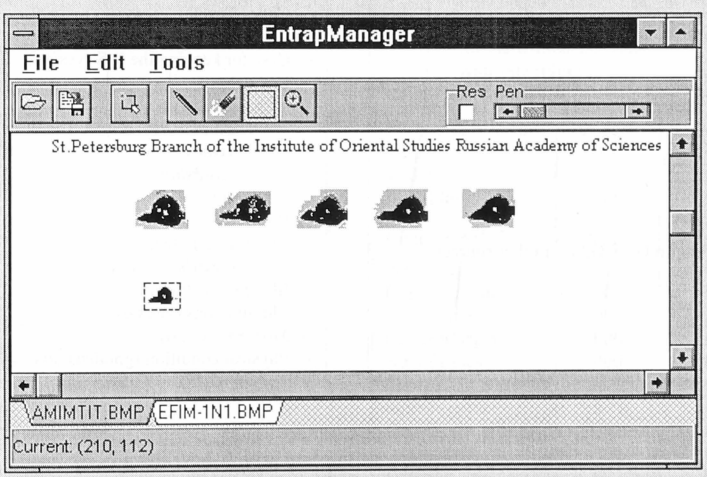


Fig. 5

— this symbol is discriminated from text by a "Rectangle" tool;

— symbol parameters are computed by pressing "Edges" button.

In *fig. 3* we can see the result. Symbol parameters are displayed.

In computing a series of symbols, average of a_s parameter will be shown on screen.

For convenience of research we can cut off necessary symbols from different papers and locate them on one screen (see *fig. 4* and *5*).

Conclusion

The statistical model was described to represent a symbol image and the parameters corresponding to the given image. They were obtained by virtue of this model. These parameters are irrespective of angle of turning.

As the images are compared and discriminated for configurations of the most indeterminate structure, not by their original picture, one can suppose that, due to a weighing function, these parameters are nearly the same for parameters of symbols of one and the same class and greatly differ from those of diverse classes.

Computation of these parameters is simple and their use makes it possible to develop a high-speed classifying algorithm.

In one of the following issues we hope to publish the results of the first tests conducted.

Tables

Table 1

Asiatic Museum (data-base)
Preliminary structure of the Arabic MSS description
(first level)

— Record No.
— Record Author
— Record Date
— Finished level (1/2/3)

— Country
— City
— Library
— Shelf number/press-mark
work = MS
majmū'a (folios 000—000)
voluminous
— Number in the Catalogue
— Number of microfilm
— Bibliography (GAL, GAS, Graf, etc.)

— Author (compiler or translator) (identified or not)
if yes:
name, date of birth/death
shuhra
— Century
Title (identified or not)
if yes:
title (as in GAL/GAS)
title (according to MS)
— Unique or not
— Autograph or not
if not:
other MSS according to GAS/GAL or other sources
— Was published (Y/N)
if yes:
bibliographical data
— Subject
— Incipit (text file)
— Excipit (text file)
— Complete (yes/no)
— Language(s) (other than Arabic)

— Date of copying
— Place of copying
— Name of copyist
— Colophone (yes/no)
— Owner's notes (yes/no)
— Certificates (*ijāza*, *samā'*, *qirā'a*, etc.) (yes/no)
— *Waqf* note(s)
— Seals (yes/no)
— Acquired from:
— Used in publication (yes/no)
— Additional information

— Number of leaves:
— Material: Paper/Parchment/Papyri
— Codex/Scroll
— Binding (yes/no)
— Case for keeping the MS (yes/no)
— Ink: colour
for the main text
rubrics
verses
ḥawāshin
etc.
— Hand
for the main text
ḥawāshin
— Illustrations (yes/no)
— Illuminations (yes/no)
— *Miṣṭāra* (yes/no)
— Physical condition (good/satisfactory/bad)

Table 2

Asiatic Museum (data-base)
Preliminary structure of the Arabic MSS description
(second level)

— Record No.
— Record Author
— Record Date
— Finished level (1/2/3)

— Country
— City
— Library
— Shelf number/press-mark
work = MS
majmū'a (folios 000-000)
voluminous
— Number in the Catalogue
— Number of microfilm
— Bibliography (GAL, GAS, Graf, etc.)

