



Extending Bidi Support on the Web



Richard Ishida, W3C
Aharon Lanin, Google

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.



Additional Requirements for Bidi in HTML

W3C Working Draft 4 March 2010

This version:

<http://www.w3.org/TR/2010/WD-html-bidi-20100304/>

Latest version:

<http://www.w3.org/TR/html-bidi/>

Editor:

Aharon Lanin, Google

Additional Contributors:

Adil Allawi, Technical Director, Diwan Software
Matthieu Allouche, Bidi Architect, IBM
Uri Bernstein, Google
Douglas Davidson, Apple
Mark Davis, Senior I18n Architect, Google; President of the Unicode Consortium
Martin J. Dürst, W3C I18n Interest Group Chair
Asmus Freytag, President, ASMUS, Inc.
Richard Ishida, I18n Lead, W3C
Shanjian Li, Google
Mohamed Mohie, IBM
Jeremy Moskovich, Google
Shachar Shemesh, Lingnu Open Source Consulting
Gaal Yahas, Google

Copyright © 2007-2010 W3C® ([MIT](#), [ERCIM](#), [Keio](#)), All Rights Reserved. W3C [liability](#), [trademark](#) and [document use](#) rules apply.

Abstract

Authoring a web app that needs to support both right-to-left and left-to-right interfaces, or to take as input and display both left-to-right and right-to-left data, usually presents a number of challenges that make it an especially laborious and bug-prone task. Some of these are due to browser bugs, but some can be traced to a gap in the specification of

these are due to browser bugs, but some can be traced to a gap in the specification of number of challenges that make it an especially laborious and bug-prone task. Some of to take as input and display both left-to-right and right-to-left data, usually presents a Authoring a web app that needs to support both right-to-left and left-to-right interfaces, or

Health warning!

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



Base direction

<p>MAKE html WORK FOR YOU</p>

LTR EKAM html UOY ROF KROW

RTL UOY ROF KROW html EKAM



Base direction

<p>10 Main st.</p>

LTR 10 Main st.

RTL .10 Main st

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

Bídí ísolatíon


Issue

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

TOP RATED RESTAURANTS

Aroma - 3 reviews
★ ★ ★ ★ ★

לילית - 5 reviews
★ ★ ★ ★ ★



TOP RATED RESTAURANTS

Aroma - 3 reviews
★ ★ ★ ★ ★

5 - לילית reviews
★ ★ ★ ★ ★



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.

לילית - 3 reviews

`לילית - 3 reviews`



Note

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

לילית - 3 reviews

`לילית`

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

ליליתLRM - 3 reviews

Problems with LRM/RLM

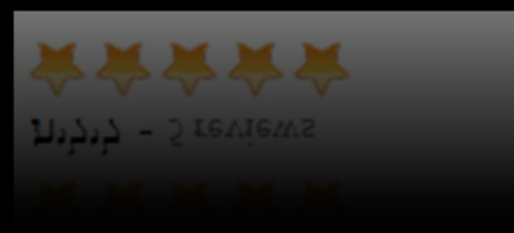
- RLM doesn't speak to the structure of the document.
- Adding logic to an application to choose the LRM vs RLM can be difficult.
- LRM/RLM is often not ignored during search or may be picked up in error during copy&paste, causing failures.
- LRM/RLM does nothing about unbalanced LRE/RLE/LRO/ RLO/PDF in the inserted string

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>לילית - 3 reviews</p>`

Reporting direction

Issue

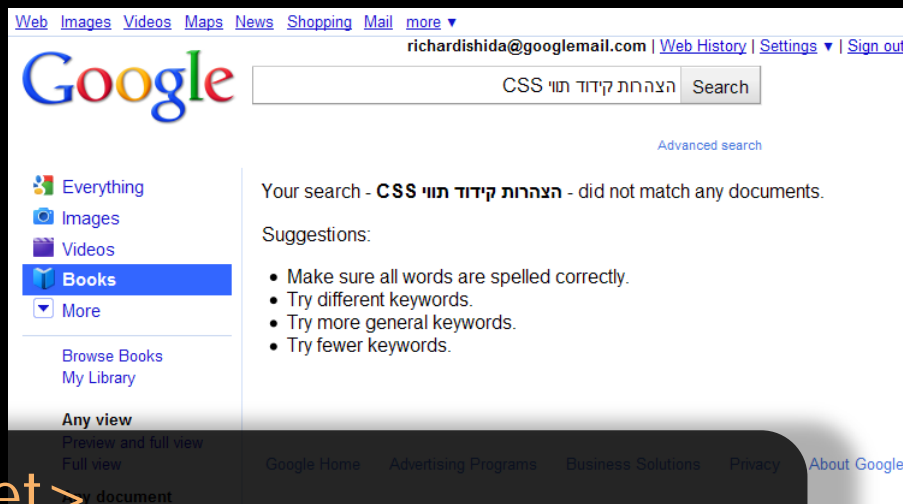
users can specify direction for form input, but it is immediately forgotten.



