

## **Text Layout Requirements of Urdu language in the W3C Standard “Arabic and Persian Gap Analysis”**

### **Introduction**

The document “Arabic and Persian Gap analysis” (<https://w3c.github.io/alreq/gap-analysis/>) describes and prioritizes gaps for the support of Arabic and Persian languages on the Web and in eBooks. In particular, it is concerned with text layout. It checks that needed features are supported in W3C specifications, in particular HTML and CSS and those relating to digital publications. It also checks whether the features have been implemented in browsers and eReaders. This document complements the [Text Layout Requirements for the Arabic Script](#), which describes the requirements for areas where gaps appear. It is linked to from the [language matrix](#) that tracks Web support for many languages. The W3C needs to make sure that the needs of scripts and languages around the world are built in to technologies such as HTML, CSS etc. so that Web pages and eBooks can look and behave as people expect around the world.

### **Problem Statement**

W3C has developed the Arabic requirements for text layout so that there will be proper and consistent text segmentation on the web. For this purpose, the current document covers the different area that requires feedback on the Urdu text segmentation so that this will be incorporated in the Standard for proper display and further implemented in the technologies such as HTML, CSS etc.

### **Recommendations**

**Text layout and presentation for languages that use the Perso-Arabic script when used by web standards and technologies such as HTML, CSS, Mobile Web, Digital Publications, and Unicode, developers can follow the Text Layout Requirements for the Perso-Arabic Script. Some following recommendations to consider include:**

**1. Using appropriate markup:**

Developers can use appropriate markup to indicate the directionality of the text, such as the HTML dir attribute, and to specify the language of the text using the lang attribute. This will ensure that the text is rendered correctly and that screen readers and other assistive technologies can properly interpret the content.

**2. Choosing appropriate fonts:**

Developers can choose appropriate fonts that support the Perso-Arabic script and ensure that they are available on their websites or applications. They can also use font fallback mechanisms to ensure that appropriate fonts are used for different operating systems and browsers.

Nastaleeq is a traditional font style that has been used for centuries in the Urdu language. There are several versions of this font style available for commercial use, both as free and open source options. Notable examples include Jameel Noori Nastaleeq, Mubeen Noori

Nastaleeq, Alvi Nastaleeq, Meher Nastaleeq, and Noto Nastaliq Urdu. Of these, Jameel Noori Nastaleeq has been widely used in print media for the past few decades, while Noto Nastaliq Urdu, developed by Google, has gained popularity as a web font. The standard Unicode encoding for Urdu text is UTF-8, which allows for the representation of all the characters and diacritical marks used in the Urdu script.

Urdu language has several commonly used fonts, such as Nastaleeq, Naskh, and Jameel Noori Nastaleeq, each with its unique style and design.

### 3. Implementing appropriate letter spacing

Letter spacing is recommended at the glyph level rather than the character level. If required, Font-kerning can be used at the character level.

اردو، جنوبی ایشیا کی ایک اہم اور بڑی زبان ہے۔ برصغیر ہند کی آزادی کے بعد سے اس زبان کی مقبولیت میں اضافہ ہی ہوتا جا رہا ہے۔ یہ ہندوستان کی اٹھارہ (18) قومی زبانوں میں سے ایک ہے۔ یہ پاکستان کی سرکاری زبان ہے۔ حالانکہ اس زبان پر عربی و فارسی کے اثرات ہیں لیکن عربی و فارسی کے برعکس یہ ہندی کی طرح ایک ہند آریائی زبان ہے جو برصغیر ہند میں پیدا ہوئی اور یہیں اس کی ترقی ہو

### 4. Implementing Nastaleeq font by Default:

Type face names/ class i.e 'Nastaleeq', 'Naskh' need to be introduced similar to 'serif', sans serif that is currently in used. 'Nastaleeq' should be used by default.

font-family: "Noto Nastaliq Urdu", "Jameel Noori Nastaleeq", "Nafees Nastaleeq", "Alvi Nastaleeq", "Urdu Nastaliq Unicode", "PDMS Nastaliq Nafees" "Pak Nastaleeq", "PDMS Saleem QuranFont", "PDMS Jauhar Urdu", "Fajer Noori Nastalique", "Fajer Quranic Nastaleeq", "Fajer Noori Nastaleeq Web", "Nafees Tahreer Nastaleeq", "Mubeen Noori Nastaleeq", "Nafees Pakistani Nastaleeq", "Nafees Web Naskh", "Segoe UI", sans-serif;

```
<!DOCTYPE html>
<html lang="ur">
<head>
  <meta charset="UTF-8">
  <title>عنوان</title>
  <style>
    @font-face {
      font-family: 'Jameel Noori Nastaleeq';
      src: url('path/to/font/file.ttf');
    }
    body {
      font-family: 'Jameel Noori Nastaleeq', serif;
      direction: rtl;
    }
  </style>
</head>
<body>
  <h1>عنوان</h1>
  اردو، جنوبی ایشیا کی ایک اہم اور بڑی زبان ہے۔ برصغیر ہند کی آزادی کے بعد سے اس زبان کی مقبولیت میں اضافہ ہی ہوتا جا رہا ہے۔ یہ ہندوستان کی اٹھارہ (18) قومی زبانوں میں سے ایک ہے۔ یہ پاکستان کی سرکاری زبان ہے۔ حالانکہ اس زبان پر عربی و فارسی کے اثرات ہیں لیکن عربی و فارسی کے برعکس یہ ہندی کی طرح ایک ہند آریائی زبان ہے جو برصغیر ہند میں پیدا ہوئی اور یہیں اس کی ترقی ہوئی۔
</h1>
</body>
</html>
```

