

# Get Your Head Around Bidirectionality!



**Behnam Esfahbod**  
Software Engineer

Quora

## Abstract

We know when the software is broken for a right-to-left languages like Arabic, Persian, or Hebrew, but often the solution is either not clear, or fixing it with out-of-place patches won't worth the costs down the road. Like other areas of i18n, bidirectional layout and right-to-left language support need deliberate design in the user-interface stack, and without good architecture it won't be useful for the developers or the users.

In this tutorial, we first learn how to think in right-to-left and how it mirrors into left-to-right directionality. We then look at the common problems in bidirectional applications and how to address them with generic solutions and standard algorithms.

This tutorial is suitable for anyone not familiar with right-to-left languages or bidirectional design, or interested to learn how to develop solutions for this area.

42nd  
Internationalization &  
Unicode Conference

September 2018

Santa Clara, CA, USA



---

## About me

- **Software Engineer @ Quora, Inc.**
- **Co-Chair of Arabic Layout Task Force @ W3C i18n Activity**
- **Virgule Typeworks**
- **Facebook, Inc.**
- **IRNIC Domain Registry**
- **Sharif FarsiWeb, Inc.**





---

## **This talk**

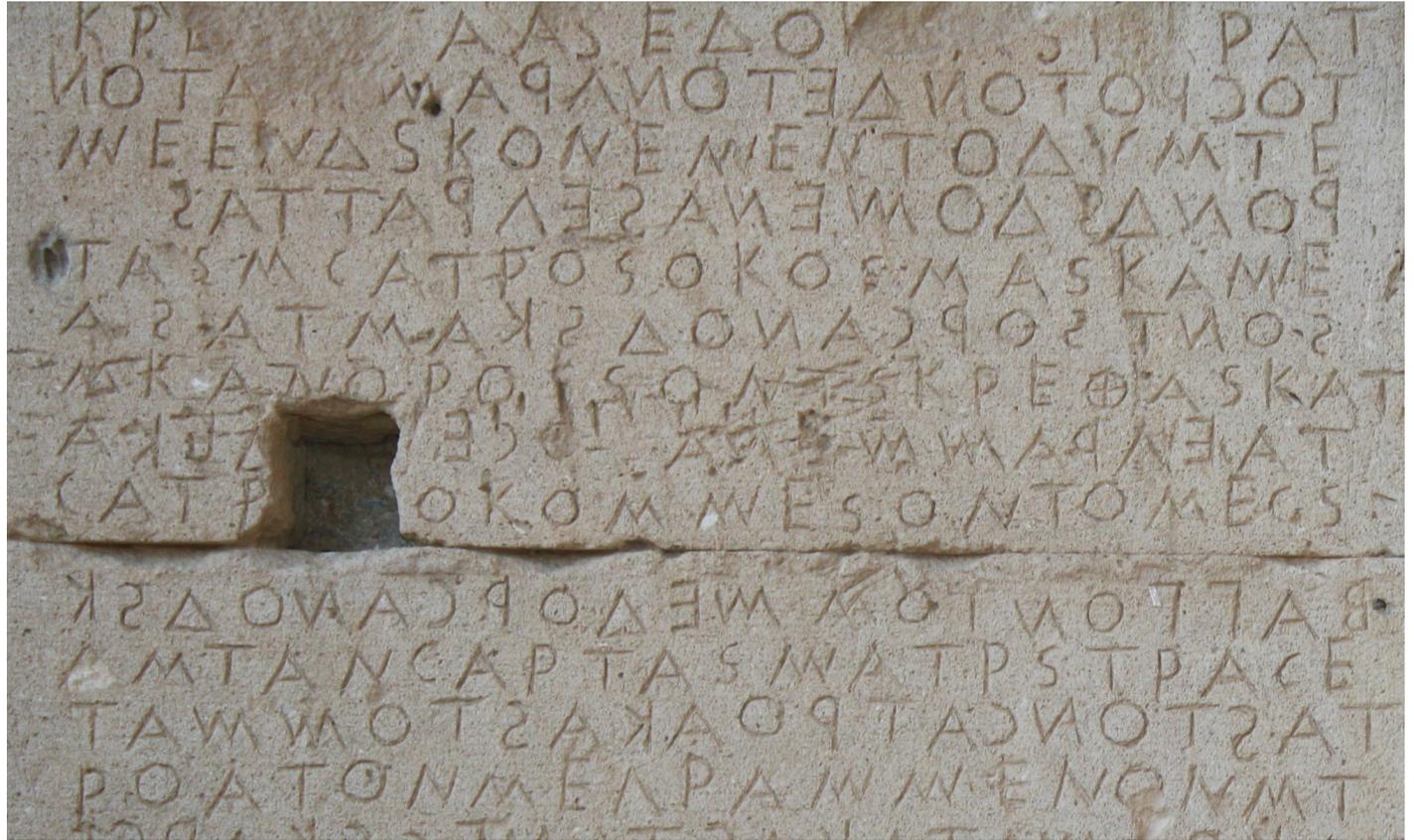
- **Bidirectional Writing Systems**
- **Bidirectional Text**
- **Bidirectional Layout**
- **Bidirectional Web Application**
- **Bidirectionality Techniques**

# **Bidirectional Writing Systems**

## History

### Boustrophedon

from Greek  
“*boustrophēdón*”  
meaning  
“ox-turning”



## History

Line direction  
alternates.  
No paragraph  
direction.

Q: Why's this useful?

Q



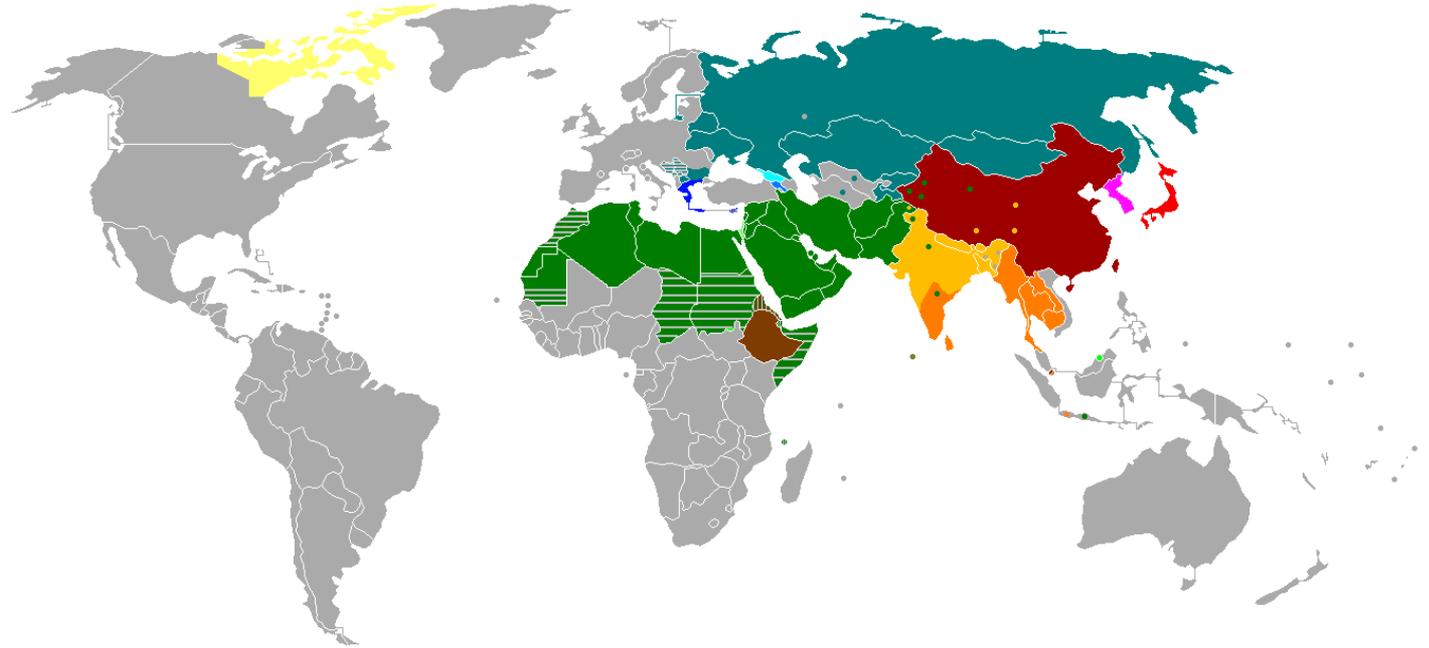
# History

## Early Writing Systems

- **Most scripts chose one way or another**
- **Small set of writing symbols**
  - Letters, e.g. Greek Alpha or Arabic Alef
  - Limited punctuations
  - No numerals: roman and abjad numbers
- **Later, Hindu-Arabic numerals**
  - Not (normally) read digit-by-digit
  - Spelled out as a (whole) number
  - Therefore: no direction in reading a numbers!