— Author (compiler or translator) (identified or not)
if yes:
name, date of birth/death
shuhra
— Century
Title (identified or not)
if yes:
title (as in GAL/GAS)
title (according to MS)
— Unique or not (Y/N)
— Autograph or not (Y/N)
if not:
other MSS according to GAS/GAL or other sources
— Published (Y/N)
if yes:
bibliographical data
— Subject
— Arrangement of the text (free description)
— Incipit (image file)
Incipit (text file)
— Excipit (image file)

(See continuation of the Table 2 on the next page)

Excipit (text file)
— Complete (yes/no)
if no:
incomplete at the beginning
at the end
number of folios missing
— Language(s) (other than Arabic)

— Date of copying
where date comes from:
colophone
title page
marginalia
indirect dating (century)
— Place of copying
— Name of copyist
— Colophone (yes/no)
if yes:
author's
copyist's
from protograph
image file
— Owner's notes (yes/no)
if yes: names and dates
— Reader's notes (yes/no)
if yes: names and dates
folios No:
— Certificates (*ijāza*, *samā'*, *qirā'a*, etc.) (yes/no)
if yes: names and dates
folios No:
— *Waqf* note(s)
if yes: names and dates
folios No:
— Seals (yes/no)
if yes: folios No:
image file(s)
— Acquired from:
— Used in publication (yes/no)
if yes: bibliographical data
— Additional information

— Number of leaves:
— Material: Paper/Parchment/Papyri
if paper:
water-marks (yes/no)
if yes: description, bibliography
image file
if parchment:
distinguishable sides-scratching: visible/not visible
hair follicles: visible/not visible
thickness: fine/medium/thick or combination
arrangement of sheets: matching sides/opposite
quire starts with: flesh-side/hair side
— Type of MSS (Codex/Scroll)
if codex:
Format: oblong/vertical
Quiring
number of gatherings,
number of folios in gatherings
structure of quires
numbering of quires
Type of sewing (if possible)
Catchword (yes/no)
if yes:
additional information
— Binding (yes/no)
if yes:
type: Oriental/European
material: leather/half-leather/cardboard
flappered (Y/N)
stamping (Y/N)
— Case for keeping the MS (yes/no)
if yes: brief description
— Ink: colour
for the main text
rubrics
verses
ḥawāshin
etc.
— Hand
for the main text
marginalia
— Illustrations (yes/no)
if yes: pp. 000—000
— Illuminations (yes/no)
if yes: pp. 000—000
— *Miṣṭāra* (yes/no)
if yes:
one or more
if more:
folios No-s for each one
type of *miṣṭāra*:
for the written area only
with side ruling
additional information

Miṣṭāra dimensions:	Page 1	Page 2	Page 3
Height	000	000	000
Width	000	000	000
No. of lines	000	000	000

Dimensions*:	Page 1	Page 2	Page 3
Page Height:	000	000	000
Upper Margin	000	000	000
Text Height	000	000	000
Lower Margin	000	000	000
No. of Lines	000	000	000
Page Width	000	000	000
Inner Margin	000	000	000
Text Width	000	000	000
Outer Margin	000	000	000
10 Lines Hgt	000	000	000

* Folios must be taken from the beginning, middle part and the end of the MS.

Proportions:	Page 1	Page 2	Page 3	Average
Page Width/Height	000	000	000	000
Text Width/Height	000	000	000	000
Page Area (.01s.qm)	000	000	000	000
Text Area (.01s.qm)	000	000	000	000
Spatial Proportion	000	000	000	000
Upper/Lower Margin	000	000	000	000
Inner/Outer Margin	000	000	000	000
Upp.+Low./Text Hgt	000	000	000	000
Inn.+Out./Text Wdt	000	000	000	000

Diacritic/Vowel signs in 5 lines: 000
Linear Density 000 Signs/ 10 cm
Spatial Density 000 Signs/ 10 cm^2

— Physical condition (good/satisfactory/bad
+ additional notes)
— Additional Bibliography
— Additional notes for the whole MS

Table 3

Asiatic Museum (data-base)
Preliminary Scheme of Early Qur'ānic MSS Description
(third level)

— Record No.
— Record Author
— Record Date
— Finished
  3rd level (yes/no)

— Country
— City
— Library
— Shelf number/press-mark
— Number in the Catalogue
— Number of microfilm
— Bibliography

— Incipit (image file)
  Incipit (text file)
— Excipit (image file)
  Excipit (text file)
— Complete (yes/no)
  if fragment:
    sūra and *āya* numbers
    number of folios missing
— Connection with any known regional counting system

