Extending Bidi Support on the Web

Richard Ishida, W3C
Aharon Lanin, Google
Bidi support on the Web

Additional Requirements for Bidi in HTML

Read the Working Draft at http://www.w3.org/International/docs/html-bidi-requirements/

You can follow all discussions on the public-i18n-bidi@w3.org list.

It's about CSS too.

And some suggestions for browser developers.
This is a report on ONGOING discussions. You MUST NOT assume that these things work or will work in the future.

22 bugs were raised against HTML5 for Oct 1st, 2010 - these are currently under discussion.

Additional proposals will be sent to CSS WG, and some are already being discussed and/or implemented.

If you feel strongly about these features, track and join the discussion!
Outline

1. Some background
2. Feature proposals
3. How to track developments/get involved
Some background
<p>MAKE html WORK FOR YOU</p>

LTR  EKAM html UOY ROF KROW

RTL  UOY ROF KROW html EKAM
<p>10 Main st. </p>

LTR 10 Main st.

RTL .10 Main st
Background

UBA 'paragraphs'

IT IS IMPORTANT TO LEARN html.
(css IS IMPORTANT TOO.)

html. NRAEL OT TNATROPMI SI TI (.OOT TNATROPMI SI (css

.html NRAEL OT TNATROPMI SI TI (.OOT TNATROPMI SI css)
New HTML & CSS proposals
Bidi isolation
Issue

Inline element content often needs to be isolated – especially when data is imported into the page.
Issue

Inline element content often needs to be isolated – especially when data is imported into the page.
Note

Existing bidi markup in HTML is not designed to solve this problem.

<span dir=rtl>ليلית</span> - 3 reviews

<span dir=rtl>ليلית</span> - 3 reviews
Bidi isolation

Note

Existing bidi markup in HTML is not designed to solve this problem.

LRM/RLM can help, but it's not as good as using markup.

Problems with LRM/RLM

• RLM doesn't speak to the structure of the document.

• Adding logic to an application to choose the LRM vs RLM can be difficult.

• LRM/RLM is often not ignored during search or may be picked up in error during copy & paste, causing failures.

• LRM/RLM does nothing about unbalanced LRE/RLE/LRO/RLO/PDF in the inserted string.
Bidi isolation

Proposal

New attribute: ubi = off | on/ubi

New CSS value unicode-bidi: isolate

The result is an embedded UBA 'paragraph' that is treated as a neutral character within the surrounding text.

<p><span ubi>ليلית</span> - 3 reviews</p>
Reporting direction
Users can specify direction for form input, but it is immediately forgotten.
Proposal

New attribute, `submitdir`, on input and textarea elements.

Standard `_dir` suffix for control name

```html
<form action=foo method=get>
  <input type=text name=mytest dir=auto submitdir />
</form>
```

`foo?mytest=texttexttext&mytest_dir=ltr`
dir=auto value
Sometimes the appropriate direction can only be determined at run-time.
Sometimes the appropriate direction can only be determined at run-time.
Proposal

New value 'auto' for dir attribute.

New attribute:
autodirmethod=first-strong | any-rtl

<html autodirmethod=first-strong>
...

<p>Your search - <span class=booktitle dir=auto>יוט דזימ תורעהה CSS</span> - did not match any documents.</p>
Flipping images
Flipping images

Issue

Handed images need to be flipped during localization. It is time-consuming and error-prone to do this using separate image files.
For images in `<img>` markup, you may be able to use CSS3 transformations.

```
.myimage:rtl { transform: scaleX(-1); }
```

This doesn't work with images inserted using CSS.
New keywords for CSS3 images for things such as background-image and list-style-image.

Option 1: rtlflip, eg.

```css
background-image: url(mygraphic.png) rtlflip;
```

Option 2: rtl | ltr, eg.

```css
background-image: url(mygraphic.png) ltr;
```
Line breaks in plain text
Browsers don't consistently treat line breaks as a UBA 'paragraph' break in pre and textarea elements, as well as script dialog text.

LINE SEPARATOR (U+2028) and PARAGRAPH SEPARATOR (U+2029) characters are not recognized by browsers.
Non-collapsed line breaks should always break paragraphs for UBA.

LINE SEPARATOR (U+2028) and PARAGRAPH SEPARATOR (U+2029) characters should behave as defined in Unicode.
<br> as a bidi separator
HTML4 defines `<br>` as equivalent to white-space, but most of the time, as used in the wild, `<br>` needs to terminate a UBA 'paragraph'.

Browsers don't do that consistently.

<p>IT IS IMPORTANT TO LEARN html. <br> (css IS IMPORTANT TOO.)</p>
Note

Using markup properly is the best solution, but it may be unrealistic to expect that to happen.

<body g_editable="true" hidefocus="true" class="editable LW-yrriRe" id=":6y" style="">IT IS IMPORTANT TO LEARN html. <br/>(css IS IMPORTANT TOO.) <br/>

...
Add a new attribute: `bidibreak=hard|soft`.

Value is inherited by `<br>` elements.

The default for the root element is 'hard'.

Value 'soft' could provide control for exceptions/legacy.
Script dialog text direction
Browsers are inconsistent in the direction they apply to dialog text.

Using Unicode controls can be problematic.
Text passed to dialog controls without explicit direction should be displayed according to the first strong character in each 'paragraph'.

The page at http://www.w3.org says:

W3C

W3C
<title> should support the dir attribute
<title> should support the dir attribute

Issue

Browsers don't use directional information from the page to display <title> text correctly.

<html dir=rtl>
<head>
<title>Egypt Israel deutschland france</title>
...

✗
✓
Proposal

The HTML specification should stipulate that title text be displayed in the `&lt;title&gt;` element's computed direction.

```html
<html dir=rtl>
<head>
<title>Egypt Israel deutschland france</title>
...
```
title and alt attribute
text direction
It should be possible to have different base directions for an element's content vs. its alt or title attributes.
It should be possible to have different base directions for an element's content vs. its alt or title attributes.
Proposal

New attribute: `titledir=ltr|rtl|auto`.

New attribute: `altdir=ltr|rtl|auto`.

Spec to state that alt and title text otherwise take on the base direction of the context.
<option> and
text direction
Most browsers don’t take into account a dir attribute on an `<option>` element. The one that does doesn’t apply it after the option has been chosen.

Also, it lets the `<option>`’s dir affect its alignment, rather than taking the alignment from the direction of the `<select>` element.

```html
<select>
  <option value=1>deutschland</option>
  <option value=2>france</option>
  <option value=3 dir=rtl>TPYGE (egypt)</option>
  <option value=4 dir=rtl>LEARSI (israel)</option>
</select>
```
The computed direction of an `<option>` element will take into account any `dir` attribute, and it will be applied to its content both in the dropdown and after being chosen.
List item rendering
Bullets and numbering are not displayed as expected, nor consistently across browsers.
Bullets and numbering should always follow the direction set on the `<ol>` or `<ul>` element.

By default, list text should be aligned according to direction.
Other topics...
• Block elements as bidi separators

• `<input type=text>` and `<textarea>` should support interoperable ‘set direction’ functionality

• A page’s overall vertical scrollbar should be on the ‘end’ side relative to the user agent chrome direction

• The vertical scrollbar of an element below `<body>` should be on the ‘end’ side relative to the element’s direction
How to track developments/get involved
Read the document:
http://www.w3.org/International/docs/html-bidi-requirements/

Follow/contribute to the discussions:
http://lists.w3.org/Archives/Public/public-i18n-bidi/

Track HTML5 bugs:
http://www.w3.org/International/reviews/html5-bidi/