# Today

Writing systems at national level



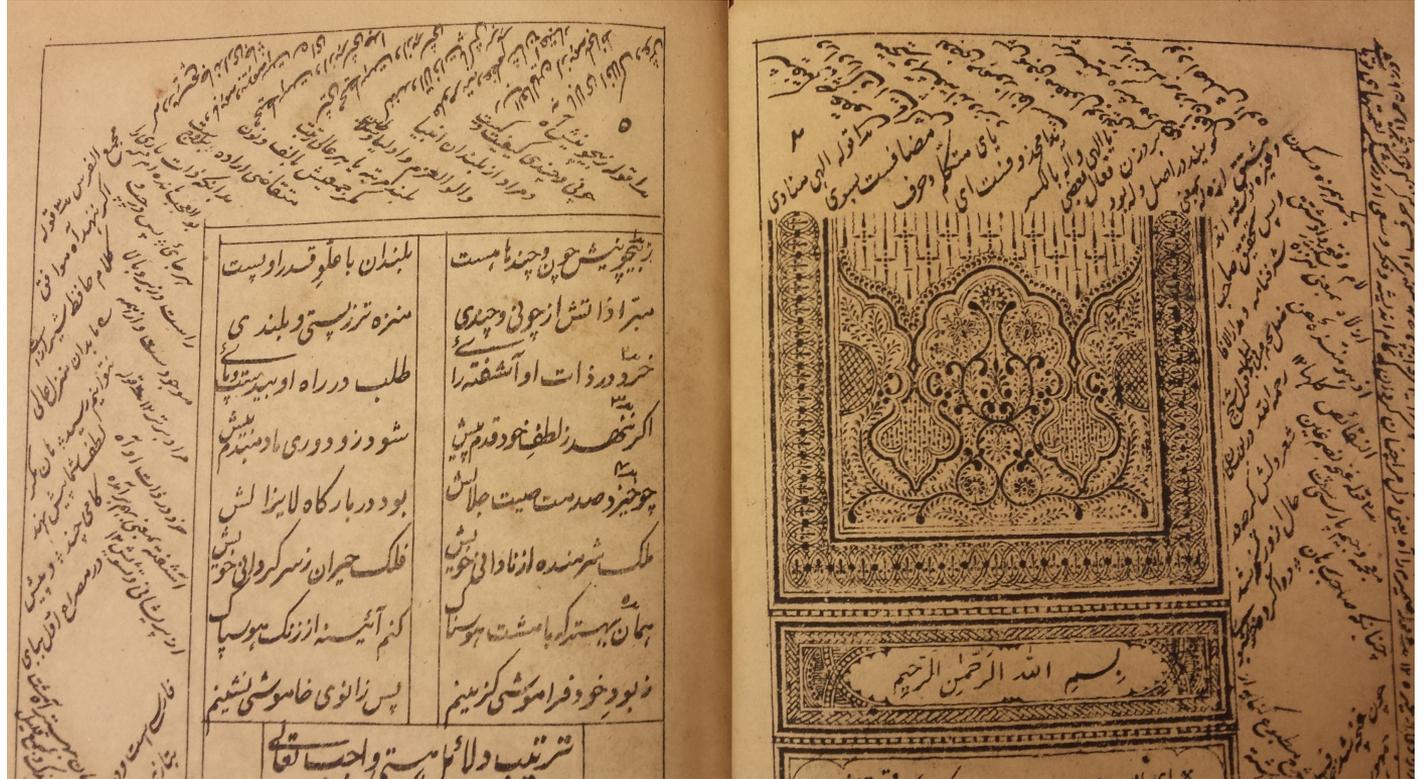
# Today

## Digital encoding

- **Unicode ≈ unique, unified, universal encoding**
- **About 150 scripts encoded in Unicode:**
  - ~110 left-to-right (**LTR**) (some could also be top-to-bottom)
  - ~30 right-to-left (**RTL**) (some are bidi...)
  - the rest are top-to-bottom, or mixed directions
- **Major *unified* scripts**
  - CJK: Chinese, Japanese, Korean
  - Arabic: Standard/Maghrebi Arabic, Persian, Urdu, Jawi, Uyghur, ...
- **Major *non-unified* scripts**
  - Latin/Greek/Cyrillic

# **Bidirectional Text**

# Manuscript text & layout



## Semantic encoding in Unicode

Store text in memory in the same order as is read/processed in mind



- **Encode concepts, not various shapes of them**
  - One Arabic Letter Alef (U+0627)
    - Most Arabic letters take at least 4 shapes depending on context
  - But, two Latin Letter A (oops!)
    - LATIN CAPITAL LETTER A (U+0041) / LATIN SMALL LETTER A (U+0061)

## Semantic encoding in Unicode

Store text in memory in the same order as is read/processed in mind



- **Encode concepts, not various shapes of them**
  - One Arabic Letter Alef (U+0627)
    - Most Arabic letters take at least 4 shapes depending on context
  - But, two Latin Letter A (oops!)
    - LATIN CAPITAL LETTER A (U+0041) / LATIN SMALL LETTER A (U+0061)
- **Some punctuations are shared, some are not**
  - Single Period/Full Stop symbol for most scripts (“.” U+002E)
  - A pair of Question Marks (“?” U+003F, “؟” U+061F)

## Semantic encoding in Unicode

Store text in memory in the same order as is read/processed in mind



- **Encode concepts, not various shapes of them**
  - One Arabic Letter Alef (U+0627)
    - Most Arabic letters take at least 4 shapes depending on context
  - But, two Latin Letter A (oops!)
    - LATIN CAPITAL LETTER A (U+0041) / LATIN SMALL LETTER A (U+0061)
- **Some punctuations are shared, some are not**
  - Single Period/Full Stop symbol for most scripts (“.” U+002E)
  - A pair of Question Marks (“?” U+003F, “؟” U+061F)
- **Some Numerals are LTR and some RTL**
  - Until 2006 (encoding of N’Ko), all numerals were LTR
    - European (ASCII): 0123456789 / Eastern Hindi-Arabic (Persian): ٠١٢٣٤٥٦٧٨٩
  - Recently-developed African systems use RTL numerals
    - N’Ko: ߠߡߣߤߥߦߧߨߩ

---

## Direction in text block

LTR paragraphs  
are usually aligned  
“flush left”, a.k.a.  
“left-aligned” or  
“ragged right”.

---

# What will be the biggest internet trends between 2016-2020?

---

## Direction in text block

RTL paragraphs are usually aligned “flush right”, a.k.a. “right-aligned” or “ragged left”.

What will be the biggest internet trends between 2016-2020?

بزرگترین روندهای اینترنتی در بین سال‌های  
۲۰۱۶-۲۰۲۰ چه خواهد بود؟

**Direction in  
text block**

→ **What will be the biggest internet  
→ trends between 2016-2020?**

**Reading direction  
is usually  
perceived  
implicitly from the  
writing system...**

← **بزرگترین روندهای اینترنتی در بین سال‌های**  
← **۲۰۱۶-۲۰۲۰ چه خواهد بود؟**

**Direction in  
text block**

...allowing reading  
“end-aligned” text  
with no problems.



→ **What will be the biggest internet  
trends between 2016-2020?**

بزرگترین روندهای اینترنتی در بین سال‌های  
۲۰۱۶-۲۰۲۰ چه خواهد بود؟ ←

---

## Direction in text block

Setting the wrong direction results in poor readability, and sometimes event close to gibberish.



