TEX-e-Parsi changes for RTL and LTR texts processing

Vafa Khalighi

December 27, 2008

Contents

1 Introduction

We need to add some features to the actual TEX engine for typesetting Persian texts. We divide these features into two groups but in fact the first group is what at the moment needs to be done:

1. tools for processing texts and RTL and LTR structure.

2. Translating TEX primitive commands and messages to Persian.

1.1 Tools for processing RTL texts

If a Persian text was just from right to left, we did not have to change TEX at all, because it was just fine to rotate the typeset text in the output about half of the page so we could obtain the required change. But Persian texts are bidirectional, which means we read words from right to left but read numbers from left to right. And nowadays, there are many Persian translations of the western text books in which in addition to Persian texts we also need to insert some Western texts as well which are typeset from left to Right. In addition to these in scientific texts, mathematical formulas like world standard, should appear from left to right.

Thus for typesetting Persian, we need to provide some tools. These tools are:

1.1.1 Commands for determining Direction of the text

With the commands `\beginR`, `\endR`, `\beginL` and `\endL` we can determine the start and the end of RTL and LTR texts. These commands can be used in any combination.

1.1.2 Swtiches for Controlling Direction

However we can typeset bidirectional texts with the above commands, but producing an input file which contains these commands is not easily achievable.From the other side, in most cases, we can determine the direction of the texts automatically based on the characters of input file, with defining a suitable collection of characters. For this purpose TEX-e-Parsi has defined some switches and the automatic text direction will be controlled based on these switches. These switches are:

1.1.3 Text direction and current font

TEX has a current font in any moment that will typeset normal characters with that font. But TEX-e-Parsi can have two current font and will typeset Persian characters with one font and Latin characters with the other font. In addition, TEX-e-Parsi can add automatically the commands for determining text direction, based on Persian and Latin characters. The usage quality of these features will depend on the value of `\curLRswch` which is a numerical parameter. This parameter has two checking bits. The first determines `\autoLRdirset` and the second one will determine `\autofontset`. The four reliable value of this parameter are defined by the following commands:

value	**Command Name**	**Detail**
0	`\manLRset`	for LTR text only
1	`\autoLRdirset`	for bidirectional text and one current font
2	`\autofontset`	for one directional text but with two current font
3	`\autoLRset`	for bidirectional text and two current font

Please note that `\curLRswch` can only be used as a number and the change of its value is only possible with the above commands.

If we have `\autoLRdirset`, then characters 0 to 127 as LTR characters and characters 128 to 255 as RTL characters, will be assumed and based on this the direction of the typesetting will be determined. For example if the typesetting direction is RTL and ز is observed as a character with number 97, then the direction of the typesetting will be LTR, and this condition will continue, until a character with a number greater than 127 is observed and in this case, the direction of the typesetting will be RTL again.

1.1.4 The direction of the lines in the page

TEX adjusts lines in the page from left to right. This means if the length of lines are different, the start of all the lines will be in the left hand side and their right hand side will be fitful. TEX-e-Parsi can adjust lines in the page both from left and right.

The quality of this job will depend on the numerical parameter value of `\curboxdir`. The values of this parameter will be determined by the commands `\rightvbox` and `\leftvbox`. The effect of these commands will be bounded in current vbox. This means if we have several vobx in which some of these vbxes are the aliquot part of some other vboxes, these commands will have effect only in the vbox that they are present in. If non of these two commands is determined, then `\leftvbox` will be assumed.

1.1.5 Adjusting the direction of vboxes

Usually, in normal texts, all vboxes are adjusted in one direction and user prefers to determine the direction of all vboxes at once. The parameter `\vboxjsutification`, which is a numerical parameter, determines the direction of all vboxes as follow:

- a value of this parameter less than 0 means all vboxes will be adjusted from left.
- a value of this parameter greater than 0 means all vboxes will be adjusted from right.
- a value of this parameter bigger than 1 means that in addition to adjusting all vboxes from right, the parameter `\leftskip` will be placed in the right hand side and the parameter `\rightskip` will be placed in the left hand side.
- it is clear that if the value of this parameter is 0, then the direction of vbox will be adjusted based on the parameter `\curvboxdir`.

It is vital to note that, we always can determine the direction of current vbox with the parameter `\curvboxdir`, independent of the parameter `\vboxjustification`.

1.1.6 Adjusting the direction of tables

Persian tables have a main difference with Latin tables and that is the oder of the placement of their columns. In Persian tables, columns are adjusted from right to left but columns of Latin tables are placed from left to right. However in simple cases, by puting `\beginR` and `\endR` at the start and the end of the lines of the table, we can obtain the order that we want in Persian tables but because of the following reasons, this approach is not usable:

- in the cases that the number of columns that we put is less than the number of normal columns, TEX will produce a line with that small number and thus the current approach of horizontal boxes from right to left can not keep the adjustment of columns. To solve this issue, TEX-e-Parsi will produce all lines of the table completely, meaning it puts empty boxes for the columns which are not inserted.
- in the actions which will be done by opening table (like environment for typesetting traditional Persian poems), the assumption is that every line of the table is only made of some boxes and if we put the commands `\beginR` and `\endR` in it, then doing those actions become impossible. So TEX-e-Parsi has defined the command `\semihalign` so that the change of orders of columns are done without the use of left and right commands.

The blow commands are defined for using fonts in TEX-e-Parsi:

`\cursemifont`: used for current Farsi font
`\curfont`: used for current Latin font
`\activefont`: used for the active current font (Farsi or Latin)
`\fonttwin`: used for twin current Farsi font
`\dblfont`: used for introducing the current Farsi font which has 256 characters and characters of second half are defined instead twin font.

The conditional commands are defined for ease in programming and creating new features:

`\ifjoinable`: not needed, because determines joinablity of the character before
`\ifLtoR`
`\ifRtoL`
`\iflatin`
`\ifleftvbox`
`\ifonesof`
`\iftens of`
`\ifhundredsof`
`\ifthousands`
`\ifmillions`
`\ifbillions`

The following commands are defined for creating special Persian language features:

`\semiday` (not needed, can be acchieved easily in the macro level)
`\semimonth` (not needed, can be achieved easily in the macro level)
`\semiyear` (not needed, can be achieved easilly in the macro level)
`\semispaceskip`
`\semixspaceskip`
`\lastcharjoinable` (not needed)
`\lastcharunjoinable` (not needed)
`\semiaccent` (not needed)
`\semiaccentdown` (not needed)
`\retainaccentchar` (not needed)

1.1.7 Features of characters

In TeX-e-Parsi the foloowing features for characters are defined:

`\lcode`
`\jattrib` (not needed)
`\accfactor` (not needed)
`\eqchar`
`\eqcharif`