

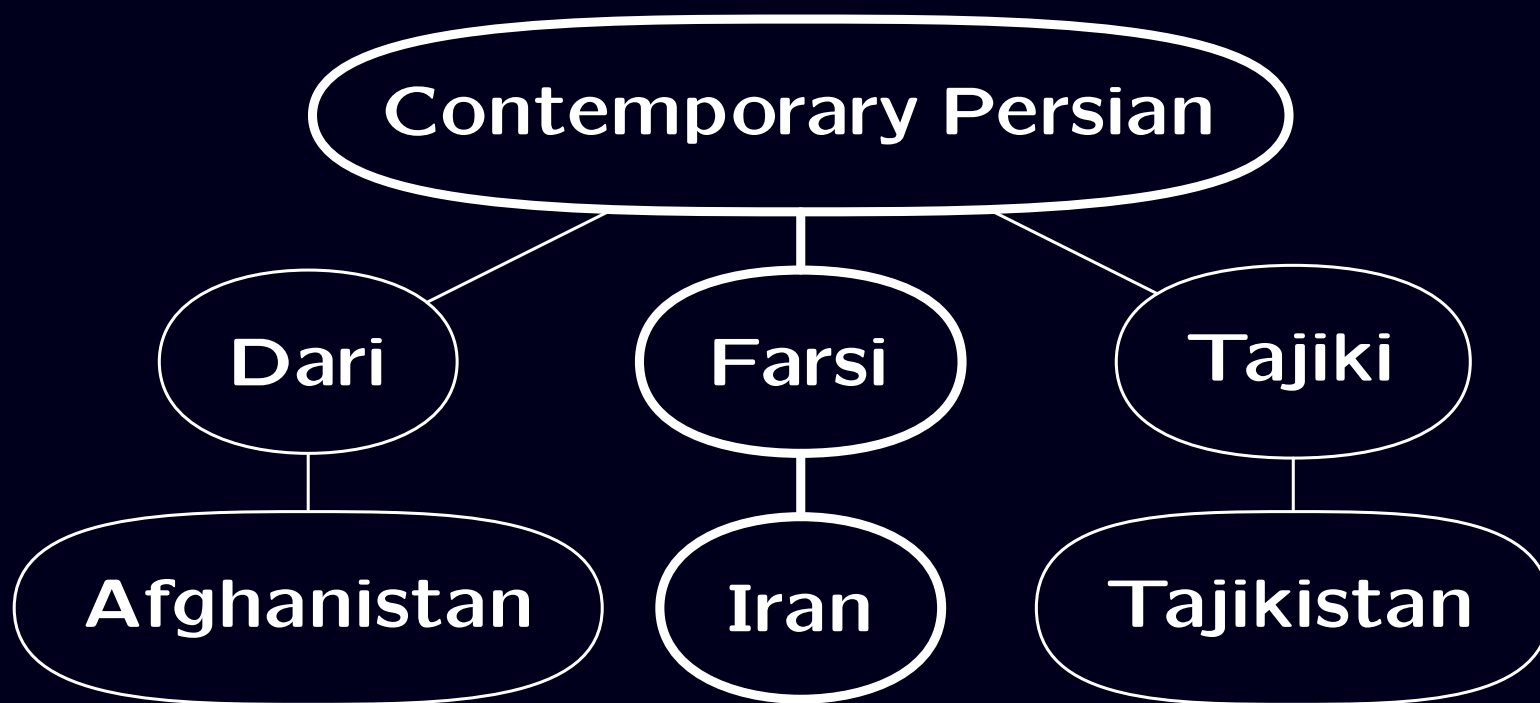
# FarsiT<sub>E</sub>X and the Iranian T<sub>E</sub>X Community

Behdad Esfahbod  
farsitex@behdad.org

Roozbeh Pournader  
roozbeh@sharif.edu

The 23rd Conference and Annual Meeting  
of the T<sub>E</sub>X Users group,  
Trivandrum, Kerala, India  
September 4, 2002

## What is Persian?



# The Modern Persian Script

- Based on the Arabic Script
- Extra letters: Peh (پ), Tcheh (چ), Jeh (ژ), and Gaf (گ)
- Modified letters:
  - Kaf (ك) → Keheh (ک)
  - Yeh (ي) → Farsi Yeh (ی)

# The History of the Script

- The switch from Pahlavi to Arabic happened in the 7th century CE
- The adaption propagated to Pakistan, Afghanistan, India, China, Malaysia, and Java where the alphabet was extended even more: 29 basic Arabic letters → 139 letters in modern use (from Kurdish to Jawi)

# The Persian Typography

- Based on calligraphic practices
  - Originally Naskh (as opposed to Kufi), the Meccan style of writing Arabic
  - Nastaliq was invented in 15th century CE and the calligraphy switched
- With lead typography it switched back to Naskh
- With late 1990s proprietary digital typography tools, Nastaliq become public again, but the popularity dropped because of unreadability

# Persian Scientific Typography

- Blossoming in 1950s by Mosahab works (who also invented *Iranic*)
- Manual typesetting using “match stick methods”
- LinoType machines in 1970s, modern publishers raised, resulting in a leap in math books

## Localized T<sub>E</sub>Xs

- T<sub>E</sub>X-e-Parsi and L<sup>A</sup>T<sub>E</sub>X-e-Farsi appearing in 1992
- T<sub>E</sub>X-e-Parsi, won the competition because of better quality

## TEX-e-Parsi

- Developed by high investment from the vendor and a few major scientific publishers, going TEXtreme
- The vendor went bankrupt in 1997
- Latest version in 1996, with pre-3.0 TEX and L<sup>A</sup>TEX 2.09 + NFSS
- A few math departments and the two original publishers who sponsored it still use it
- The price was very high