---

What will be the biggest internet  
?trends between 2016-2020

- بزرگترین روندهای اینترنتی در بین سال‌های
- چه خواهد بود؟ ۲۰۱۶-۲۰۲۰

**Direction in  
text block**

→ **What will be the biggest internet  
→ trends between 2016-2020?**

**Let's now look at  
how sequences of  
shapes are  
perceived.**

← **بزرگترین روندهای اینترنتی در بین سال‌های**  
← **۲۰۱۶-۲۰۲۰ چه خواهد بود؟**

**Direction in  
text block**

→ **What will be the biggest internet  
→ trends between 2016-2020?**

*LTR runs* ⇒ orange

*RTL runs* ⇒ green

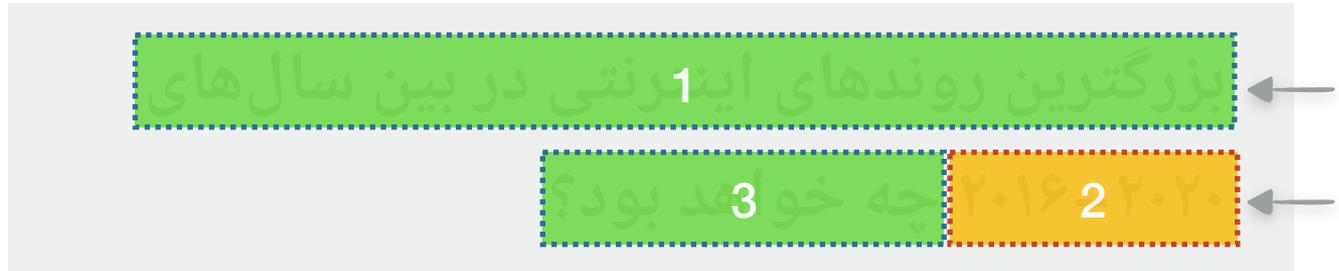
بزرگترین روندهای اینترنتی در بین سال‌های ←

← ۲۰۱۶-۲۰۲۰ چه خواهد بود؟

## Direction in text block



On the line level, the runs are read in order, in the direction of the paragraph (base direction)



---

# Unicode Bidirectional Algorithm (UBA)

Annex #9 to the  
Unicode Standard  
(UAX #9)

- Converting a *semantic in-memory* string of chars into a reordering suitable for presentation (visual output)

---

## Unicode Bidirectional Algorithm (UBA)

Annex #9 to the  
Unicode Standard  
(UAX #9)

- **Converting a *semantic in-memory* string of chars into a reordering suitable for presentation (visual output)**
- **Every Unicode Character has a Bidi Class**
  - Strong, such as letters
  - Weak, such as numbers
  - Neutral, such as whitespace, punctuation and symbols

# Unicode Bidirectional Algorithm (UBA)

Annex #9 to the  
Unicode Standard  
(UAX #9)

- **Converting a *semantic in-memory* string of chars into a reordering suitable for presentation (visual output)**
- **Every Unicode Character has a Bidi Class**
  - Strong, such as letters
  - Weak, such as numbers
  - Neutral, such as whitespace, punctuation and symbols
- **Some characters are *Mirrored* if in an RTL run**
  - Parenthesis are mirrored: “(” is an *open parens* in both LTR & RTL
  - Question Marks do not mirror: “?” is always closed on the right.

---

# Unicode Bidirectional Algorithm (UBA)

High-level steps of  
the algorithm

- ***Input: string of characters & base direction***
  - Both inputs should be set correctly to achieve the correct presentation

---

## Unicode Bidirectional Algorithm (UBA)

### High-level steps of the algorithm

- ***Input: string of characters & base direction***
  - Both inputs should be set correctly to achieve the correct presentation
- ***Output: chars' levels (evens are LTR, odds are RTL) & position***

# Unicode Bidirectional Algorithm (UBA)

## High-level steps of the algorithm

- **Input: string of characters & base direction**
  - Both inputs should be set correctly to achieve the correct presentation
- **Output: chars' levels (evens are LTR, odds are RTL) & position**
- **First, explicit direction levels are calculated**
  - Based on special directional formatting characters
    - Embedding (LRE, RLE), Isolate (LRI, RLI, FSI), Override (LRO, RLO)
  - Higher-level protocol
    - HTML (`dir="rtl"`)
    - CSS (`direction: rtl;`)

# Unicode Bidirectional Algorithm (UBA)

## High-level steps of the algorithm

- **Input: string of characters & base direction**
  - Both inputs should be set correctly to achieve the correct presentation
- **Output: chars' levels (evens are LTR, odds are RTL) & position**
- **First, explicit direction levels are calculated**
  - Based on special directional formatting characters
    - Embedding (LRE, RLE), Isolate (LRI, RLI, FSI), Override (LRO, RLO)
  - Higher-level protocol
    - HTML (`dir="rtl"`)
    - CSS (`direction: rtl;`)
- **Then, implicit dir. levels are calculated using chars' Bidi Class**
  - Implicit formatting characters (LRM, RLM, ALM) take effect here

# Unicode Bidirectional Algorithm (UBA)

## High-level steps of the algorithm

- **Input: string of characters & base direction**
  - Both inputs should be set correctly to achieve the correct presentation
- **Output: chars' levels (evens are LTR, odds are RTL) & position**
- **First, explicit direction levels are calculated**
  - Based on special directional formatting characters
    - Embedding (LRE, RLE), Isolate (LRI, RLI, FSI), Override (LRO, RLO)
  - Higher-level protocol
    - HTML (`dir="rtl"`)
    - CSS (`direction: rtl;`)
- **Then, implicit dir. levels are calculated using chars' Bidi Class**
  - Implicit formatting characters (LRM, RLM, ALM) take effect here
- **Finally, having the bidi levels, reordering can be done, when needed**

## Directional embeddings

How directions are mixed when sentences get more complicated?

- They translated the question
- into “بزرگترین روندهای اینترنتی در بین”
- “سالهای ۲۰۱۶-۲۰۲۰ چه خواهد بود؟” on
- Quora!

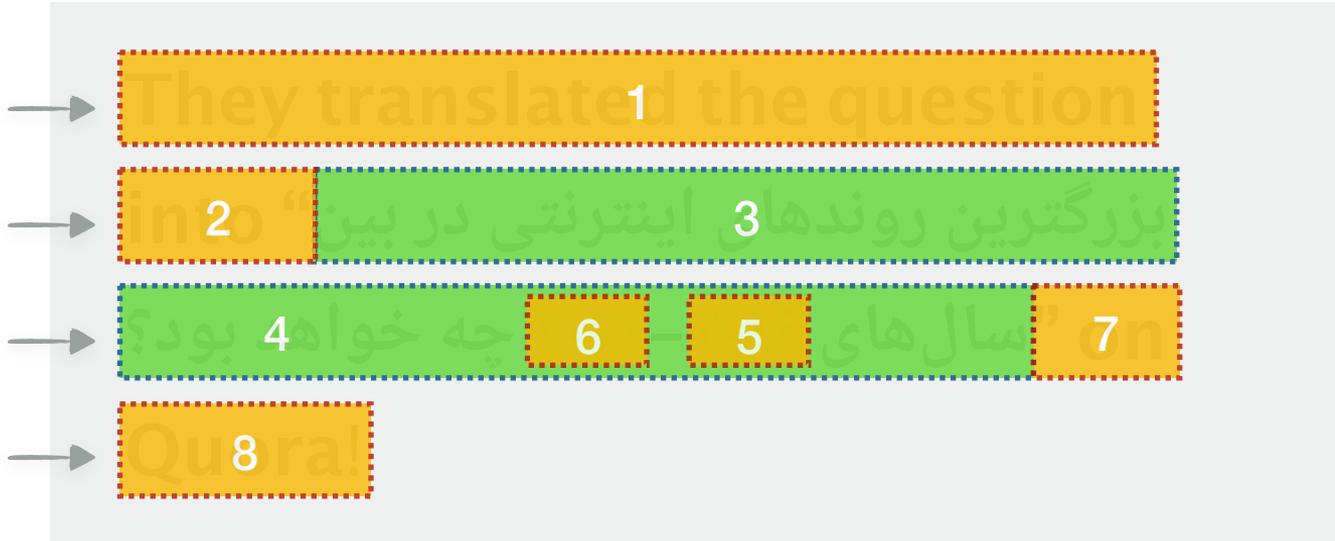
## Directional embeddings

We get opposite-direction runs embedded in runs, running opposite to the paragraph direction.

- They translated the question
- into “بزرگترین روندهای اینترنتی در بین”
- “سالهای ۲۰۱۶–۲۰۲۰ چه خواهد بود؟” on
- Quora!

