Arabic scientific document composition

Mustapha Eddahibi* and Azzeddine Lazrek**

*m.eddahibi@ucam.ac.ma **lazrek@ucam.ac.ma

http://www.ucam.ac.ma/fssm/rydarab

ICITNS2003

Plan

1 Introduction

Goal: Multilingual scientific document composition

mathematical expressions in an Arabic presentation

composed with specific symbols & flowing from right to left

There is no system for typesetting such documents

Example: ادرس الدالة التالية :

$$\text{د}(\text{س}) = \begin{cases} \displaystyle\sum_{\text{ب}=1}^{\text{ص}} \text{س}^{\text{ب}} & \text{إذا كان } \text{س} > 0 \\ \displaystyle\prod_{1}^{\text{ص}} \text{س}^{\text{ب}} & \text{إذا كان } \text{س} < 0 \\ \text{جا } \pi & \text{غير ذلك} \end{cases}$$

Importance: e-document

standardization, diffusion, production, evolution, updating

Difficulties: - direction of writing

- big amount of symbols in different forms and sizes that can vary dynamically - processing

2 Arabic math document

2.1 Mathematical document

Consists on three different components consisting on:

- textual component: texts, link-words, ...
- symbolic component: symbols, expressions, ...
- graphic component: arrays, graphs, ...

2.2 Mathematical language

Special features:

- high precision of the expressions composition
  - ⋆ integration of the expressions into the text
  - ⋆ any change in the position of a symbol in the streamline has a meaning
  - ⋆ size of the symbol will have to change according to its position
  - ⋆ change of the shape of the symbol according to the size and the need
- size variation of the symbols according to the context
- use of a significant number of features of writing:
  - ⋆ types: text, symbols, extensible symbols, . . .
  - ⋆ attributes: bold, italic, . . .

2.3 Arabic mathematical language

Arabic writing

- direction of writing from right to left
- Arabic alphabet (ا ب ت)
- cursivity of the writing (محمد)
- consonants are marked as letters while vowels are diacritics (مُحَمَّدٌ)
- certain letters differ from each other only by dots (ب ت ث)
- certain letters differ from each other only by parts of glyphs (ل ن س ص)
- shape of letter depends on its position in the word (م مـ مـ م)

- superposition of the letters of a word through ligatures (محمد محمد)
- stretching of letters in a curvilinear way through kashida (محـــمد)
- punctuation marks in Arabic: orientation (؟) & glyph (·)
- superposition of the words
- several calligraphic styles are in use but not yet formalized: distinguish in shape, ligature, calligraphic rules, . . .
  (Farisi, Koufi, Maghribi, Naskh, Thuluth, Rouq‘a, DywAni, . . .)

Arabic notation system

- letters are with or without dots (ب as i or ب as *ı*)
- another order for classifying alphabetic symbols (ابحد)
- no cursivity of alphabetic symbols (ابح)
- several glyphs for digits (٢١٠ and 210)
- superscript and subscript on the left ($_2$ب ^{2}ب)
- abbreviations use connected letters with ligatures but without vowels; with or without dots (جا as sin or حا as *sın*)
- two punctuation systems ((3 ، 14)ح 3,14)

- some symbols are extensible $\left(\sqrt{3 + \frac{\text{ص}^{\text{ب}}}{\text{ب}}}\right)$
- kashida is curvilinear

$$\Big(\text{ص}^{\text{ب}}\ \underset{1 - \text{ا} = \text{ب}}{\overset{\text{س}}{\text{ـــــــــــمجـ}}}\Big) \text{ instead of } \Big(\text{ص}^{\text{ب}}\ \underset{1-\text{ا}=\text{ب}}{\overset{\text{س}}{\text{ـــــمجـ}}}\Big)$$

Sources of the standards

- handbooks
- conventions such as the Amman's convention
- standards of the AMS and the ISO
- practice of TEX and `MathML`

Existing systems

- free: TEX (LATEX, ArabTEX, Ω, NTS, ...)
  (Scientific Work, LyX, ...)
  high quality, normative, open source, evolutionary, format text, free, scientific societies, ...
- owner: MathType (Equation Editor, TEXaide, MathEdit, ...)