— Owner's notes (yes/no)
  if yes: names and dates
— Reader's notes (yes/no)
  if yes: names and dates
— *Waqf* note (yes/no)
  if yes:
    who donated
    commissioned by
    where donated
    when donated
    Hijra date:
    A.D. date:
    who copied
    when copied
    Hijra date:
    A.D. date:
    where copied
    Formulas:
    at head
    within the text
    at end
    is *waqf* note contemporary to the text (yes/no)
    if yes: (limitation of data reliability, if obtained without physical methods) — all *waqf* notes must form the image file.
— Seals (yes/no)
  if yes: pages No:
    image file(s)
— Acquired from:
— Additional information

— Number of leaves
— Material: Parchment/Papyri
  if parchment:
  distinguishable sides-scratching: visible/not visible
  hair follicles: visible/not visible
  thickness: fine/medium/thick or combination
  arrangement of sheets: matching sides/opposite
  quire starts with: flesh-side/hair side
  parchment sheets: outer/inner/both
    outer (yes/no)
    if yes: outer sheet starts with:
    inner (yes/no)
    if yes: inner sheet starts with:
  if papyri:

— Codex/Scroll
  if codex:
    Format: oblong/vertical
    Quiring
      number of gatherings,
      number of folios in gatherings
      structure of quires
      numbering of quires
    Type of sewing (if possible)

Binding (yes/no)
  if yes:
    type: Oriental/European
    material: leather/half-leather/cardboard
      flappered (Y/N)
      stamping (Y/N)
— Case for keeping the MS (yes/no)
  if yes: brief description
— **Palaeographical Data:**
  shapes of *alif*, *lām*, *qāf*, existing ligatures form image files to be anlysed by specialised software.
— Ink
  for the main text
    diacritics
    vowel marks
    āya separators
    illuminations and decorations
— *** *āya* separators (yes/no)
  if yes: their position:
    between all *āya*
    each five *āya*
    each ten *āya*
    other groups of *āya* (yes/no)
      if yes: what groups:
if yes (any *āya* separators): are they contemporary to the text (yes/no) (limitation of data reliability, if obtained without physical methods)
  if yes (any *āya* separators): what is the shape of separators:
  abjad
  decorative (form the file of images, which must contain also additional comparative material of ornaments, decorations and illuminations from architectural and written monuments dated by II/VIII—III/X centuries — common file with the points of description marked by ***)
— *** *juz'* and *ḥizb* separators (yes/no)
  if yes: are they contemporary to the text (yes/no) (limitation of data reliability, if obtained without physical methods)

(See continuation of the Table 3 on the next page)

shape: (form the file of images, which must contain also additional comparative material of ornaments, decorations and illuminations from architectural and written monuments dated by II/VIII-III/X centuries — common file with the points of description marked by ***)
— *** decorative *sūra* divisions (yes/no = empty space as separator)
if yes: are they contemporary to the text (yes/no) (limitation of data reliability, if obtained without physical methods)
shapes: (form the file of images, which must contain also additional comparative material of ornaments, decorations and illuminations from architectural and written monuments dated by II/VIII—III/X centuries — common file with the points of description marked by ***)

— *Sūra* titles (yes/no)
if yes: are they contemporary to the text (yes/no) (limitation of data reliability, if obtained without physical methods)
— *Miṣṭāra* (yes/no)
if yes:
type of *miṣṭāra*:
for the written area only
with side ruling

Miṣṭāra dimensions:	Page 1	Page 2	Page 3
Height	000	000	000
Width	000	000	000
No. of lines	000	000	000

— Filling up and preventing exceeding lines:
breaking off the word
leaving space after last word
dilation of letters, if yes: what letters
truncated letters, if yes: what letters
— Diacritic marks (yes/no)
if yes: are they contemporary to the text (yes/no)
(limitation of data reliability, if obtained without physical methods)
— Vowel marks (yes/no)
if yes: are they contemporary to the text (yes/no) (limitation of data reliability, if obtained without physical methods)
shape: dots or traditional marks
if dots: what are the colours for
a
i
u
if traditional marks: what are the colours for
a
i
u
— Belonging to any known system of orthography
— *al-qirā'āt* differed from *rasm 'uthmānī*
if yes: *sūra* and *āya* numbers
if yes: to what known system of text transmission it belongs

Dimensions*:	Page 1	Page 2	Page 3
Page Height:	000	000	000
Upper Margin	000	000	000
Text Height	000	000	000
Lower Margin	000	000	000
No. of Lines	000	000	000
Page Width	000	000	000
Inner Margin	000	000	000
Text Width	000	000	000
Outer Margin	000	000	000
10 Lines Hgt	000	000	000

