

# Towards Spell checking in FarsiTeX

Youssef Rezvan  
 Islamic Azad University-Ahar Branch  
 Dept. Electrical Eng.&Computer  
 Ahar, IRAN

Karim Ivaz  
 Tabriz University  
 Dept. Mathematical Sciences  
 Tabriz, IRAN

Farshad Rezvan  
 Tabriz University  
 Dept. Mathematical Sciences  
 Tabriz, IRAN

*Abstract:* This paper contains a brief introduction to FarsiTeX, a document preparation tool in Farsi Language based on L<sup>A</sup>T<sub>E</sub>X, as well as explaining the difficulties of spell cheking in that.

*Key-Words:* Farsi manuscripts, FarsiTeX, Spelling in Farsi

## 1 Introduction

The most powerful formatting program for producing scientific and technical texts in book quality is that of Donald E. Knuth (1984-86). The program is called T<sub>E</sub>X [1, 3].

The enormous power of T<sub>E</sub>X has its price: its application and indeed the exploitation of all its potential demand considerable experience with programming techniques. Its application thus remains the exclusive domain of programming professionals. For this reason, the program package L<sup>A</sup>T<sub>E</sub>X has been developed by the American computer scientist Lesli Lamport (1985), which uses an exhaustive set of T<sub>E</sub>X commands to create a more user-friendly interface between T<sub>E</sub>X and user [2].

FarsiTeX software package (Fig. 1) has been developed by Dr Mohammad Ghodsi and his FarsiTeX project group [4].

## 2 FarsiTeX

This package is a bi-language typesetting package, Farsi and English, the former writes from right to left and the latter vice versa. This system is based on a paper of Knuth and Mackay [5] solving right to left and left to right writing problem. In Farsi Language texts are written from right-to-left but digits and arithmetics from left-to-right. So if we also mix English to this mixture, Farsi digits should be treated as well as English texts. This is an important feature and has been solved very well in FarsiTeX.

Font design for FarsiTeX was not so simple. Some letters in Farsi language has four forms, so FarsiTeX uses four char code to represent them. Some others just has two forms. however in keyboard there is just one key to any letter. There exists some rules in Farsi language to manage these char codes of each letter. When a writer types letters in FarsiTeX to construct a Farsi word, FarsiTeX according to the place of the letter in the word recognizes to use which form of it.

As depicted in fig. 2, FarsiTeX invokes e-TeX, to produce a dvi file. For this, there is a format file for the extended mode of e-TeX. In every Installation of FarsiTeX, this format file is constructed by initexmf using Lplain.tex (a file of Latex 2.09)

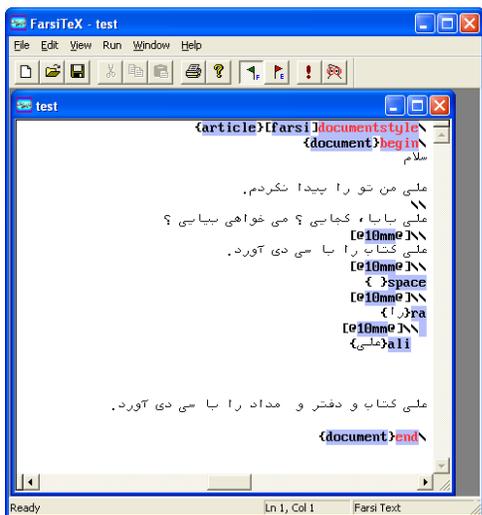
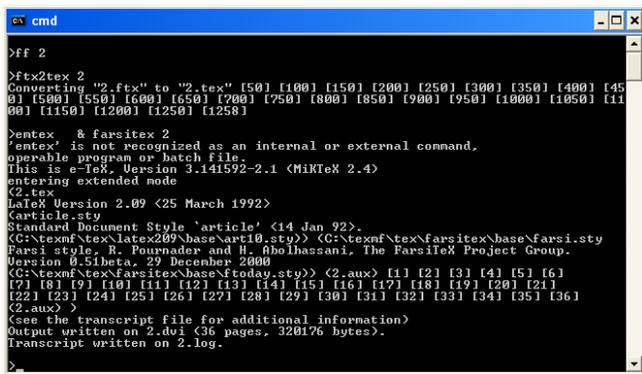


Fig 1: FarsiTeX editor environment

and FarsiTeX.tex (written by the team [4], see Appendix). Here every char code has a special font to be stoned in the dvi file.

The farsi.sty is a style file for input files of FarsiTeX. It is loaded by documetstyle in every input file. It is this file that makes a typeout of a message pointing to the authors in fig. 2. In this style file there is some extensions that is needed for FarsiTeX over an standard style file (such as the article style file of 1992 ). This version of article.sty should be used with FarsiTeX. With the MikTeX setup of LaTeX, this article style file is in the Latex 2.09 folder. FarsiTeX needs this article of 1992, so one usual way to handle this is to copy this style file to the working directory.



```

cmd
>ff 2
>ftx2tex 2
Converting "2.ftx" to "2.tex" [50] [100] [150] [200] [250] [300] [350] [400] [450] [500] [550] [600] [650] [700] [750] [800] [850] [900] [950] [1000] [1050] [1100] [1150] [1200] [1250] [1258]
>entex & farsiTeX 2
'entex' is not recognized as an internal or external command,
operable program or batch file.
This is eTeX, Version 3.141592-2.1 (MikTeX 2.4)
entering extended mode
<2.tex>
LaTeX Version 2.09 <25 March 1992>
Article.sty
Standard Document Style 'article' <14 Jan 92>.
(C:\texmf\tex\latex\209\base\art10.sty) (C:\texmf\tex\fasitex\base\fasir1.sty)
Farsi style, R. Pourmader and H. Abolhassani, The FarsiTeX Project Group.
Version 0.51beta, 29 December 2000
(C:\texmf\tex\fasitex\base\ftoday.sty) <2.aux> [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29] [30] [31] [32] [33] [34] [35] [36] <2.aux> >
(see the transcript file for additional information)
Output written on 2.dvi <36 pages, 320176 bytes>.
Transcript written on 2.log.