Proposal

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

Standard `_dir` suffix for control name



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



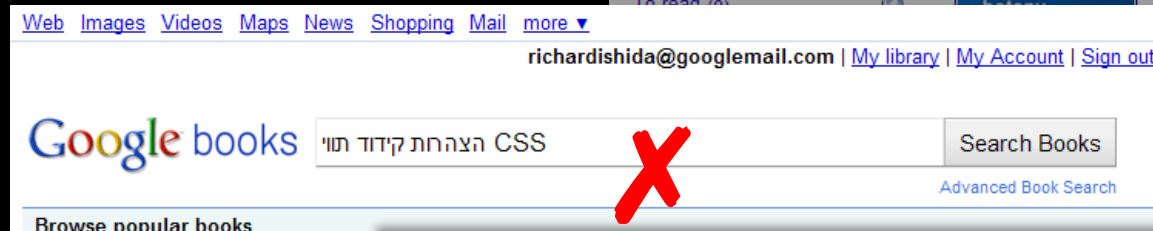
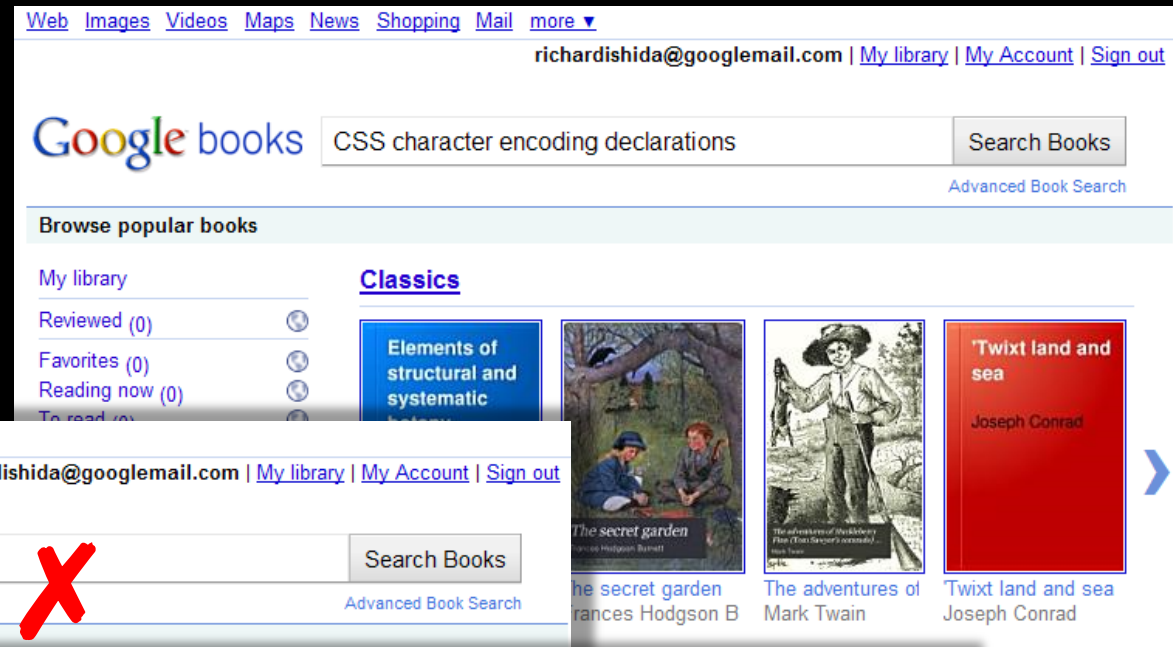
```
foo?mytest=texttext&mytest_dir=ltr
```