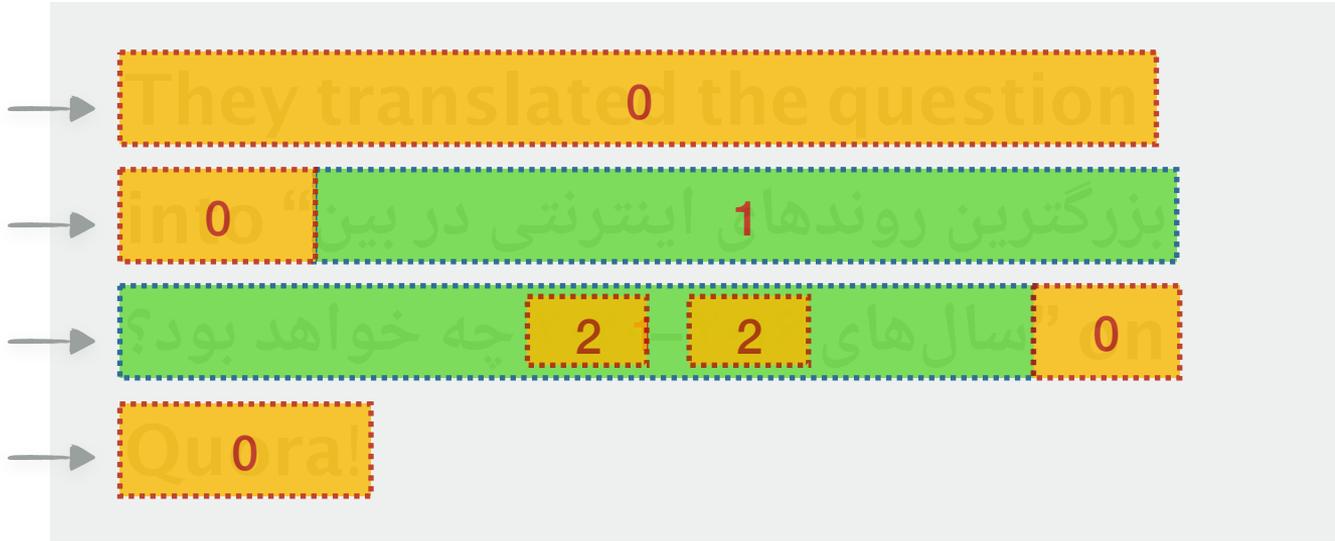
## Directional embeddings

In order, these will be...



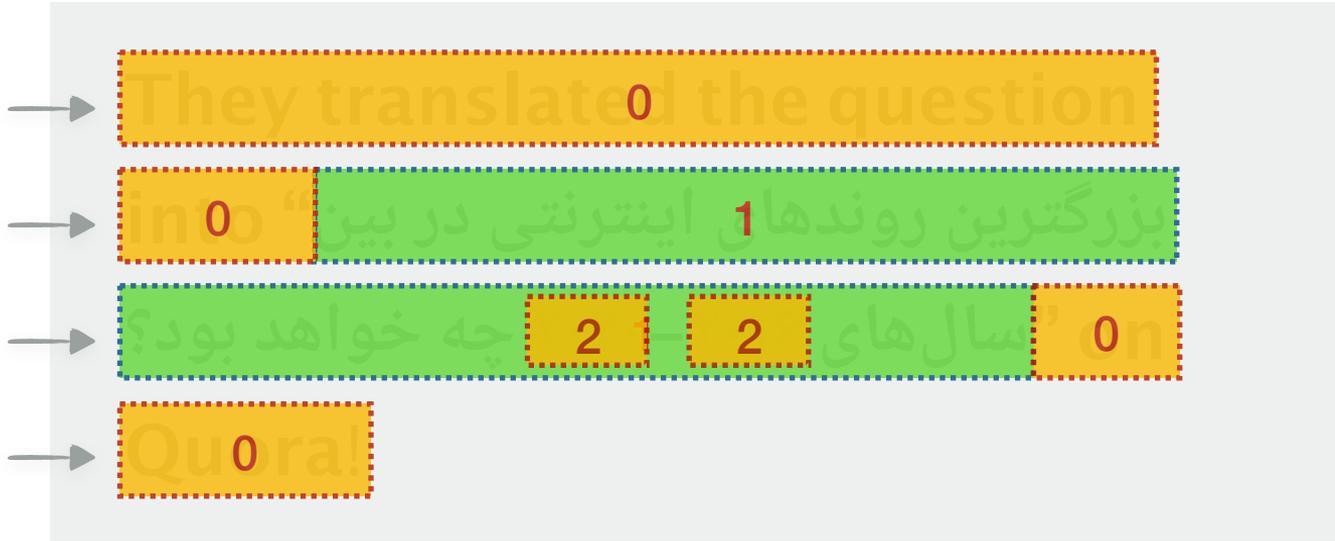
## Directional embeddings

In terms of UBA embedding levels, they would be...



## Directional embeddings

In terms of UBA embedding levels, they would be...



Can go up to 126 levels!

# **Bidirectional Layout**



## به ویکی‌پدیا خوش آمدید

دانشنامه‌ای آزاد که همه می‌توانند آن را ویرایش کنند؛  
با ۶۳۹٬۹۷۷ مقاله به زبان فارسی

آشنایی با دانشنامه • آشنایی با اصول ویرایش • کارهای قابل انجام • سیاست‌ها و رهنمودها • فهرست الفبایی مقاله‌ها • نسخهٔ تلفن همراه

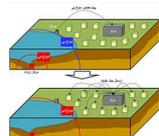
### از میان خبرها



ان.کی. جمیسین در سال ۲۰۱۵

- ان. کی. جمیسین (تصویر) برای سومین مرتبه متوالی برندهٔ جایزه ادبی هوگو برای بهترین رمان برای **آسمان سنگی** شد.
  - **عمران خان**، رسماً نخست‌وزیر جدید پاکستان شد.
  - **کوفی عنان**، دبیرکل پیشین سازمان ملل متحد در سن ۸۰ سالگی درگذشت.
  - طی **تهاجم طالبان** برای تسلط بر غزنی و درگیری با نیروهای مسلح افغانستان و آمریکا چند صد تن از طرفین کشته شدند.
  - در نشست سران کشورهای حاشیه دریای خزر در آق‌تاو، **کنوانسیون رژیم حقوقی خزر** به امضا رسید.
- رویدادهای کنونی: نبرد جدید • بازگشایی آسیایی**  
۳۰۱۸

### مقاله برگزیده



**سامانه هشدار زمین‌لرزه** سامانه‌ای برای اطلاع‌رسانی سریع زمین‌لرزه‌های بزرگ قریب‌الوقوع است. این فناوری هم‌اینک تنها فناوری دارای قابلیت پیش‌بینی زمین‌لرزه در لحظاتی پیش از وقوع آن است. افزایش جمعیت و تراکم صنایع در مناطق آسیب‌پذیر از چند دهه آخر قرن بیستم به بعد، نرخ مرگ‌ومیر و زیان‌های اقتصادی ناشی از بلایای طبیعی را افزایش می‌دهد و اهمیت سامانه‌های هشدار زمین‌لرزه، به عنوان مهم‌ترین ابزار کاهش شدت فاجعه را بیشتر نموده‌است. نحوه عملکرد انواع جدید سامانه به این صورت است: در هنگام وقوع زمین‌لرزه، از کانون آن چندین موج با سرعت‌های مختلف منتشر می‌شود (در هر زلزله‌ای، ۴ نوع اصلی از امواج ارتجاعی، قابل احساس بوده و قادر به ایجاد خسارت می‌باشند). امواج سریع‌تر (**امواج پی**) با سرعتی حدود ۷ کیلومتر بر ثانیه منتشر می‌شوند. این امواج به‌طور کلی آسیب کمی می‌رسانند. امواج مخرب (**موج اس**) با سرعتی حدود ۴ کیلومتر بر ثانیه منتشر می‌شوند. البته سرعت دقیق و واقعی امواج اولیه (P) و ثانویه (S)، وابسته به تراکم زمین و جنس سنگ‌های موجود در منطقه انتشار است. به عنوان مثال، در بستر گرانیتی، امواج P با سرعت ۵/۵ کیلومتر در ثانیه و امواج S با سرعت ۳ کیلومتر در ثانیه منتشر می‌شوند. در حالیکه درون آب (فرضاً اقیانوس‌ها)، سرعت انتشار امواج  $P = ۵/۱$  کیلومتر در ثانیه و سرعت انتشار  $S = ۳$  کیلومتر در ثانیه است. بنابراین (در اغلب زمین‌لرزه‌ها) موج پی زودتر از موج اس قابل حس

# Web-based layout

Top to bottom, right to left



۱ صفحه اصلی بحث

۲ خواندن نمایش مبدأ نمایش تاریخچه جستجو در ویکی‌پدیا

ویکی‌پدیا  
دانشنامه آزاد

## برویکی‌پدیا خوش آمدید

دانشنامه‌ای آزاد که همه می‌توانند آن را ویرایش کنند؛  
با ۶۳۹٬۹۷۷ مقاله به زبان فارسی

آشنایی با دانشنامه • آشنایی با اصول ویرایش • کارهای قابل انجام • سیاست‌ها و رهنمودها • فهرست القاب‌های مقاله‌ها • نسخه تلفن همراه

از میان خبرها

ان. کی، حمیسین (تصویر) برای سومین مرتبه متوالی برنده جایزه ادبی هوگو برای بهترین رمان برای **آسمان سنگی** شد.

• **عمران خان**، رسماً نخست‌وزیر جدید پاکستان شد.

• **کوفی عنان**، دبیرکل پیشین سازمان ملل متحد در سن ۸۰ سالگی درگذشت.