* Pages must be taken from the beginning, middle part and the end of the MS

Proportions:	Page 1	Page 2	Page 3	Average
Page Width/Height	000	000	000	000
Text Width/Height	000	000	000	000
Page Area (.01s.qm)	000	000	000	000
Text Area (.01s.qm)	000	000	000	000
Spatial Proportion	000	000	000	000
Upper/Lower Margin	000	000	000	000
Inner/Outer Margin	000	000	000	000
Upp.+Low./Text Hgt	000	000	000	000
Inn.+Out./Text Wdt	000	000	000	000

Diacritic/Vowel signs in 5 lines: 000
Linear Density 000 Signs/10 cm
Spatial Density 000 Signs/10 cm^2

— Possibility of being early imitation (yes/no) (limitation of data reliability)
— Dating based on the physical methods
— Physical condition (good/satisfactory/bad + additional notes)
— Additional Bibliography

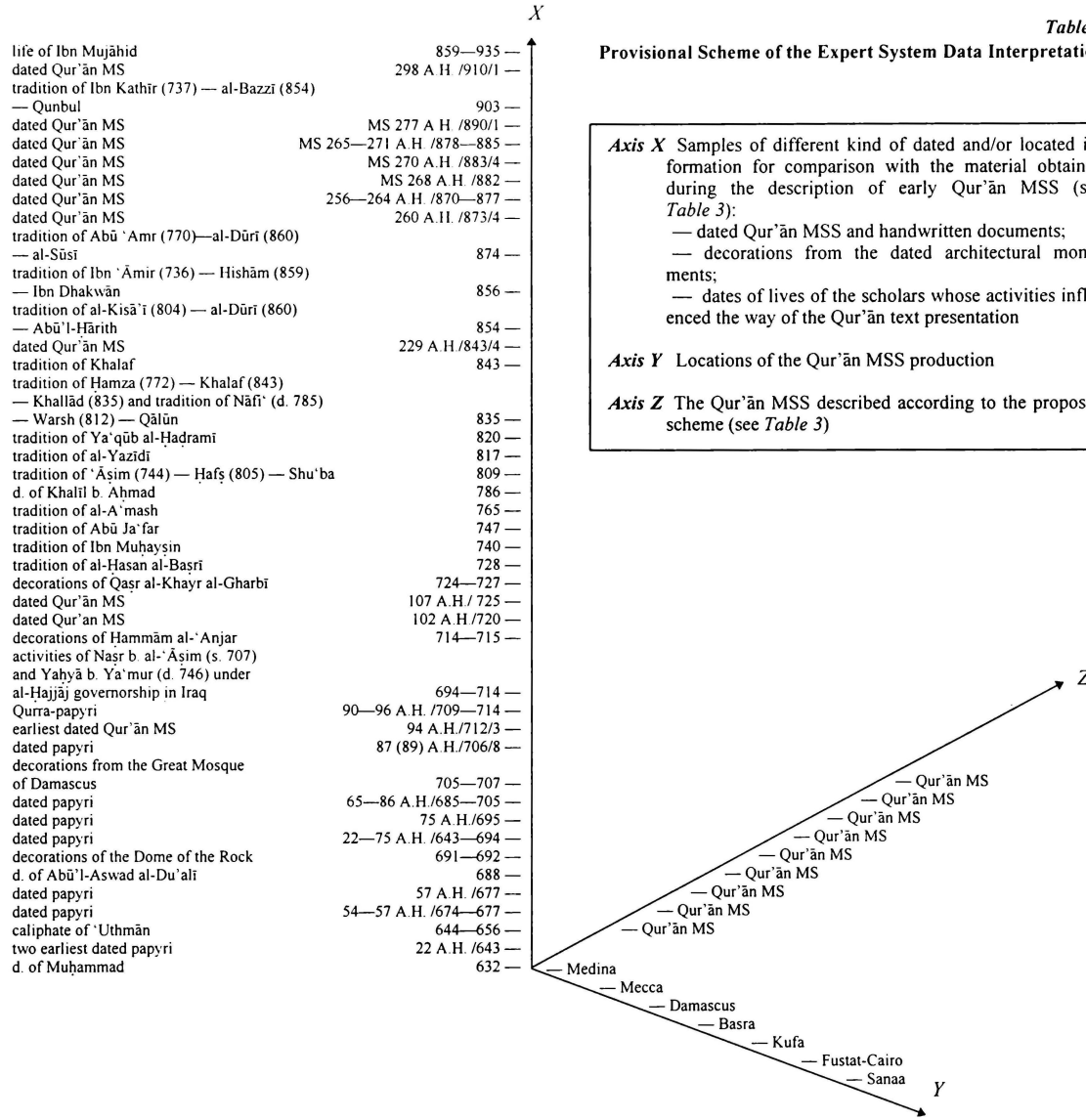
Table 4

Provisional Scheme of the Expert System Data Interpretation

Axis X Samples of different kind of dated and/or located information for comparison with the material obtained during the description of early Qur'ān MSS (see *Table 3*):
— dated Qur'ān MSS and handwritten documents;
— decorations from the dated architectural monuments;
— dates of lives of the scholars whose activities influenced the way of the Qur'ān text presentation