3 RyDArab

3.1 Fonts

Mathematical font with features:

- a significant number of symbols
- a significant number of features of the writing
- change of the symbols size according to the position
- extension of the symbols size according to the context
- a significant number of typographical, didactic and aesthetic constraints

Family fonts static/dynamic

- Family fonts amxnsh: adapted from Naskh used in ArabTEX and developped by K. Lagally

- symbols without dots neither cursivity neither ligature nor vowels (ا ب ح)
- changing automatically the size of the letters according to their positions (ـبـ ـبـ)
- symbols in several forms: isolated, initial and with tail (ب ـب ـد)

Used for alphabetic symbols, digits and punctuation signs (؟ ٧ ب)

- New font NasX: built independently

Used for multiple alphabetic symbols in isolated, initial, tail and stretched forms in simple or double-struck (ـب ـب ـب ـب ـب ـب)

- family fonts amcm (amcmr, amcmsy, amcmex): adapted from Computer Modern used in TEX and developped by D. E. Knuth

- glyph inversion: $\int$ becomes $\int\!\!\!\!\!\!\backslash$
- semantics inversion: $<$ becomes $>$
- command inversion: $\in$ becomes $\ni$
- macro commands change: $\angle$ becomes $\backslash$
- interpretation inversion: { becomes }
- composed glyph and macro inversion: $\sqrt{}$ becomes $\overline{\backslash}$

Used for Latin and Greek alphabetic symbols, basic symbols and linear extensible symbols ($A\alpha \in \overline{\backslash}$)

- Dynamic font through CurExt

Used for curvilinear extensible symbols

$$\text{ص}_{\text{ب}}\ \underset{1-\text{ا}=\text{ب}}{\overset{\text{س}}{\text{مجـ}}} \quad \text{instead of} \quad \text{ص}_{\text{ب}}\ \underset{1-\text{ا}=\text{ب}}{\overset{\text{س}}{\text{مجـ}}}$$

$$\left(\begin{array}{ccc} 5 & \text{ص}2 & 1 \\ 7 & 5 & \text{ب}4 \\ 1 & 8 & \text{ا}3 \\ 7 & 5 & \text{ب}4 \\ 1 & 8 & \text{ا}3 \end{array}\right) \quad \text{instead of} \quad \begin{pmatrix} 5 & \text{ص}2 & 1 \\ 7 & 5 & \text{ب}4 \\ 1 & 8 & \text{ا}3 \\ 7 & 5 & \text{ب}4 \\ 1 & 8 & \text{ا}3 \end{pmatrix}$$

3.2 Options

arabmath : expressions in an Arabic presentation

latinmath : expressions in a Latin presentation

warabnum : standard western Arabic digits

oldstylenum : old style western Arabic digits

earabnum : eastern Arabic digits

alpwithoutdots : alphabetic symbols without dots

alpwithdots : alphabetic symbols with dots

funwithdots : abbreviation representing elementary functions with dots

funwithoutdots : abbreviation representing elementary functions without dots

3.3 Elements and processing

⋆Alphabetic symbols:

ا ، ب ، ح ، د ، و ، ر ، ط ، ى ، ك ، ل ، م ، ن ، س ، ع

ف ، ص ، ق ، ء ، لا ، ے ، ہ

ر ، ح ، ھ ، ل ، م ، س ، ع ، و ، ص

ب ، ح ، ھ ، ط ، ى ، ك ، ل ، م ، س ، ع ، ف ، ص

or

ا ، ب ، ح ، د ، و ، ر ، ط ، ى ، ل ، م ، ن ، س ، ع

ف ، ص ، ق ، ﴿ ، لا ، ے ، ہ ، ر ، م

ر ، ح ، ھ ، ك ، ل ، م ، س ، ع ، ف ، ص

ب ، ح ، ھ ، ط ، ل ، م ، س ، ع ، ف ، ص ، ك

ب ، ح ، ھ ، ك ، ل ، م ، س ، ع ، ف ، ص ، ط

ا ، ب ، ح ، د ، و ، ر

ط ، ى ، ل ، م ، ن ، س ، ع ، ف ، ص ، ق

★Accents:

ا′ ، ب′ ، ح′ ، . . . ؛ ا١ ، ب١ ، ح١ ، . . .