```

Fig 2: *FarsiTeX* compilation

### 3 Spell checking in Farsi

Every spell checker needs to have a dictionary to finds spell errors. To have a good spell checker, it should be able to give some suggestion for each error it finds. As mentioned in Section 2 there are various forms for each Farsi letter. Some errors come up because of these various forms. These forms are called :

- o Starting Form
- o Between Form
- o Glued Ending Form
- o Singular Ending Form

In some cases instead of Singular Ending Form of a letter, its Glued Ending Form is left. It is a job of spell checker to correct this error.

### 3.1 The dictionary of the spell checker

Almost every day, Farsi Language is updated to the new words in the translation : New technical words, new medical words, and new constructed words. In some of these cases, non-Farsi words is written just with Farsi alphabet. For this non-Farsi words, there exists some variants. All of these variants can be accepted. So in the design of the dictionary of the spell checker, these cases should be considered.

### 3.2 Meaning related spell errors

In Farsi language there are some couple of words with the same pronunciations, but with the different meanings and the different spellings (Fig. 3). In writing, because of the same pronunciations, a writter may be uses the wrong one instead of the right one. The difficult part of a Farsi spell checker is may be related to the errors that stems from this phenomena. Farsi Spell checker should consider a littel bit of the meaning of the sentences to overcome this problem.

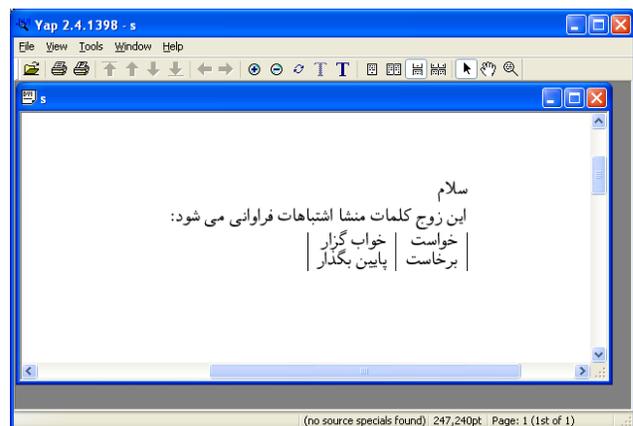


Fig 3: *Two pairs of Farsi words with the same pronunciations*

## 4 Conclusion

In the previous sections FarsiTeX is introduced. The various design issues of having a bi-language typesetting system is shown. Difficulties of spell checking and giving suggestions in Farsi language is discussed.

**Acknowledgements:** We say our best thanks to Dr Majid Ghandchi, for his encourage and reading the manuscript. The research was supported in

part by the Islamic Azad University, Ahar branch and in the case of the second author, it was also supported by the Department of Mathematical Sciences of Tabriz University.

#### References:

- [1] Donald E. Knuth, *TeX : The program*, Reading, Mass: Addison-Wesley, 1986.
- [2] H. Kopka and P. Daly, *A Guide to L<sup>A</sup>T<sub>E</sub>X, Document Preparation for Beginners and Advanced Users*, Addison-Wesley, 1993.
- [3] Donald E. Knuth, *The TeXBook*, illustrations by Duane Bibby, Reading, Mass: Addison-Wesley, 1986.
- [4] M. Ghodsi, *FarsiTeX Guide*, in Farsi Language, Computer department, Sharif University, 1997
- [5] D. E. Knuth and P. Mackay, *Mixing right-to-left texts with left-to-right texts*, TUGboat vol. 8, No. 1, 1987.

## 5 Appendix

This is the first part of the FarsiTeX.tex file, including its license, that is available at :

<http://www.farsitex.org>

```
%%
%% This file is a part of FarsiTeX, a Persian/English typesetting system.
%%
%% Copyright (C) 1996 Hassan Abolhassani
%% Copyright (C) 1996-2001 Roozbeh Pournader <roozbeh@sharif.edu>
%%
%% FarsiTeX is free software; you can redistribute it and/or modify
%% it under the terms of the GNU General Public License as published by
%% the Free Software Foundation; either version 2 of the License, or
%% (at your option) any later version.
%%
%% FarsiTeX is distributed in the hope that it will be useful,
%% but WITHOUT ANY WARRANTY; without even the implied warranty of
%% MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
%% GNU General Public License for more details.
%%
%% You should have received a copy of the GNU General Public License
%% along with FarsiTeX; if not, write to the Free Software
%% Foundation, 59 Temple Place, Suite 330, Boston, MA 02111-1307,
%% USA.
%%
%% Any licensing or usage questions should be directed to Roozbeh
%% Pournader <roozbeh@sharif.edu>.
%%
:
```