

Direction

Arabic script is written from right to left. Numbers, even Arabic numbers, are written from left to right, as is text in a script that is normally left-to-right.

When the main script is Arabic, the layout and structure of pages and documents are also set from right to left.

[Unicode Standard Annex #9, Unicode Bidirectional Algorithm](#) details an algorithm for rendering right-to-left text and covers a myriad of situations in mixing different kinds of characters. A simpler explanation of the basics of the algorithm exists in the W3C article [Unicode Bidirectional Algorithm basics](#). You can refer to these documents for more information about Unicode's bidirectional algorithm.

A brief overview of the bidirectional ("bidi" for short) algorithm follows, because direction is an essential part of how Arabic script is used.

The logical order of a text is the order in which its characters are usually typed and read. This is the order used for saving texts in digital format. This visual order is the order of characters when the text is represented in its written form. The goal of the bidi algorithm is to convert the logical order to visual order to prepare the texts for rendering.

TODO: image showing logical and visual orders

An initial step of the process involves determining each paragraph's "base direction": whether the paragraph is left-to-right or right-to-left. The base direction is either explicitly set by the author, inherited from the page, or (typically for user-generated content) detected based on the content of the paragraph. The base direction has two important uses later in the process. The next step is to split the text into "directional runs". Each directional run is a sequence of characters with the same direction.

TODO: image showing a text made of different runs

Inside each run, all the characters follow the same order. The runs themselves are ordered for visual representation from left to right or from right to left, depending on the base direction of the paragraph. This is the first effect of the base direction.

TODO: image showing the order of runs depending on the base direction

Unicode has a “bidi category” property defined for each character that is used to determine the direction of each character. All the Arabic letters are marked as right-to-left characters, while Latin characters have the left-to-right category.

Some characters, mostly punctuation, are “neutral”. The direction of these characters is derived from their surrounding characters. If a neutral character is surrounded by characters of the same direction (e.g. an space surrounded by Arabic letters), it gets the direction of its neighbors. Otherwise (e.g. a space between an Arabic and a Latin, or a neutral character appearing at the start or the end of a paragraph), the neutral character gets its direction from the paragraph’s base direction. This is the another effect of the base direction in the bidi algorithm.

The above explanation of the bidi algorithm is highly simplified, to convey only the essentials of how Arabic text is transformed for rendering. The actual algorithm deals with many more character types and edge cases.