dir=auto value

dir=auto value

Issue

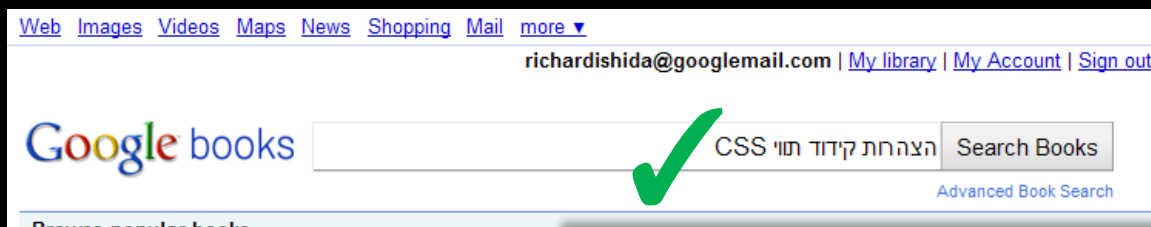
Sometimes the appropriate direction can only be determined at run-time.



dir=auto value

Issue

Sometimes the appropriate direction can only be determined at run-time.



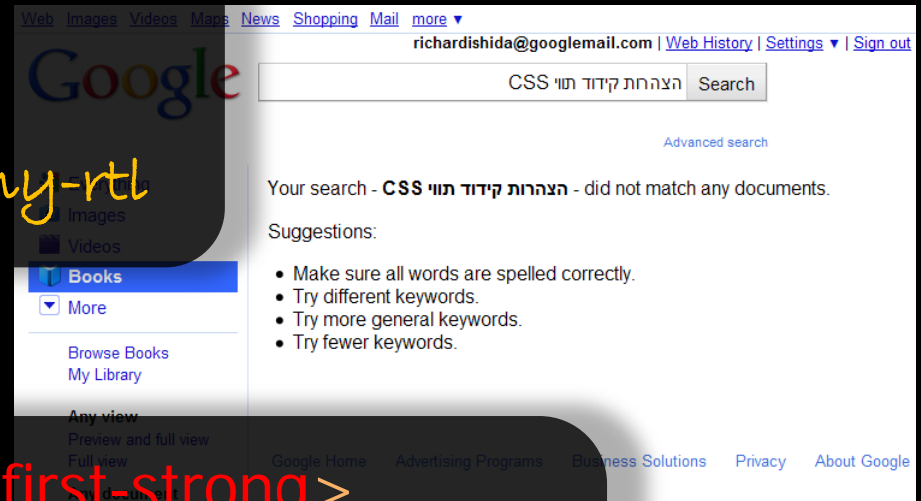
dir=auto value

Proposal

New value 'auto' for dir attribute.

New attribute:

autodirmethod=first-strong | any-rtl



`<html autodirmethod=first-strong>`

...

`<p>Your search - CSS - did not match any documents.</p>`

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.

Anna Karenina



★★★★☆

4415 Reviews

[Write review](#)

[About this book](#)

Anna Karenina

★☆☆☆☆

4415 ביקורות

כתוב ביקורת

אודות ספר זה



Note

For images in `` markup, you may be able to use CSS3 transformations.



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

This doesn't work with images inserted using CSS.

Proposal

New keywords for CSS3 Images for things such as background-image and list-style-image.



Option 1: rtlflip, eg.

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

Option 2: rtl | ltr, eg.

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

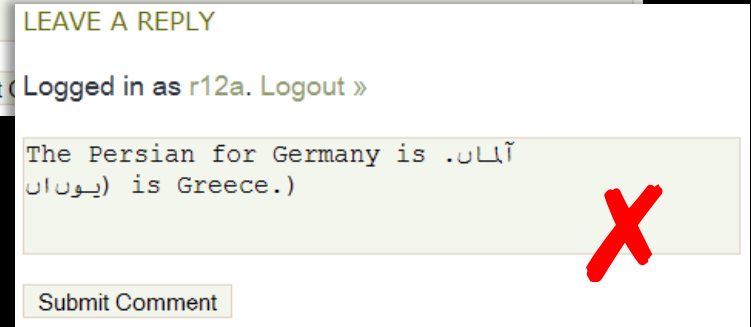
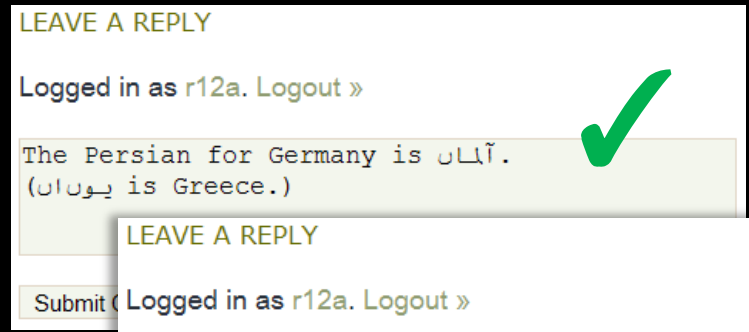
Line breaks in
plain text