Axis Y Locations of the Qur'ān MSS production

Axis Z The Qur'ān MSS described according to the proposed scheme (see *Table 3*)



Notes

1. E. Rezvan, I. Tikhonova, "Bazy dannykh po rukopisnym sobraniiam: problemy i perspektivy (k nachalu osushchestvleniia programmy)" ("The data-bases on manuscripts' collections: the problems and perspectives (on the beginning of the programme)"), *Bazy dannykh po istorii Evrazii v Srednie veka*, fasc. 1 (Moscow, 1992), pp. 55—63.

2. The schemes for the first and the second levels were jointly proposed by E. Rezvan, Val. Polosin, Vl. Polosin. As for the description of the Qur'ānic manuscripts (the sample of the third level description), it was elaborated by E. Rezvan. Codicological information for the third and for the second levels is treated on the basis of approach elaborated within the Hebrew Paleography Project by Professor Malachi Beit-Arié and his colleagues.

3. E. Rezvan, "The data-base on the early Qur'ān MSS: new approach to the text history reconstruction", *Proceedings of the 3rd International Conference and Exhibition on Multi-Lingual Computing* (*Arabic and Roman Script*) (Durham, 1992), 3.3.1—3.3.17. Also see *idem*, "Computer methods in Qur'ānic studies", *Proceedings of the Conference on Bilingual Computing in Arabic and English* (University of Cambridge, 1990), pp. 1—7.

4. Because of the financial shortages only little part of the project have been realized up to now. About the work on the project see, in particular, *Manuscripta Orientalia*, I/3 (1995), pp. 47—62 (Arabic OCR project); *ibid.*, I/1 (1995), pp. 53—5 (the data-base on Muslim seals); *ibid.*, II/1 (1996), pp. 51—3 (Tibetan data-base) and the present article.

5. E. Rezvan, "The Qur'ān between *textus receptus* and critical edition", *Les problèmes posés par l'éditions critique des textes anciens et médiévaux* (Louvain-la-Neuve, 1992), pp. 291—310.

6. *Masāḥif San'ā'* (Kuwait, 1985); Graf von Bothmer, "*Masāḥif San'ā'*. Qur'ānic calligraphy and illumination as shown in exhibitions in Sanaa and Kuwait", *Ur. International Magazine of Arab Culture*, II (1986); *idem.*, "Frühislamische Koranillumintionen. Meisterwerke aus dem Handschriftenfund der Grossen Moschee in Sanaa/Yemen", *Kunst und Antiquitaten*, I (1986), pp. 22—33; J. Sourdel-Thomin and D. Sourdel, "Nouveaux documents sur l'histoire religieuse et sociale de Damas au Moyen Âge", *Revue des Études Islamiques*, XXXII (1964), pp. 1—25; *idem.*, "À propos des documents de la Grand Mosquée de Damas", *ibid.*, XXXIII (1965), pp. 73—85; S. Ory, "Un nouveau type de mushaf", *ibid.*, pp. 87—149; Fr. Déroche, "Collections de manuscripts anciens du Coran à Istambul. Raport préliminaire", *Études médiévales et patrimoine turc* (Volume publié à l'occasion du 100^{e} anniversaire de la naissance de Kemal Atatürk) (Paris, 1983), pp. 145—65.

7. A. Zahour, B. Taconet, A. Faure, "Machine recognition of Arabic cursive writing", *International Workshop on Frontiers in Handwriting Recognition (from pixels to features III)*. Chateon de Banas, France 23—27 September 1991 (Chateon de Banas, 1991); *Proceedings of the 3rd International Conference and Exhibition on Multi-Lingual Computing*, Part 7.

8. U. Grenander, *Lektsii po teorii obrazov. Analiz obrazov* (Lectures in Pattern Theory. Pattern Analysis) (Moscow, 1992), ii, p. 636.

9. *Rekonstruktsiia izobrazhenii* (Image Recovery), ed. H. Stark (Moscow, 1992), p. 636.

10. R. L. Stratonovich, *Teoriia informatsii* (The Theory of Information) (Moscow, 1975), p. 424.