

Hidden Problems in Bi-Directional Word Processing

Problem 1:

Undirected (Neutral) spaces, international punctuation, tabs, etc.

The Unicode Bi-Directional Algorithm <<http://www.unicode.org/reports/tr9/>> makes assumptions about the direction of these neutral characters that are correct most of the time, but there are exceptions.

E.g., lower case is L-R, UPPER CASE is R-L, and # is space
Yellow is LR and Blue is RL

TIME ORDER:

he#said#SHALOM#DANIEL#BERY#to#daniel.

First, words get directions based on their characters

he#said#SHALOM#DANIEL#BERY#to#daniel.

Then each neutral character gets its direction based on an algorithm that looks at the directions of its neighbors.

he#said#SHALOM#DANIEL#BERY#to#daniel.

That is, the spaces between two RL words is RL and the space at a boundary between EMBEDDING LR text and EMBEDDED RL text is LR, and the period at the end of a LR word is LR.

So that "SHALOM#DANIEL#BERY" is the chunk or run of LR characters, and the visual order of the line (assuming no line breaks in the middle and LR document) is:

he said YREB LEINAD MOLAH'S to daniel.

Note that the algorithm does what is USUALLY wanted.

But suppose you want the space or tab between two RL words to be LR, as in a LR table of RL items

@ is a tab.

TIME ORDER:

blah#ONE@TWO@THREE#wow

By the default algorithm, each space is LR, but each tab is RL

blah#ONE@TWO@THREE#wow

causing visual order of:

blah EERHT OWT ENO wow

but we want visual order of

blah ENO OWT EERHT wow

i.e., each word is rotated in its place in a LR document

Unicode provides three codes for FORCING a direction regardless of what direction, LR, RL, or neutral, the text has

LRO starts a region of LR text that is ended by the next RLO or PDF

RLO starts a region of RL text that is ended by the next LRO or PDF

Each of LRO, RLO, and PDF is zero width.. invisible in output.

So to get the desired behaviour with

blah#ONE@TWO@THREE#wow

you replace each "@" with "LRO@PDF". Then the @s are forced to be LR,

blah#ONE@TWO@THREE#wow

and you get the desired visual order:

blah ENO OWT EERHT wow.

Note that LRO, RLO, and PDF can be used to force ANY character to have ANY direction. So you could use it if you want to make a font of Latin letters be treated as RL characters, .e.g, to prepare the "Brief Bi-Directional Text Reading Lesson". This forcing is Problem 2.

Problem 3:

The strange behavior of the selected region when moving a mouse across a text direction change boundary.

The characters that are selected in a region should be adjacent to each other in the

time ordered storage.

Look at what happens as the selected region that was growing L-R in a sweep of the mouse starting in LR text as the sweep crosses the directional boundary:

1. VISUAL ORDER:

he said YREB LEINAD MOLAHS to daniel.

TIME ORDER:

he#said#SHALOM#DANIEL#BERY#to#daniel.

2. VISUAL ORDER:

he said YREB LEINAD MOLAHS to daniel.

TIME ORDER:

he#said#SHALOM#DANIEL#BERY#to#daniel.

3. VISUAL ORDER:

he said YREB LEINAD MOLAHS to daniel.

TIME ORDER:

he#said#SHALOM#DANIEL#BERY#to#daniel.

4. VISUAL ORDER:

he said YREB LEINAD MOLAHS to daniel.

TIME ORDER:

he#said#SHALOM#DANIEL#BERY#to#daniel.

5. VISUAL ORDER:

he said YREB LEINAD MOLAHS to daniel.

TIME ORDER:

he#said#SHALOM#DANIEL#BERY#to#daniel.

6. VISUAL ORDER:

he said YREB LEINAD MOLAHS to daniel.

TIME ORDER:

he#said#SHALOM#DANIEL#BERY#to#daniel.

Solutions?:

1. Just live with the cognitive dissonance caused by the selection region growing in the direction opposite to that in which the mouse is moving.
2. If the user hits the ESCAPE key while the mouse is crossing a directional boundary, the user may move the mouse in the other direction to continue the sweep in time order.
3. Show the portion that is selected in time order temporarily until the sweep stops and then switch to showing the disconnected selected region in time order.
4. Do 1 above, but independently provide a way for the user to show the current view in time order. Since the file is stored in time order, visual order is ONLY a view and it does not change the file to change the view. This view switch would be in the "View" menu.

This ability to show the time order is useful in general, when the input is complex.