Line breaks in plain text

Issues

Browsers don't consistently treat line breaks as a UBA 'paragraph' break in pre and textarea elements, as well as script dialog text.



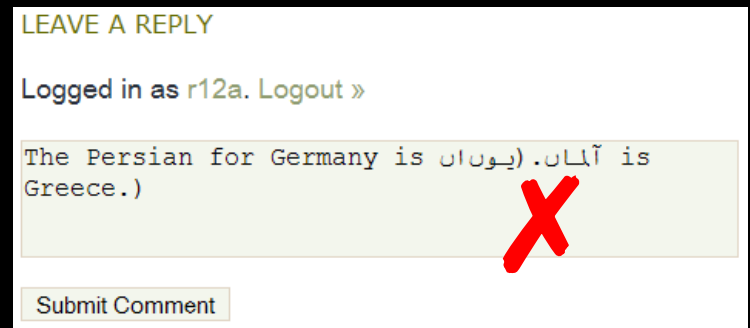
LINE SEPARATOR

(U+2028) and

PARAGRAPH

SEPARATOR (U+2029)

characters are not recognized by browsers.





Line breaks in plain text

Proposal

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.

LEAVE A REPLY

Logged in as r12a. [Logout »](#)

The Persian for Germany is آلمان.
(یونان is Greece.)

 as a
line separator

Issue

HTML4 defines

 as equivalent
to white-space, but
most of the time, as
used in the wild,

 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>
```

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

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

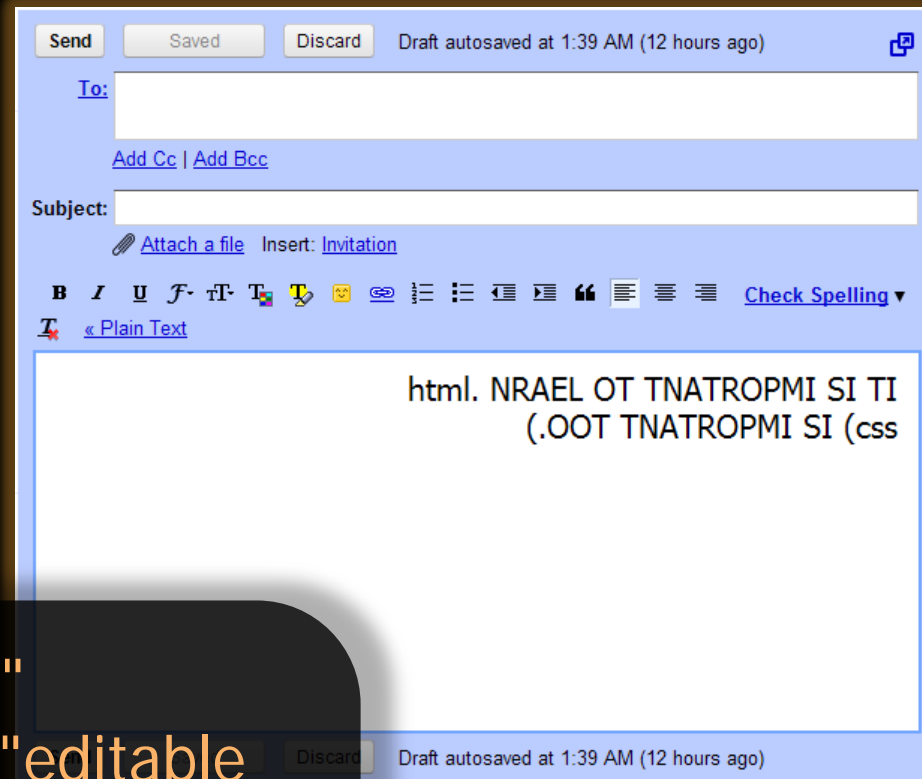

 as a bidi separator

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>
```

...



 as a bidi separator

Proposals

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

value is inherited by

 elements.