ب′ ، ب″ ، ب‴ ، . . . ؛ ب١ ، ب١١ ، ب١١١ ، . . .

ب̂ ، ب̌ ، ب̃ ، ب́ ، ب̀ ، ب̇ ، ب̈ ، ب̄ ، ب̆ ، $\overrightarrow{\text{ب}}$ ، $\overleftarrow{\text{ب}}$

$\underbrace{\text{ابـح}}$ ، $\overbrace{\text{ابـح}}$ ، $\overrightarrow{\text{ابـح}}$ ، $\overleftarrow{\text{ابـح}}$ ، $\widetilde{\text{ابـح}}$ ، $\widehat{\text{ابـح}}$ ، $\overline{\text{ابـح}}$ ، $\underline{\text{ابـح}}$

★Common functions:

جا ، جتا ، ظا ، ظتا ، قا ، قتا ؛

زجا ، زجتا ، زظا ، زظتا ، زقا ، زقتا ؛

جاز ، جتاز ، ظاز ، ظتاز ، قاز ، قتاز

زجاز ، زجتاز ، زظاز ، زظتاز ، زقاز ، زقتاز

لو ، قه

★New function

$$\text{صغر}(\text{س}) = \text{جتا}(\text{س}^2) - 6$$

★Digits:

- Western Arabic digits: {9 ، 8 ، 7 ، 6 ، 5 ، 4 ، 3 ، 2 ، 1 ، 0}
- Western old Arabic digits: {9 ، 8 ، 7 ، 6 ، 5 ، 4 ، 3 ، 2 ، 1 ، 0}
- Eastern Arabic digits: {٩ ، ٨ ، ٧ ، ٦ ، ٥ ، ٤ ، ٣ ، ٢ ، ١ ، ٠}

★Numbers: 3,14 ، 12.345 ، 107 ، 8− ، 92+ ، 5 ، 7

٣,١٤ ، ١٢.٣٤٥ ، ١٠٧ ، ٨− ، ٩٢+ ، ٥ ، ٧

★Punctuation:

… — ؟ ! : ؛ . ، ؛ , . ; : ! ? — …

★Delimiters:

{، ⟨، (، | ، [، ‖ ، {، ⋯ ،} ، ‖ ،] ، | ،) ،⟩ ،}

★Basic symbols:

… ، = ، × * ، − ، +

… ، ؉ ، ٪ ؛ ‰ ، % ؛ /

… ، $\overset{\text{تعر}}{\Longleftrightarrow}$ ، $\overset{\text{تعر}}{=}$ ، =: ؛ ∈ ، ∋ ؛ < ، >

… ، ∀ ، E ؛ \

… ، ∌ ، ≯ ، ≠

⋆Size variation

- Upper script: $^{\text{ب}}\,^{\text{ب}}\,^{\text{ب}}\;^{b}\,^{b}\,^{b}\;^{2}\,^{2}\,^{2}\;^{\vee}\,^{\vee}\,^{\vee}\,^{\vee}\;^{+}\,^{+}\,^{+}\,^{+}\;^{>}\,^{>}\,^{>}\;^{>}$
- Lower script: $_{\text{ب}}\,_{\text{ب}}\,_{\text{ب}}\;_{b}\,_{b}\,_{b}\;_{2}\,_{2}\,_{2}\,_{2}\;_{\vee}\,_{\vee}\,_{\vee}\,_{\vee}\;_{+}\,_{+}\,_{+}\,_{+}\;_{>}\,_{>}\,_{>}\;_{>}$

$$\prod_{\text{ب}}^{\text{ب}}\sum_{\text{ب}}^{\text{ب}}\int_{\text{ب}}^{\text{ب}} \qquad \prod_{\text{ب}}^{\text{ب}}\sum_{\text{ب}}^{\text{ب}}\int_{\text{ب}}^{\text{ب}}$$