• طی **تهاجم طالبان** برای تسلط بر غزنی و درگیری با نیروهای مسلح افغانستان و آمریکا چند صد تن از طرفین کشته شدند.

• در نشست سران کشورهای حاشیه دریای خزر در آق‌تاو، **کنوانسیون رژیم حقوقی خزر** به امضا رسید.

• **رویدادهای کنونی: نبرد جدید • بازگهای آسیایی**

مقاله برگزیده

**سامانه هشدار زمین‌لرزه سامانه‌ای برای اطلاع‌رسانی سریع زمین‌لرزه‌های بزرگ قریب‌الوقوع است. این فناوری هم‌اینک تنها فناوری دارای قابلیت پیش‌بینی زمین‌لرزه در لحظاتی پیش از وقوع آن است. افزایش جمعیت و تراکم صنایع در مناطق آسیب‌پذیر از چند دهه آخر قرن بیستم به بعد، نرخ مرگ‌ومیر و زیان‌های اقتصادی ناشی از بلایای طبیعی را افزایش می‌دهد و اهمیت سامانه‌های هشدار زمین‌لرزه، به عنوان مهمترین ابزار کاهش شدت فاجعه را بیشتر نموده است. نحوه عملکرد انواع جدید سامانه به این صورت است: در هنگام وقوع زمین‌لرزه، از کانون آن چندین موج با سرعت‌های مختلف منتشر می‌شود (در هر زلزله‌ای، ۴ نوع اصلی از امواج ارتجاعی، قابل احساس بوده و قادر به ایجاد خسارت می‌باشند). امواج سریع‌تر (امواج پی) با سرعتی حدود ۷ کیلومتر بر ثانیه منتشر می‌شوند. این امواج به‌طور کلی آسیب کمی می‌رسانند. امواج مخرب (موج اس) با سرعتی حدود ۴ کیلومتر بر ثانیه منتشر می‌شوند. البته سرعت دقیق و واقعی امواج اولیه (P) و ثانویه (S)، وابسته به تراکم زمین و جنس سنگ‌های موجود در منطقه انتشار است. به عنوان مثال، در بستر گرانیتی، امواج P با سرعت ۵/۵ کیلومتر در ثانیه و امواج S با سرعت ۳ کیلومتر در ثانیه منتشر می‌شوند. در حالیکه درون آب (فرضاً اقیانوس‌ها)، سرعت انتشار امواج P = ۵/۱ کیلومتر در ثانیه و سرعت انتشار S = صفر است. بنابراین (در اغلب زمین‌لرزه‌ها) موج پی زودتر از موج اس قابل حس**

39

# Web-based layout

Every block has a direction



🔔 به سامانه وارد نشده‌اید. بحث مشارکت‌ها ایجاد حساب کاربری ورود

صفحه اصلی بحث

خواندن نمایش مبدأ نمایش تاریخچه جستجو در ویکی‌پدیا

2

## ویکی‌پدیا

دانشنامه آزاد

# برویکی‌پدیا خوش آمدید

دانشنامه‌ای آزاد که همه می‌توانند آن را ویرایش کنند؛  
با ۶۳۹٬۹۷۷ مقاله به زبان فارسی

آشنایی با دانشنامه • آشنایی با اصول ویرایش • کارهای قابل انجام • سیاست‌ها و رهنمودها • فهرست القاب‌های مقاله‌ها • نسخهٔ تلفن همراه

3

صفحه اصلی  
رویدادهای کنونی  
مقالهٔ تصادفی  
کمک مالی

همکاری  
تغییرات اخیر  
ویکی‌نویس شوید!  
راهنما  
نمایش با ویکی‌پدیا

1

نسخه برداری  
ایجاد کتاب  
دریافت به صورت PDF  
نسخه قابل چاپ

🌍 **از میان خبرها**

- ان. کی. جمیسین (تصویر) برای سومین مرتبه متوالی برندهٔ جایزه ادبی هوگو برای بهترین رمان برای **آسمان سنگی** شد.
- **عمران خان**، رسماً نخست‌وزیر جدید پاکستان شد.
- **کوفی عنان**، دبیرکل پیشین سازمان ملل متحد در سن ۸۰ سالگی درگذشت.
- طی **تهاجم طالبان** برای تسلط بر غزنی و درگیری با نیروهای مسلح افغانستان و آمریکا چند صد تن از طرفین کشته شدند.
- در نشست سران کشورهای حاشیه دریای خزر در آق‌ناو، کنوانسیون رژیم حقوقی خزر به امضا رسید.
- رویدادهای کنونی: نبرد جدید • بازی‌های آسیایی

📖 **موضوعات خالجه**

ویکی‌سفر راهنمای سفر به ژاپن

ویکی‌سفر راهنمای سفر به چین

ویکی‌سفر راهنمای سفر به هند

ویکی‌سفر راهنمای سفر به آمریکا

ویکی‌سفر راهنمای سفر به اروپا

ویکی‌سفر راهنمای سفر به آسیا

ویکی‌سفر راهنمای سفر به اقیانوس آرام

ویکی‌سفر راهنمای سفر به آفریقا

ویکی‌سفر راهنمای سفر به اقیانوس منجمد جنوبی

ویکی‌سفر راهنمای سفر به اقیانوس منجمد شمالی

ویکی‌سفر راهنمای سفر به اقیانوس منجمد جنوبی

ویکی‌سفر راهنمای سفر به اقیانوس منجمد شمالی

40

---

## Direction in layout blocks

Here, we limit the discussion to horizontal *writing mode* with upright *line orientation* and downward *block flow direction*.

Q



---

## Direction in layout blocks

Here, we limit the discussion to horizontal *writing mode* with upright *line orientation* and downward *block flow direction*.



- 
- Converting an LTR layout to an RTL one is called *Mirroring*



## Direction in layout blocks

Here, we limit the discussion to horizontal *writing mode* with upright *line orientation* and downward *block flow direction*.



- Converting an LTR layout to an RTL one is called *Mirroring*
- *Flow of movement is reversed in mirroring*
  - Start/previous/past is on the righthand-side (RHS)
  - End/next/future is on the lefthand-side (LHS)



## Direction in layout blocks

Here, we limit the discussion to horizontal *writing mode* with upright *line orientation* and downward *block flow direction*.



- Converting an LTR layout to an RTL one is called *Mirroring*
  - *Flow of movement is reversed in mirroring*
    - Start/previous/past is on the righthand-side (RHS)
    - End/next/future is on the lefthand-side (LHS)
- 
- **Layout direction works very similar to text direction**
    - Blocks are set from *start* to *end*, depending on the contextual dir.
    - Table columns are also ordered from *start* to *end*
    - Any sequence, such as images, is also ordered from *start* to *end*

## Direction in layout blocks

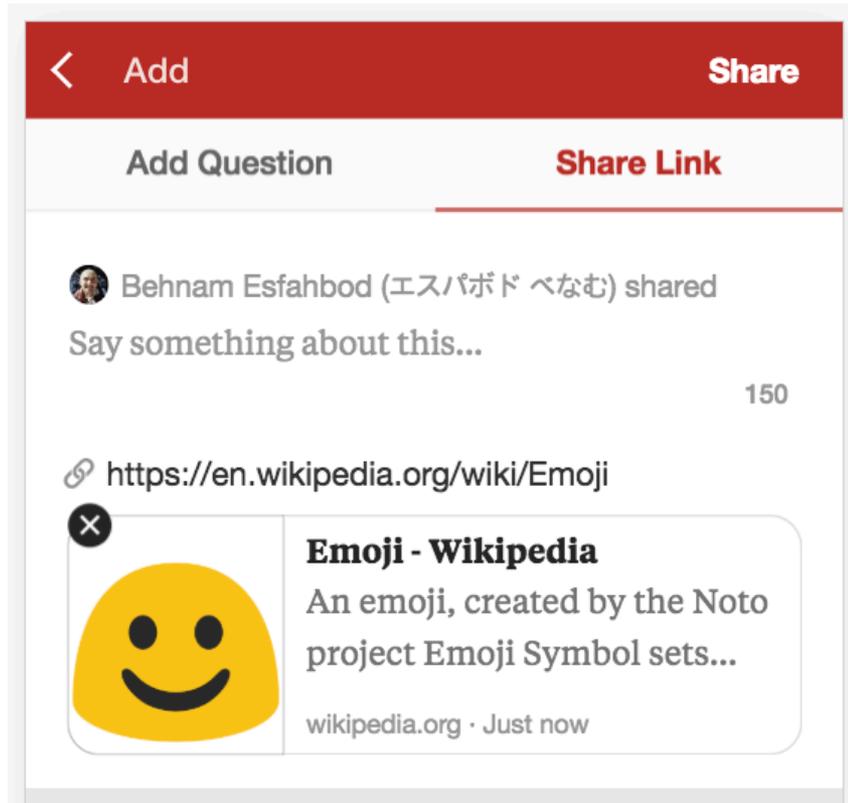
Here, we limit the discussion to horizontal *writing mode* with upright *line orientation* and downward *block flow direction*.