The default for the root
element is 'hard'.

value 'soft' could provide
control for
exceptions/legacy.

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

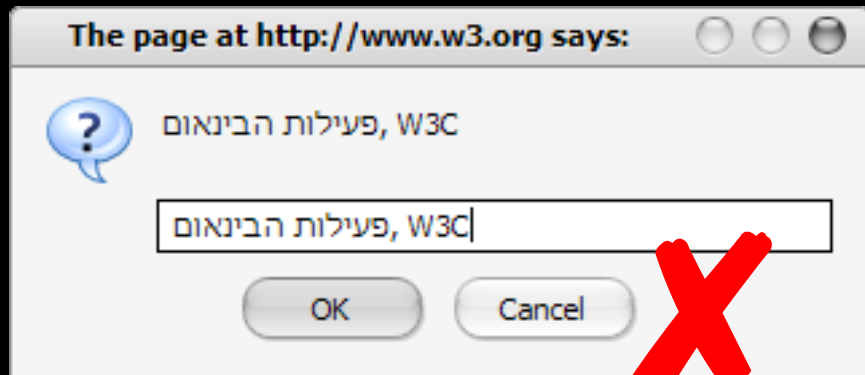
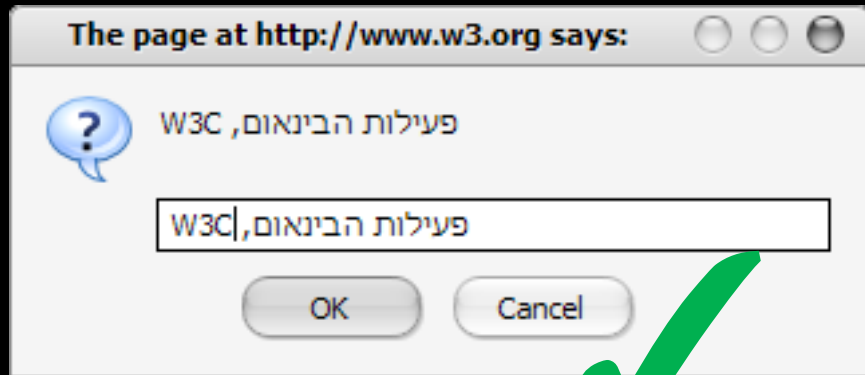
```
<html bidibreak=soft lang=he>
```

Script diálogo text dirección

Issue

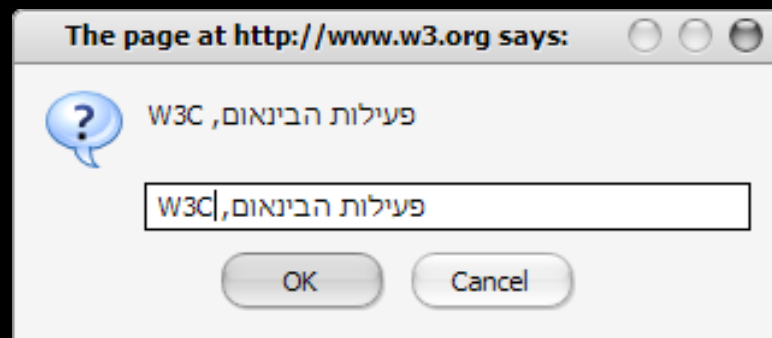
Browsers are inconsistent in the direction they apply to dialog text.

Using unicode controls can be problematic.