عنوان

اردو، جنوبی ایشیا کی ایک اہم اور بڑی زبان ہے۔ برصغیر ہند کی آزادی کے بعد سے اس زبان کی مقبولیت میں اضافہ ہی ہوتا جا رہا ہے۔ یہ ہندوستان کی اٹھارہ (18) قومی زبانوں میں سے ایک ہے۔ یہ پاکستان کی سرکاری زبان ہے۔ حالانکہ اس زبان پر عربی و فارسی کے اثرات ہیں لیکن عربی و فارسی کے برعکس یہ ہندی کی طرح ایک ہند آریائی زبان ہے جو برصغیر ہند میں پیدا ہوئی اور یہیں اس کی ترقی ہوئی۔

## 5. Implementing appropriate line breaking:

Developers can use appropriate line-breaking rules to ensure that the text is displayed correctly and that words are not broken in the middle. For example, the word "السلام" (As-Salaam) should not be broken in the middle, as it changes the meaning of the word.

## 6. Implementing appropriate ligatures:

Developers can use appropriate ligatures to ensure that the Urdu script is rendered correctly. For example, the letters "ل" and "م" should be joined together to form the ligature "لم" when they appear together in a word. Ligatures, such as "ٹ" and "ہ" to create "ٹھ" or "پ" and "ہ" to create "پھ", are commonly used to improve the flow and legibility of Urdu text. Ligatures also help in reducing the space between letters while maintaining proper legibility. Urdu script includes several ligatures, which are combinations of two or more letters that form a single glyph. Ligatures, such as "ٹ" and "ہ" to create "ٹھ" or "پ" and "ہ" to create "پھ", are commonly used to improve the flow and legibility of Urdu text. Ligatures also help in reducing the space between letters while maintaining proper legibility.

Different types of ligatures used in Urdu language:

1. Two-letter ligatures: These are ligatures formed by combining two Urdu letters to create a single glyph. Examples include "کا" (ka), "میں" (main), "کی" (ki), and "تھا" (tha).
2. Three-letter ligatures: These are ligatures formed by combining three Urdu letters to create a single glyph. Examples include "گھر" (ghar), "جھاڑ" (jhaar), and "کھا" (kha).
3. Half-letter ligatures: These are ligatures formed by combining the first half of one letter with the second half of another letter. Examples include "آباد" (abad), "آغاز" (aaghaaz), and "خوشبو" (khushbu).
4. Initial-middle ligatures: These are ligatures formed by combining the initial letter of one word with the middle letter of another word. Examples include "رکھتے ہیں" (rakhtay hain), and "پڑھنے والا" (parhney wala).
5. Medial-final ligatures: These are ligatures formed by combining the medial letter of one word with the final letter of another word. Examples include "جانوروں کو" (janwaron ko), "سمجھتے ہیں" (samajhtay hain), and "بچوں کے" (bachon ke).
6. Diacritical marks in Urdu are marks used to modify the pronunciation of Urdu letters. They are also called "aeraab" in Urdu.

## 7. Using anchor tag

If there is no space defined between the anchor tags, the declared links merged. There is a need to introduce the space so that links can display separate and meaningful.