- **Converting an LTR layout to an RTL one is called *Mirroring***
  - ***Flow of movement is reversed in mirroring***
    - Start/previous/past is on the righthand-side (RHS)
    - End/next/future is on the lefthand-side (LHS)
- 
- **Layout direction works very similar to text direction**
    - Blocks are set from *start* to *end*, depending on the contextual dir.
    - Table columns are also ordered from *start* to *end*
    - Any sequence, such as images, is also ordered from *start* to *end*
  - **There are a few exceptions, though!**
    - Modern mathematics notation (usually) stays LTR
    - Some well-known interfaces, like audio/video back/play/forward set

## Mixed directions

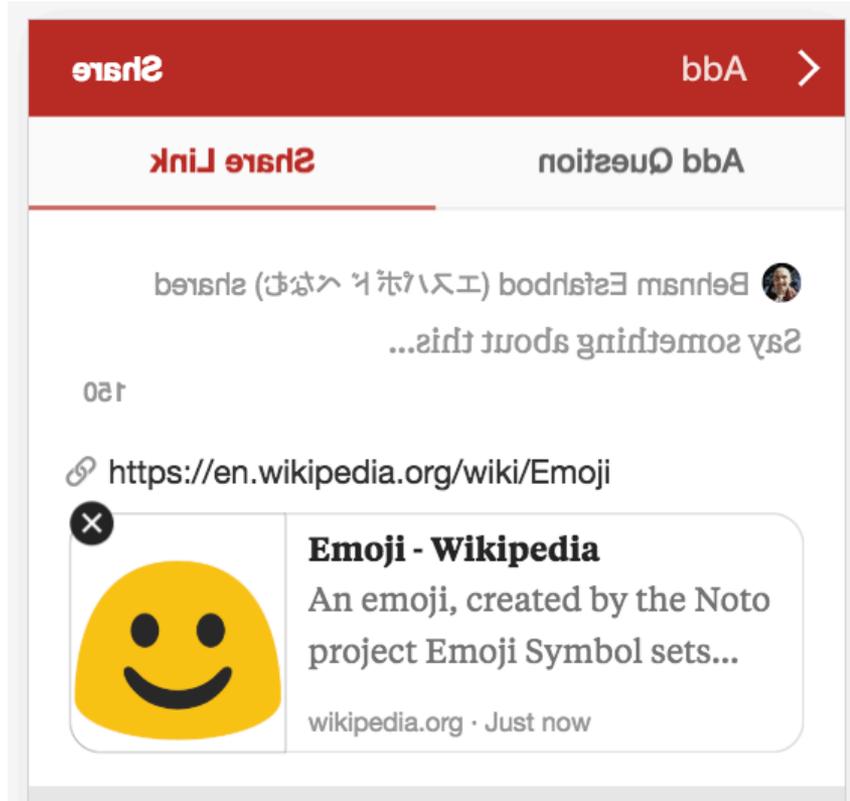
Let's look at a basic example...



## Mixed directions

Most elements mirror...

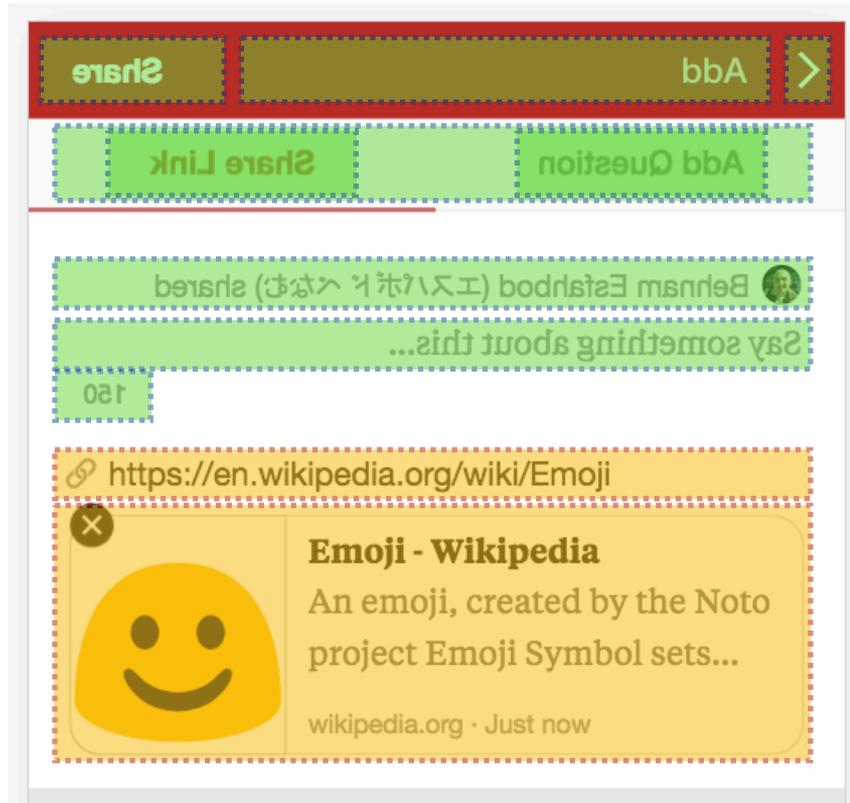
Some, don't.



## Mixed directions

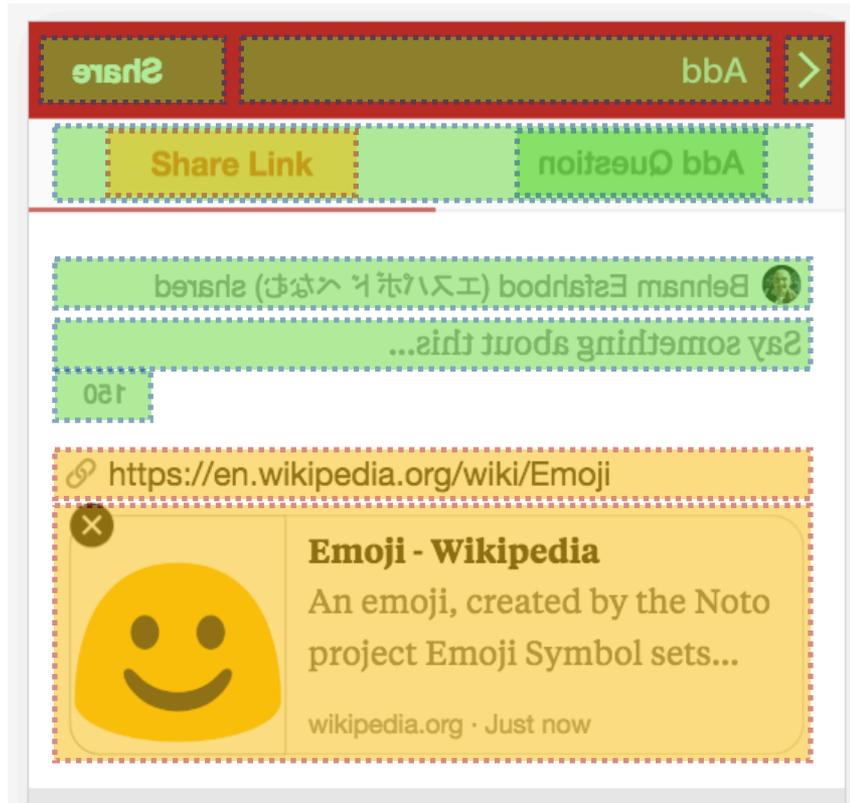
Many levels of implicit or explicit directionality

In a sample RTL  
Top-level direction...



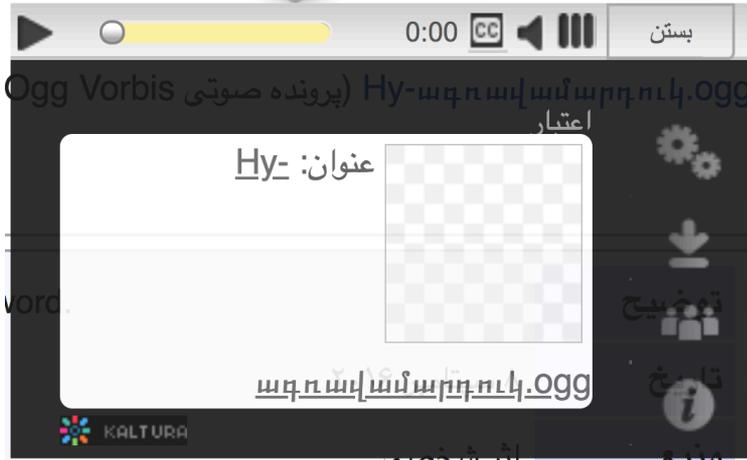
## Mixed directions

What if an interface message is not translated?