Proposal

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

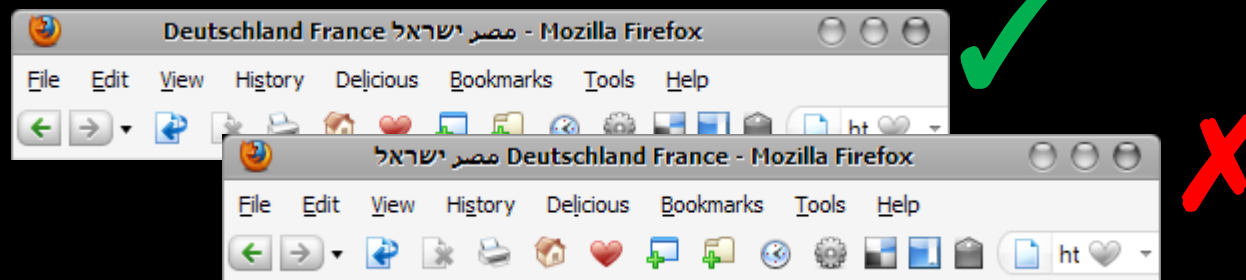


<title> should support
the dir attribute

<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>
```

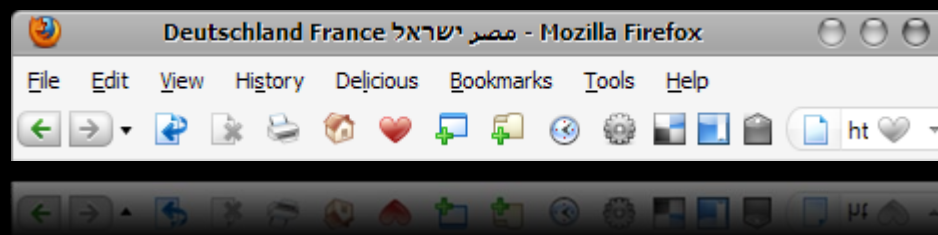
```
<title>EGYPT ISRAEL deutschland france</title>
```

```
...
```

 <title> should support the dir attribute

Proposal

The HTML specification should stipulate that title text be displayed in the <title> element's computed direction.



```
<html dir=rtl>
```

```
<head>
```

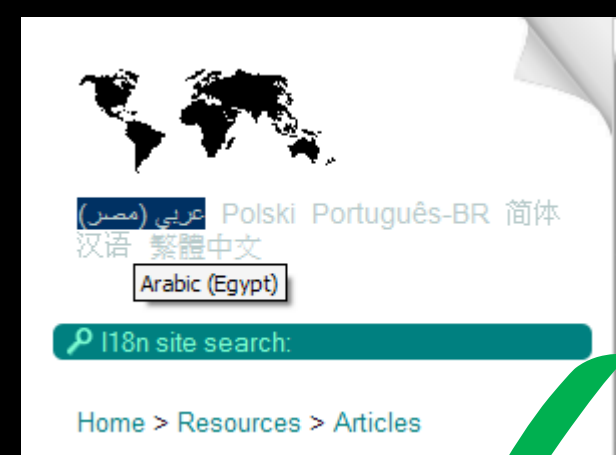
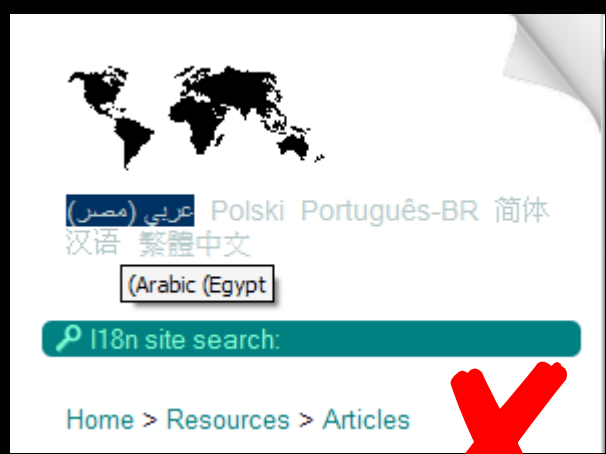
```
<title>EGYPT ISRAEL deutschland france</title>
```

```
...
```

title and alt attribute
text direction

Issue

It should be possible to have different base directions for an element's content vs. its alt or title attributes.



Issue

It should be possible to have different base directions for an element's content vs. its alt or title attributes.

2009/05/26: مكتب W3C المغرب يشارك جمعية الإنترنت المغرب في اليوم الوطني 10 لإنترنت، تحت شعار "الملتقى الوطني للإنترنت: القضايا والتحديات التي تواجه المغرب"، يوم الخميس 26 مايو 2009، الرباط، المغرب. المواضيع الرئيسية التي ستناقش تدور حول الإدارة العالمية للإنترنت والفجوة الرقمية والمفاوضات الدولية الجارية بشأن إدارة أسماء النطاقات، و المعايير و الخدمات التقنية ومستقبل الإنترنت.

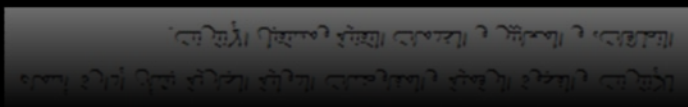
2009/05/26: مكتب W3C المغرب يشارك جمعية الإنترنت المغرب في اليوم الوطني 10 لإنترنت، تحت شعار "الملتقى الوطني للإنترنت: القضايا والتحديات التي تواجه المغرب"، يوم الخميس 26 مايو 2009، الرباط، المغرب. المواضيع الرئيسية التي ستناقش تدور حول الإدارة العالمية للإنترنت والفجوة الرقمية والمفاوضات الدولية الجارية بشأن إدارة أسماء النطاقات، و المعايير و الخدمات التقنية ومستقبل الإنترنت.

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

Issue

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.