⋆Fraction: $\frac{\text{ص}*2}{\text{ب}}$ ؛ $\frac{1}{2}$

⋆Root: $\sqrt[\text{ب}]{5}$ ؛ $\sqrt[2]{\text{س}}$ ؛ $\sqrt[\frac{\text{ب}*9}{\text{لس}}]{}$ ؛ $\sqrt[3]{\frac{\text{ب}*9}{\text{س}}+2}$

⋆Integral: $\int_{1}^{\text{ط}}\text{س}^{\text{ب}}\,\text{ءس}$ ؛ $\int_{1}^{\text{ط}}\text{س}^{\text{ب}}\,\text{ءس}$ ؛ $\int_{1}^{\text{ط}}\text{س}^{\text{ب}}\,\text{ءس}$

★Sum: $\sum_{\text{ب}=1}^{\text{ص}} \text{س}_{\text{ب}}$ ؛ $\sum_{\text{ب}=\text{ا}-1}^{\text{ص}} \text{س}_{\text{ب}}$ ؛ $\sum_{\text{ب}=1}^{\text{ص}} \text{س}_{\text{ب}}$ ؛ $\sum_{\text{ب}=\text{ا}-1}^{\text{ص}} \text{س}_{\text{ب}}$

$\text{مجـــــ}_{\text{ب}=1}^{\text{ص}} \text{س}_{\text{ب}}$ ؛ $\text{مجـــــ}_{\text{ب}=\text{ا}-1}^{\text{ص}} \text{س}_{\text{ب}}$ ؛ $\text{مجـــــ}_{\text{ب}=1}^{\text{ص}} \text{س}_{\text{ب}}$ ؛ $\text{مجـــــ}_{\text{ب}=\text{ا}-1}^{\text{ص}} \text{س}_{\text{ب}}$

★Product: $\prod_{\text{ب}=1}^{\text{ص}} \text{س}_{\text{ب}}$ ؛ $\prod_{\text{ب}=\text{ا}-1}^{\text{ص}} \text{س}_{\text{ب}}$ ؛ $\prod_{\text{ب}=1}^{\text{ص}} \text{س}_{\text{ب}}$ ؛ $\prod_{\text{ب}=\text{ا}-1}^{\text{ص}} \text{س}_{\text{ب}}$

$\text{جـــــذ}_{\text{ب}=1}^{\text{ص}} \text{س}_{\text{ب}}$ ؛ $\text{جـــــذ}_{\text{ب}=\text{ا}-1}^{\text{ص}} \text{س}_{\text{ب}}$ ؛ $\text{جـــــذ}_{\text{ب}=1}^{\text{ص}} \text{س}_{\text{ب}}$ ؛ $\text{جـــــذ}_{\text{ب}=\text{ا}-1}^{\text{ص}} \text{س}_{\text{ب}}$

★Limit:

$\text{نهـــــا}_{\text{س}\to 0} \text{س}^2$ ؛ $\text{نهـــــا}_{\text{س}\to +\infty} \text{س}^2$ ؛ $\text{نهـــــا}_{\text{س}\to +\infty} \text{س}^2$ ؛ $\text{نهـــــا}_{\text{س}\to 0} \text{س}^2$

⋆Function defined with cases:

$$\text{د}(\text{س}) = \left\{ \begin{array}{ll} -4\text{س} & \text{إذا كان س} < 0 \\ 4\text{س} & \text{إذا كان س} > 0 \\ -2 & \text{غير ذلك} \end{array} \right.$$