# Static directionality

Mostly concepts  
with static  
behavior IRL



# **Bidirectional Web Application**

## Text input

Can't make assumption about the script of every character of user-generated content.



H1 H2 H3 H4 H5 H6 Blockquote UL OL Code Block  
Bold Italic Underline Monospace

2008 – CERN's **Large Hadron Collider** (*section pictured*), the world's largest and highest-energy particle accelerator, was first powered up beneath the Franco-Swiss border near Geneva.

برخورددهنده هادرونی بزرگ در ژنوسوئیس آغاز به کار کرد - ۲۰۰۸

## Text input

Heuristic methods often result in unexpected behavior.

H1 H2 H3 H4 H5 H6 Blockquote UL OL Code Block  
Bold Italic Underline Monospace

CERN's **Large Hadron Collider** (*section pictured*), the – 2008 world's largest and highest-energy particle accelerator, was first .powered up beneath the Franco-Swiss border near Geneva

۲۰۰۸ - برخورددهنده هادرونی بزرگ در ژنوسوئیس آغاز به کار کرد.

## Text input

Giving control of every text block to the user has the least friction.

H1 H2 H3 H4 H5 H6 Blockquote UL OL Code Block  
Bold Italic Underline Monospace

2008 – CERN's **Large Hadron Collider** (*section pictured*), the world's largest and highest-energy particle accelerator, was first powered up beneath the Franco-Swiss border near Geneva.

۲۰۰۸ - برخورددهنده هادرونی بزرگ در ژنوسوئیس آغاز به کار کرد.

---

## Text processing

- The top advantage of *semantic* encoding of RTL/bidi text is the ease of processing

## Text processing

- The top advantage of *semantic* encoding of RTL/bidi text is the ease of processing
- Most Unicode characters represent a *linguistic* element
  - Although encoding of Arabic script has extra complexities

## Text processing

- **The top advantage of *semantic* encoding of RTL/bidi text is the ease of processing**
- **Most Unicode characters represent a *linguistic* element**
  - Although encoding of Arabic script has extra complexities
- **Finding the first letter, splitting into words, truncating a paragraph, all work very similar to LTR scripts**

## Text output

- **Most apps depend on the system/platform to render a bidi text**
  - Get good results *iff* play well with the text and layout algorithms

Plaintext



## Text output

### Plaintext

- **Most apps depend on the system/platform to render a bidi text**
  - Get good results *iff* play well with the text and layout algorithms
- **For plaintext, use Unicode bidi formatting chars**
  - Implicit: Marks (LRM, RLM, ALM)
    - Useful when the problem is local and asymmetric
    - e.g. positioning of a single symbol is not correct in an isolated box
  - Explicit: Embedding (LRE, RLE) & Isolate (LRI, RLI)
    - Embedding is the old method, Isolate is more recent
    - Useful at the boundaries of languages/scripts, also data and its surrounding sentence.

## Text output

### Plaintext

- **Most apps depend on the system/platform to render a bidi text**
  - Get good results *iff* play well with the text and layout algorithms
- **For plaintext, use Unicode bidi formatting chars**
  - Implicit: Marks (LRM, RLM, ALM)
    - Useful when the problem is local and asymmetric
    - e.g. positioning of a single symbol is not correct in an isolated box
  - Explicit: Embedding (LRE, RLE) & Isolate (LRI, RLI)
    - Embedding is the old method, Isolate is more recent
    - Useful at the boundaries of languages/scripts, also data and its surrounding sentence.
  - Explicit: Overrides (LRO, RLO)
    - For legacy systems
    - There's almost **no** good reason to use these in modern systems

## Text output

- **Use formatting Marks for implicit matters**
  - As encoded characters, or
  - As entities, `&lrm;` and `&#8206;`

HTML



## Text output

### HTML

- **Use formatting Marks for implicit matters**
  - As encoded characters, or
  - As entities, `&lrm;` and `&#8206;`
- **For blocks and explicit directions**
  - Use proper attributes
    - HTML (`dir="rtl"`)
    - CSS (`direction: rtl;`)
  - Leverage the default *inheritance* of these properties from parent nodes to children
    - Set `dir` attribute on the `<html>` or `<body>` tags

## Text output

### HTML

- **Use formatting Marks for implicit matters**
  - As encoded characters, or
  - As entities, `&lrm;` and `&#8206;`
- **For blocks and explicit directions**
  - Use proper attributes
    - HTML (`dir="rtl"`)
    - CSS (`direction: rtl;`)
  - Leverage the default *inheritance* of these properties from parent nodes to children
    - Set `dir` attribute on the `<html>` or `<body>` tags
- **Use CSS flipping tools to make a RTL version of LTR rules**

## Text output

### HTML

- **Use formatting Marks for implicit matters**
  - As encoded characters, or
  - As entities, `&lrm;` and `&#8206;`
- **For blocks and explicit directions**
  - Use proper attributes
    - HTML (`dir="rtl"`)
    - CSS (`direction: rtl;`)
  - Leverage the default *inheritance* of these properties from parent nodes to children
    - Set `dir` attribute on the `<html>` or `<body>` tags
- **Use CSS flipping tools to make a RTL version of LTR rules**
  - As of 2018, you still cannot do that natively in CSS!



# Interface

## Non-textual elements



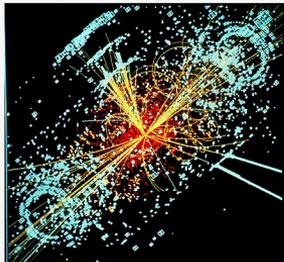
به سامانه وارد نشده‌اید. بحث مشارکت‌ها ایجاد حساب کاربری ورود

مقاله بحث خواندن ویرایش نمایش تاریخچه جستجو در ویکی‌پدیا

انتشار تغییرات... درج یادکرد

مختصات: ۳۶°۱۳′ شمالی ۰۶°۰۳′ شرقی﻿ / ﻿۳۶٫۲۱۷°شمالی ۰۶٫۰۵۰°شرقی﻿ / ۳۶.۲۱۷; ۶.۰۵۰

### فرا تر از مدل استاندارد



شبه‌سازی داده‌های آشکارساز ذرات برخورددهنده هادرونی بزرگ که یک بوزون هیگز که از برخورد پروتون‌هایی به‌وجود آمده است که به فواره‌های هادرونی و الکترونی واپاشی می‌شود.

#### مدل استاندارد

شواهد

- مسئله سلسله‌مراتب ماده تاریک
- مسئله ثابت کیهانی
- مسئله سی‌پی قوی
- نوسان نوترینو

نظریه‌ها

تکنی کالر



ویکی‌پدیا  
دانشنامهٔ آزاد

- صفحهٔ اصلی
- رویدادهای کنونی
- مقالهٔ تصادفی
- کمک مالی

همکاری

- تغییرات اخیر
- ویکی‌نویس شوید!
- راهنما
- تماس با ویکی‌پدیا

در دیگر پروژه‌ها

ویکی‌انبار

ابزارها

- پیوندها به این صفحه
- تغییرات مرتبط
- بارگذاری پرونده
- صفحه‌های ویژه
- اطلاعات صفحه
- آیتم ویکی‌داده



به زبان‌های دیگر

- Deutsch
- English
- Español
- Français
- Italiano

## برخورددهنده هادرونی بزرگ

از ویکی‌پدیا، دانشنامهٔ آزاد

**برخورددهندهٔ هادرونی بزرگ** (به انگلیسی: Large Hadron Collider) یا به طور مختصر **ال‌اچ‌سی** (**LHC**) یک شتاب‌دهندهٔ ذره‌ای و برخورددهنده مستقر در سازمان تحقیقاتی *سرن* در نزدیکی ژنو سوئیس است. این پروژه در ۱۰ سپتامبر ۲۰۰۸ میلادی (۲۰ شهریور ۱۳۸۷ هجری شمسی) پس از ۲۰ سال آماده‌سازی، آغاز به کار کرد. هدف از ساختن آن شناخت اجرام ماده در حد فاصل <sup>23</sup>–10 سانتی‌متر، آزمون مدل استاندارد ذرات، کشف اجزای یافت نشدهٔ مدل استاندارد، آزمون نظریهٔ ابرتقارن و نظریه وحدت بزرگ است. از دیگر اهداف مهم این پروژه، کشف ذرهٔ بنیادی **هیگز** است که فیزیک‌دانان ذرات بنیادی وجود آن را پیش‌گویی کرده‌اند. ذرهٔ هیگز یا بوزون هیگز، دخیل در ایجاد جرم در ذرات بنیادی است.