```
<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>
```



Proposal

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



Issue

Bullets and numbering are not displayed as expected, nor consistently across browsers.

Default



- Deutschland
- France

•

مصر

•

ישראל

1. Deutschland
2. France

.3

مصر

.4

ישראל

With text-align



- Deutschland
- France

•

مصر

•

ישראל

1. Deutschland
2. France

.3

مصر

.4

ישראל

Proposal

Bullets and numbering should always follow the direction set on the `` or `` element.


By default, list text should be aligned according to direction.

Default

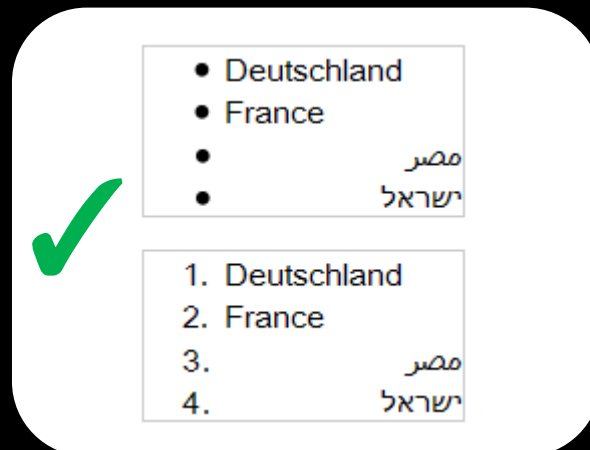


A rounded rectangle containing two list boxes. The top box is a bulleted list with items 'Deutschland' and 'France' on the left, and 'مصر' and 'إسرائيل' on the right. The bottom box is a numbered list with items '1. Deutschland' and '2. France' on the left, and '3. مصر' and '4. إسرائيل' on the right. A large red 'X' is on the left side of the rectangle.

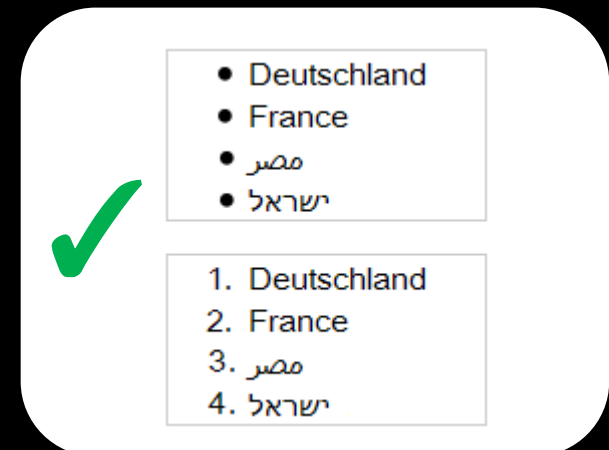
With text-align



A rounded rectangle containing two list boxes. The top box is a bulleted list with items 'Deutschland' and 'France' on the left, and 'مصر' and 'إسرائيل' on the right. The bottom box is a numbered list with items '1. Deutschland' and '2. France' on the left, and '3. مصر' and '4. إسرائيل' on the right. A large red 'X' is on the left side of the rectangle.



A rounded rectangle containing two list boxes. The top box is a bulleted list with items 'Deutschland' and 'France' on the left, and 'مصر' and 'إسرائيل' on the right. The bottom box is a numbered list with items '1. Deutschland' and '2. France' on the left, and '3. مصر' and '4. إسرائيل' on the right. A large green checkmark is on the left side of the rectangle.



A rounded rectangle containing two list boxes. The top box is a bulleted list with items 'Deutschland' and 'France' on the left, and 'مصر' and 'إسرائيل' on the right. The bottom box is a numbered list with items '1. Deutschland' and '2. France' on the left, and '3. مصر' and '4. إسرائيل' on the right. A large green checkmark is on the left side of the rectangle.

Other topics...

- Block elements as bidirectional 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/>