⋆System of equations:

$$\left. \begin{array}{r} 5\text{س} - 4\text{ص} + 6 = 8\text{ص} \\ 5\text{س} - 7\text{ص} + 9 = 3 \\ 5\text{س} - 7\text{ص} + 9\text{ط} = 12\text{س} \end{array} \right\} \qquad (1)$$

⋆Matrix:

$$\text{م} = \begin{pmatrix} 1 & 2\text{ص} & 3 & +5 \\ 4 & 5 & 6 & -2 \\ 4 & 5 & 6 & -2 \end{pmatrix}$$

⋆Array:

المبيان		
النقط	الإحداثيات السينية	الصادية
ا	5	7
ب	8	9

⋆Numbering:

(ا3) $2س^2 - 3س + 5 = 0$ ؛

(ب3) $2ص^2 - 3س + 5 = 0$.

⋆Date:

2003/10/16 16 أكتوبر 2003

٢٠٠٣/١٠/١٦ ١٦ تشرين الأول ٢٠٠٣

⋆Translation:

```
$$d(c) = {\cases{
 {{\ssum\limits_{b=1}^{s}} c{{}^b}} &
   {{\amhbox{AYZa kan}} c < 0} \cr
 {{\sint\limits_{1}^{s}} c{{}^b}} &
   {{\amhbox{AYZa kan}} c > 0} \cr
 {{\sin} \pi} & {{\amhbox{Gyr Zlk}}}\cr
 }}$$
```

$$\text{د}(\text{س}) = \begin{cases} \sum\limits_{\text{ب}=1}^{\text{ص}} \text{س}^{\text{ب}} & \text{إذا كان } \text{س} < 0 \\ \int\limits_{1}^{\text{ص}} \text{س}^{\text{ب}} & \text{إذا كان } \text{س} > 0 \\ \text{جا } \pi & \text{غير ذلك} \end{cases}$$

$$d(c) = \begin{cases} \sum_{b=1}^{s} c^b & \text{إذا كان} \quad c < 0 \\ \int_{1}^{s} c^b & \text{إذا كان} \quad c > 0 \\ \sin\pi & \text{غير ذلك} \end{cases}$$

$$d(c) = \begin{cases} \sum_{b=1}^{s} c^b & \text{if} \quad c < 0 \\ \int_{1}^{s} c^b & \text{if} \quad c > 0 \\ \sin\pi & \text{otherwise} \end{cases}$$

$$d(c) = \begin{cases} \sum_{b=1}^{s} c^b & \text{si} \quad c < 0 \\ \int_{1}^{s} c^b & \text{si} \quad c > 0 \\ \sin\pi & \text{sinon} \end{cases}$$

3.4 Features

In addition to the distinguished qualities of TEX, the system RyDArab is:

- compatible with all other extensions of TEX
- its commands are structurally analogous to those of TEX
- runs with Windows or Unix/Linux platforms
- used in conjunction with any TEX implementation or format (`plain` TEX or LATEX)
- can be adapted according to the areas and levels
- ...