<pre> !DOCTYPE html&gt; &lt;html&gt; &lt;head&gt; &lt;meta charset="UTF-8"&gt; &lt;title&gt;My Web Page&lt;/title&gt; &lt;style&gt;  @font-face { font-family: 'nastaleeq-nafees'; src: url('https://cdn.jsdelivr.net/gh/tariq-abdullah/urdu-web-font-CDN/PDMS_NastaliqNafees.ttf') format('opentype'); }  body { font-family: "Amiri", serif; font-size: 20px; line-height: 1.5; }  /* CSS rule for Arabic text with right-to-left direction */ .arabic-text { direction: rtl; letter-spacing: 0.2em; }  /* CSS rule for Urdu text with right-to-left direction */ .urdu-text { direction: rtl;  font-family: "nastaleeq-nafees", cursive; font-kerning: normal; }  .urdu-text2 { direction: rtl; letter-spacing: 0.7pt; font-family: "nastaleeq-nafees", cursive; font-kerning: normal; } &lt;/style&gt; &lt;/head&gt; &lt;body&gt;  &lt;h2&gt;With Any Nastaleeq Font &lt;/h2&gt;  &lt;h3&gt;Correct Form &lt;/h3&gt; &lt;a href="#a" class="urdu-text2"&gt;فاصلہ&lt;/a&gt; &lt;a class="urdu-text2" href="#d"&gt;پے&lt;/a&gt;  &lt;h3&gt;In correct Form &lt;/h3&gt; &lt;p&gt;when there is no space character between two links, the fonts are merged&lt;/p&gt; &lt;a href="#b" class="urdu-text2" &gt;فاصلہ&lt;/a&gt;&lt;a class="urdu-text2" href="#c"&gt;پے&lt;/a&gt;  &lt;h2&gt;Without Font &lt;/h2&gt;  &lt;h3&gt;Correct Form &lt;/h3&gt; &lt;a href="#a"&gt;فاصلہ&lt;/a&gt; &lt;a href="#d"&gt;پے&lt;/a&gt;  &lt;h3&gt;In correct Form &lt;/h3&gt; &lt;p&gt;when there is no space character between two links, the fonts are merged&lt;/p&gt; &lt;a href="#b" &gt;فاصلہ&lt;/a&gt;&lt;a href="#c"&gt;پے&lt;/a&gt;  &lt;/body&gt; &lt;/html&gt; </pre>	<h2>With Any Nastaleeq Font</h2> <h3>Correct Form</h3> <p>فاصلہ پے</p> <h3>In correct Form</h3> <p>when there is no space character between two links, the fonts are merged</p> <p>فاصلہپے</p> <h2>Without Font</h2> <h3>Correct Form</h3> <p>فاصلہ پے</p> <h3>In correct Form</h3> <p>when there is no space character between two links, the fonts are merged</p> <p>فاصلہپے</p>
--	--

## 8. Initial letter styling

Initial letter styling should be applicable on the final un-joined form or standalone from the character of the word .ie

## Final Un-joined Form

ہندوستان کے دینی مدارس دنیا کا ایک ایسا سب سے بڑا غیر سرکاری نظام (این جی او، نیٹ ورک) ہیں جو عام لوگوں کے تعاون سے، عام لوگوں کے ذریعے اور عام لوگوں کے لیے چلایا جاتا ہے۔ اس کے منتظمین کوئی امر اور روسا اور خواص نہیں ہوتے۔

کا ان نامی ہاتھی سنہ 2012ء سے اس کی ساتھی کی ہتھنی کی وفات کے بعد تنہائی کا شکار ہے۔

یہ زبان شمال مغربی ہندوستان کی علاقائی اپ بھرنشوں سے رابطے کی زبان کے طور پر ابھری

طلباء کو برخاست کیے جانے کی مخالفت میں مظاہرہ کر رہے اسٹوڈنٹس نے مڈ سیمسٹر کے امتحان کا بائیکاٹ کر دیا ہے

## Standalone From

اردو، جنوبی ایشیا کی ایک اہم اور بڑی زبان ہے۔ برصغیر ہند کی آزادی کے بعد سے اس زبان کی مقبولیت میں اضافہ ہی ہوتا جا رہا ہے۔ یہ ہندوستان کی اٹھارہ (18) قومی زبانوں میں سے ایک ہے۔ یہ پاکستان کی سرکاری زبان ہے۔ حالانکہ اس زبان پر عربی و فارسی کے اثرات ہیں لیکن عربی و فارسی کے برعکس یہ ہندی کی طرح ایک ہند آریائی زبان ہے جو برصغیر ہند میں پیدا ہوئی اور یہیں اس کی ترقی ہوئی

یاستی اکادمیاں جو ہندوستان میں اردو کے لیے کام کر رہی ہیں۔

### 9. Alphabetic Lists & counters:

Urdu alphabet and Abjad (Alpha, Beta, Gamma) are widely used as alphabetic listing as shown below:

ا	عشق مجھ کو نہیں وحشت ہی سہی
ب	میری وحشت تری شہرت ہی سہی
پ	قطع کیجے نہ تعلق ہم سے
ت	کچھ نہیں ہے تو عداوت ہی سہی
ٹ	میرے ہونے میں ہے کیا رسوائی
ث	اے وہ مجلس نہیں خلوت ہی سہی
ج	ہم بھی دشمن تو نہیں ہیں اپنے
چ	غیر کو تجھ سے محبت ہی سہی
ح	اپنی ہستی ہی سے ہو جو کچھ ہو
خ	آگہی گر نہیں غفلت ہی سہی
د	عمر ہر چند کہ ہے برق خرام
ڈ	دل کے خون کرنے کی فرصت ہی سہی
ذ	ہم کوئی ترک وفا کرتے ہیں
ر	نہ سہی عشق مصیبت ہی سہی
ڑ	کچھ تو دے اے فلک ناانصاف
ز	آہ و فریاد کی رخصت ہی سہی
ژ	ہم بھی تسلیم کی خو ڈالیں گے
س	ہے نیازی تری عادت ہی سہی
ش	یار سے چھیڑ چلی جائے اسد
ص	عشق مجھ کو نہیں وحشت ہی سہی
ض	میری وحشت تری شہرت ہی سہی
ط	قطع کیجے نہ تعلق ہم سے
ظ	کچھ نہیں ہے تو عداوت ہی سہی
ع	میرے ہونے میں ہے کیا رسوائی
غ	اے وہ مجلس نہیں خلوت ہی سہی
ف	ہم بھی دشمن تو نہیں ہیں اپنے
ق	غیر کو تجھ سے محبت ہی سہی
ک	اپنی ہستی ہی سے ہو جو کچھ ہو
گ	آگہی گر نہیں غفلت ہی سہی

ل	عمر ہر چند کہ ہے برق خرام
م	دل کے خوں کرنے کی فرصت ہی سہی
ن	ہم کوئی ترک وفا کرتے ہیں
ں	نہ سہی عشق مصیبت ہی سہی
و	کچھ تو دے اے فلک ناانصاف
ہ	آہ و فریاد کی رخصت ہی سہی
ھ	ہم بھی تسلیم کی خو ڈالیں گے
ء	بے نیازی تری عادت ہی سہی
ی	یار سے چھیڑ چلی جائے اسد

In the ordered list, if the characters (defined in CSS 'symbol' property) repeat, they should not join.

۱۱	دل کے خوں کرنے کی فرصت ہی سہی
ب ب	ہم کوئی ترک وفا کرتے ہیں
ت ت	نہ سہی عشق مصیبت ہی سہی
ث ث	کچھ تو دے اے فلک ناانصاف
ج ج	آہ و فریاد کی رخصت ہی سہی

Alphabetic Listing		Alpha, beta Listing	
U+0627	ا	U+0627	ا
U+0628	ب	U+0628	ب
U+067E	پ	U+062C	ج
U+062A	ت	U+062F	د
U+0679	ث	U+06C1	ہ
U+062B	ٹ	U+0648	و
U+062C	ج	U+0632	ز
U+0686	چ	U+062D	ح
U+062D	ح	U+0637	ط
U+062E	خ	U+06CC	ی
U+062F	د	U+06A9	ک

U+0688	ڈ	U+0644	ل
U+0630	ذ	U+0645	م
U+0631	ر	U+0646	ن
U+0691	ڑ	U+0633	س
U+0632	ز	U+0639	ع
U+0698	ژ	U+0641	ف
U+0633	س	U+0635	ص
U+0634	ش	U+0642	ق
U+0635	ص	U+0631	ر
U+0636	ض	U+0634	ش
U+0637	ط	U+062A	ت
U+0638	ظ	U+062B	ث
U+0639	ع	U+062E	خ
U+063A	غ	U+0630	ذ
U+0641	ف	U+0636	ض
U+0642	ق	U+0638	ظ
U+06A9	ک	U+063A	غ
U+06AF	گ		
U+0644	ل		
U+0645	م		
U+0646	ن		
U+06BA	ں		
U+0648	و		
U+06C1	ہ		
U+06BE	ھ		
U+0621	ء		
U+06CC	ی		
	۱۱		
	ب ب		
	ت ت		
	ث ث		
	ج خ		

## 10. Forms & user interaction

Where a page allows users to input information in forms or interact with a page, there may be special requirements for the layout and orientation of the input fields and associated labels.

in Urdu text layout, Form, checkbox, radio buttons etc. has orientation right to left as shown below:



# ملازمتوں کے لیے درخواست

نام:

ایمیل:

عمر:

جنس:

مرد

مطلوبہ تسخّل:

سافٹ ویئر انجینئر

تجربہ:

مبتدی

متوسط

ماہر

مہارتیں:

پیغام:

جمع کریں

## 11. References

1. **Gap Analysis:** <https://w3c.github.io/alreq/gap-analysis/>
2. **Arabic Text Layout Requirements:** <https://w3c.github.io/alreq/>
3. **W3C Language Matrix:** <https://w3c.github.io/typography/gap-analysis/language-matrix.html>
4. **Language Enablement Index:** <https://www.w3.org/TR/typography/>
5. **Readymade Counter Styles:** <https://www.w3.org/TR/predefined-counter-styles/>