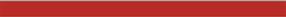
در این آزمایشگاه، **پروتونها**، در یک تونل ۲۷ کیلومتری شتاب گرفته و به اندازهٔ ۱۲ تریلیون الکترون ولت انرژی می‌گیرند و با هم برخورد می‌کنند تا این برخورد، ردی از **بوزون هیگز** را نشان‌دهد.<sup>[۱]</sup> این شتاب‌دهنده در تاریخ ۱۰ سپتامبر ۲۰۰۸ راه‌اندازی شد، ولی نه روز بعد به علت نقص فنی و بالا رفتن دمای آهن‌ریا‌های ابر رسانا که باید در دماهای پایین کار کنند متوقف شد. این شتاب‌دهنده بعد از ۱۴ ماه وقفه در تاریخ ۲۱ نوامبر ۲۰۰۹ مجدداً راه‌اندازی شد.<sup>[۲]</sup>

مرکز کنترل سرن هرگونه کنترل و سازماندهی اساسی را بر روی این شتاب‌دهنده انجام می‌دهد.<sup>[۳]</sup> در برخورد دهندهٔ بزرگ هادرونی تونل‌ها طوری طراحی و برنامه‌ریزی شده‌اند که در سراسر مسیر حلقه‌ای شکل، چهار برخورد برای پروتون‌ها صورت می‌گیرد، که این نقاط محلّ فرارگیری آزمایش‌ها هستند.<sup>[۳]</sup>

### آزمایش‌ها

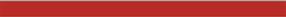


# **Bidirectionality Techniques**



## **Directionality context**

- 
- **Direction of text runs/blocks & layout blocks is a contextual property**



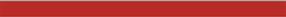
## **Directionality context**

- 
- **Direction of text runs/blocks & layout blocks is a contextual property**
  - **Techniques for managing *directionality context***
    1. Embedding
    2. Inheritance
    3. Cascading
    4. Propagation

## Directionality context

- **Direction of text runs/blocks & layout blocks is a contextual property**
- **Techniques for managing *directionality context***
  1. Embedding
  2. Inheritance
  3. Cascading
  4. Propagation
- **Abstractions to provide/absorb *directionality context***
  - Interface translation
    - Text processing
  - Interface components
    - HTML/platform elements and custom abstractions





## Embedding technique

Inline runs  
(intra-block)

- 
- **If not clear about directional, set isolation boundaries**
    - Skip isolation for same-direction embeddings, if known

## Embedding technique

### Inline runs (intra-block)

- **If not clear about directional, set isolation boundaries**
  - Skip isolation for same-direction embeddings, if known
- **Single block (start-to-end)**
  - One base direction per block
  - Limited to 126 levels (usually)

## Embedding technique

### Inline runs (intra-block)

- **If not clear about directional, set isolation boundaries**
  - Skip isolation for same-direction embeddings, if known
- **Single block (start-to-end)**
  - One base direction per block
  - Limited to 126 levels (usually)
- **Examples**
  - Plaintext embedding using Bidi Control Characters
  - HTML embedding using inline markups

---

## **Inheritance technique**

### **Block level**

- **Inherit the direction of parent block**
  - Unless there's more evidence
    - Static directionality
    - Propagation (Technique #4)

# Inheritance technique

## Block level

- **Inherit the direction of parent block**
  - Unless there's more evidence
    - Static directionality
    - Propagation (Technique #4)
- **Top-down**
  - One single top-level direction
  - Unlimited

# Inheritance technique

## Block level

- **Inherit the direction of parent block**
  - Unless there's more evidence
    - Static directionality
    - Propagation (Technique #4)
- **Top-down**
  - One single top-level direction
  - Unlimited
- **Examples**
  - Default behavior in HTML and most native interface stacks

# Cascading technique

- **If no strong direction, fallback on the previous block's**
  - Continue fallback until there's a strong direction
  - First block falls back onto the parent block (inheritance)

Block level



---

## Cascading technique

### Block level

- **If no strong direction, fallback on the previous block's**
  - Continue fallback until there's a strong direction
  - First block falls back onto the parent block (inheritance)
- **Same layer**
  - Unlimited

# Cascading technique

## Block level

- **If no strong direction, fallback on the previous block's**
  - Continue fallback until there's a strong direction
  - First block falls back onto the parent block (inheritance)
- **Same layer**
  - Unlimited
- **Examples**
  - Paragraph direction setting
    - GNOME Text Editor
    - Draft.js

# Cascading technique

Example from Draft.js (React WYSIWYG text editor)



H1 H2 H3 H4 H5 H6 Blockquote UL OL Code Block  
Bold Italic Underline Monospace

## Article 1

All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood.

...

2, 3, 4, ..., 30.

## مادهٔ اول

تمام افراد بشر آزاد بدنیا می‌آیند و از لحاظ حیثیت و حقوق با هم برابرند. همه دارای عقل و وجدان می‌باشند و باید نسبت بیکدیگر با روح برادری رفتار کنند.

...

۲، ۳، ۴، ...، ۳۰.

|

---

## Propagation technique

Block level  
& inline level



- 
- **Direction of an element depend on a child element**
    - In inline, the (outer) element is *perceived* as an *inline block*.

## Propagation technique

Block level  
& inline level

- **Direction of an element depend on a child element**
  - In inline, the (outer) element is *perceived* as an *inline block*.
- **Bottom-up**
  - Usually limited to within a *component* boundary

## Propagation technique

Block level  
& inline level

- **Direction of an element depend on a child element**
  - In inline, the (outer) element is *perceived* as an *inline block*.
- **Bottom-up**
  - Usually limited to within a *component* boundary
- **Examples**
  - Hashtags (inline)

Welcome to the i18n Conference! #unicode #یونی‌کد

به کنفرانس بین‌المللی‌سازی خوش آمدید! #یونی‌کد #unicode

## Propagation technique

Block level  
& inline level

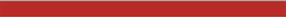
- **Direction of an element depend on a child element**
  - In inline, the (outer) element is *perceived* as an *inline block*.
- **Bottom-up**
  - Usually limited to within a *component* boundary
- **Examples**
  - Hashtags (inline)

Welcome to the i18n Conference! #unicode **#یونی‌کد**

**#یونی‌کد** به کنفرانس بین‌المللی‌سازی خوش آمدید! #unicode

- Link attachment preview (block)





## **Other challenges**

- 
- **Can't expect everyone to know UBA details by heart**



## **Other challenges**

- 
- **Can't expect everyone to know UBA details by heart**
  - **Some systems/platforms lack some bidi features**

## Other challenges

- **Can't expect everyone to know UBA details by heart**
- **Some systems/platforms lack some bidi features**
- **Some systems/platforms behave differently in corner cases**
  - e.g. UI components for Apple & Android

## Other challenges

- **Can't expect everyone to know UBA details by heart**
- **Some systems/platforms lack some bidi features**
- **Some systems/platforms behave differently in corner cases**
  - e.g. UI components for Apple & Android
- **Mixing data with interface messages is always a challenge**
  - Strict abstraction is needed to make sure every data, such as phone numbers, are always presented in the right order.

## Other challenges

- **Can't expect everyone to know UBA details by heart**
- **Some systems/platforms lack some bidi features**
- **Some systems/platforms behave differently in corner cases**
  - e.g. UI components for Apple & Android
- **Mixing data with interface messages is always a challenge**
  - Strict abstraction is needed to make sure every data, such as phone numbers, are always presented in the right order.
- **Unresolved culturally questions in bidi behavior**

## Summary

- **How writing systems got directionality**
- **How bidi text works in written form, and is encoded & represented**
- **How text and layout structures work in different directionalities**
- **Special application behaviors to support bidi locales &/or content**
- **Additional problems that require better system & i18n architecture**

---

## **Additional Reads**

### **Unicode Consortium**

- **[Unicode® Standard Annex #9—Unicode Bidirectional Algorithm \(UBA\)](#)**

### **W3C WG Notes and Articles**

- **[Text Layout Requirements for the Arabic Script](#)**
- **[Authoring HTML: Handling Right-to-left Scripts](#)**
- **[Additional Requirements for Bidi in HTML & CSS](#)**
- **[Unicode Bidirectional Algorithm basics](#)**
- **[Strings and bidi](#)**

### **Libraries**

- **[Draft.js](#)**



질문 ?

質問 ?

שאלות?

سؤال؟

پرسش؟

प्रश्न?

Questions?

Quora