## Zarnegar, the alternative

- Appearing in early 1995
- Original design, using a visual markup language
- Splendid fonts, and the vendor's knowledge of the market
- Still in wide use: may be the second popular software after MS Word
- Main Problems: Unbearable math typesetting, and a proprietary and closed file format

# FarsiT<sub>E</sub>X

- Started as an academic project by Mohammad Ghodsi in 1991, called FaT<sub>E</sub>X in the first year
- Three BSc projects provided the foundation in 1992 and 1993
- Two master theses in 1994, shaped the current macros, and the Scientific Farsi (sf) family of fonts
- Some Arabic script specific works, like contextual shaping of letters, was done in a pre-processor

# The Old Releases

- A new team was gathered in 1996
- The team created a new syntax and character set
- Wrote some converters, and an MS-DOS editor
- The engine was based on em $\text{T}_\text{E}\text{X}$ , and  $\text{L}\text{A}\text{T}_\text{E}\text{X}$  2.09
- Released Farsi $\text{T}_\text{E}\text{X}$  for MS-DOS under GNU GPL
- The last release of this era is dated October 1998

# The New Releases

- After a meeting in 2000, the team become semi-active again
- A MS Windows editor was almost ready
- Packaged engine based on MiK<sub>T</sub>E<sub>X</sub>
- Released the MS Windows version

## Other Released Stuff

- Localized version of MakeIndex
- FarsiT<sub>E</sub>X to HTML converter tool, written from scratch
- . . . which are just some prototypes

# Never Released Material

- Azin fonts, as an alternative to the original Scientific Farsi font family
- The  $\text{\LaTeX} 2_{\epsilon}$  macros
- $\text{teTeX}$  based engine (Linux & friends finally)
- Farsi $\text{\TeX}2\text{HTML}$ , based on  $\text{\LaTeX}2\text{HTML}$

## Never Released Material (continued)

- PostScript Type 1 Scientific Farsi fonts
- Popular public domain Persian fonts, converted to both METAFONT and PS Type 1
- FarsiT<sub>E</sub>X2Unicode character set converter

# Linux Editor?

- Not yet. Many people promised to write one, but possibly forgot it!
- The current MS Windows editor runs using WINE
- There's a Persian LyX
- What about transliteration-based input?



# Problems with the Current Version

The current version, being based on  $\text{\LaTeX}$  2.09, has many problems, a barrier to further development:

- $\text{\LaTeX}$  2.09 is not supported anymore
- Lack of NFSS support, which makes using other Persian fonts too hard
- The design is dirty, and overrides many  $\text{\LaTeX}$  internals, so that hardly any  $\text{\LaTeX}$  package would work with Farsi $\text{\TeX}$ , unless some tailoring is done

# T<sub>E</sub>Xnical Details

- Having it's own character set, FarsiT<sub>E</sub>X needs it's own special editor
- Some converters are needed to pre-process the input
- And finally, the macros (and the T<sub>E</sub>X--X<sub>E</sub>T engine) take care of bidirectional rendering

# Arabic Script Rendering

Input text	Logical order	م ا ل س
After Bidirectional Algorithm	Visual order	س ل ا م
After Arabic Joining Algorithm	Glyph list	س د ل ا م
After Ligation	Glyph list	س د ل ا م
When Rendered	Output	سلام

With enough care, the above algorithms can be applied in some different order.

# Bidirectional Algorithm

- Main issue to tackle
- $\text{T}_{\text{E}}\text{X} \rightarrow \text{X}_{\text{E}}\text{T}$  can render bidirectional text
- ... but only when subtext directions are known explicitly!
- The editor or the pre-processor should specially mark the directions for the  $\text{T}_{\text{E}}\text{X} \rightarrow \text{X}_{\text{E}}\text{T}$  engine

## Bidirectional Algorithm (continued)

- A very simplified bidirectional algorithm, but powerful
- The editor converts between logical and visual orders
- Two code points for some punctuation marks
- Identify the direction (using the background color in the editor)
- Pre-processor marks different directions by inserting `\InE`, `\EnE`, `\InF`, and `\EnF`

# Joining & Shaping Algorithms

- Two adjacent letters may *join* to each other, or may not
- ... forming 1, 2, or 4 glyphs for each character (for example  
س, سس, سس, سس)
- The Joining Algorithm is for deciding if two adjacent letters do join or not
- The Shaping Algorithm is for selecting the proper glyph, based on the results of the Joining Algorithm
- The pre-processor and the editor are responsible for them

# Line Justification

- It is common to stretch the joining line between letters
- No inter-letter spacing, no hyphenation
- The pre-processor inserts a *stretchable Kashida* character between the connected letters
- The active inserted character, then, expands to a horizontal glue filled by horizontal rules

# FarsiT<sub>E</sub>X Forever

- FarsiT<sub>E</sub>X is not released as a part of any T<sub>E</sub>X distribution yet, mainly because the team members still think that it's not stable
- The team is going to cleanup and release the current code base, with PostScript Type 1 fonts, based on MiK<sub>T</sub><sub>E</sub>X and t<sub>e</sub>T<sub>E</sub>X, for both MS Windows and Linux platforms?



## FarsiT<sub>E</sub>X Forever (continued)

- The system should be redesigned, restructured, and rewritten, which needs breaking backwards compatibility, that is the reason it is not happened yet
- And “The Ultimate Solution”, is moving to Unicode and using Omega

# Iranian T<sub>E</sub>X Community

- There is no real community
- There are people using (Farsi)T<sub>E</sub>X daily and professionally
- Some are active in mailing lists too
- But it is far from an active community: nobody contributes (*has ever contributed*) patches!

# The Team

(The new FarsiT<sub>E</sub>X team in 1999)



<http://www.farsitex.org/>

# Questions?