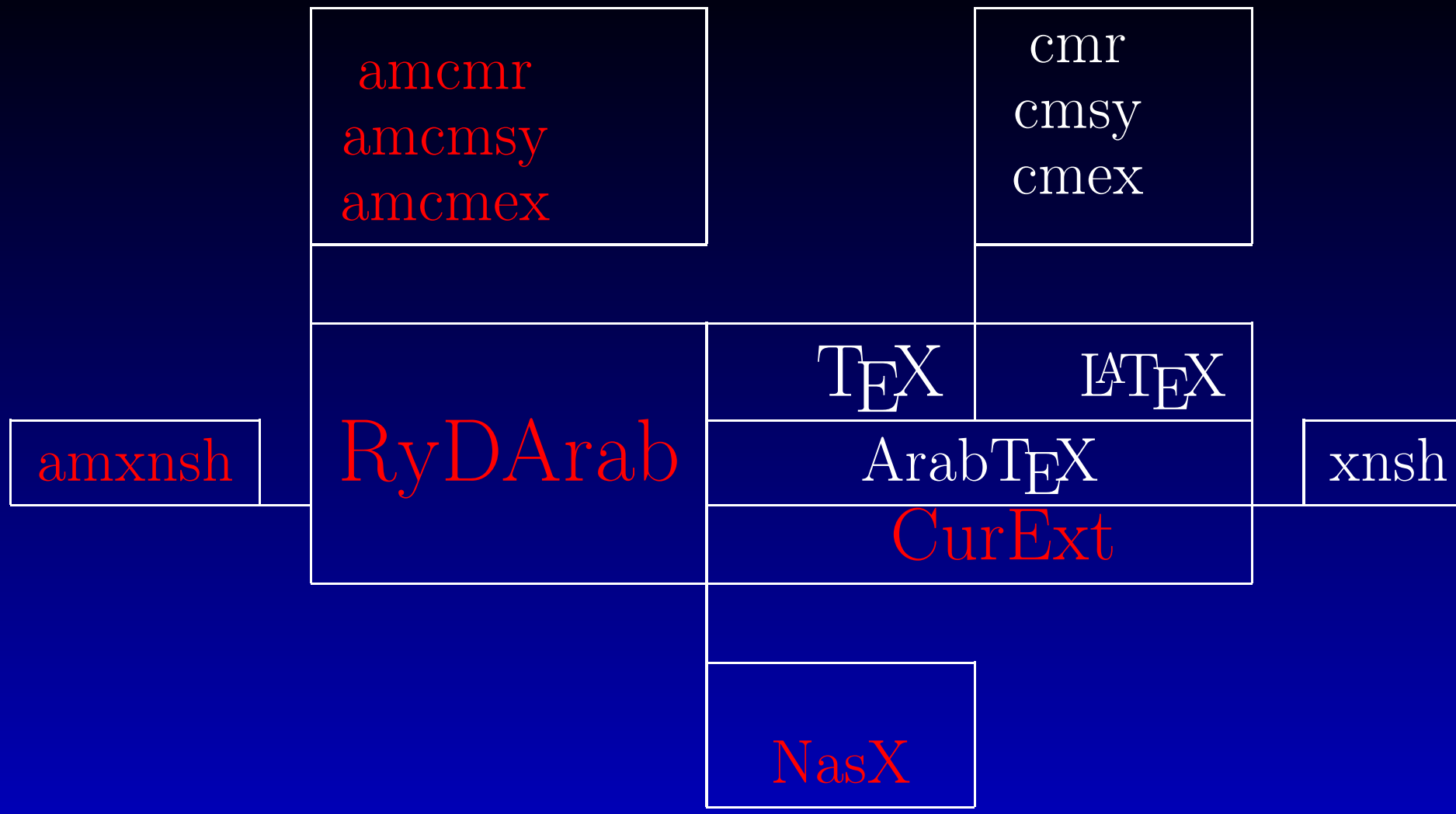

Table 1: Integration of the new systems in environment TEX

4 Conclusion

4.1 Conclusions

- Arabic mathematical language: notation system & typographic rules
- operational Arabic system: unification of semantics, syntax, interpretation, ...
- there are still some adaptation improvements:
  - ⋆ automatic management features: spaces
  - ⋆ inversion of direction: left/right, before/after
  - ⋆ heterogeneity: size, glyphs, bold, position
  - ⋆ indirect input: transliteration
  - ⋆ calligraphy completion of symbols: formalization
  - ⋆ need of the environment TEX

4.2 Applications

- e-edition & e-publication
- applications of formal computing
- didactic of mathematics
- automatic reasoning, translation and correction

4.3 Prospects

- provide to the user a suitable interface
- build a complete Arabic mathematical font
- integrate the Arabic math characters to Unicode
- justify the Arabic cursive text with the kashida
- translation Arabic/Latin
- adaptation for `XML` and `MathML`

The system will always be in evolution...