

This is a brief tutorial on the Arabic, Urdu or Persian writing system. These languages are written using scripts which go from right to left, quite different from the normal left to right writing for most languages of the world. Arabic, Persian and Urdu relate to the Semitic language group which include others such as Hebrew, Avestan which are also written right to left.

This tutorial has been prepared using the special editor developed at the Indian Institute of Technology Madras in India. This Institute has developed useful multilingual software to work with all the Indian languages . Urdu is one of the languages of India and hence the motivation for developing this editor.

Basic observations on the writing system for Arabic and Urdu.

The Alphabet consists of only shapes for the consonants. The vowel with a consonant is indicated by a mark above or below the shape for the consonant. In Arabic, only three vowels are seen normally corresponding to the most basic vowels "a,i,u" observed in most languages. The long vowels corresponding to "a,,," are also used. Traditionally the Arabic writing employed in the Quran explicitly shows the vowel marks so that reading is proper. Modern Arabic writing system tends to omit the vowel marks but the natives will be able to read the text from experience.

Before we look at the details of the consonants, let us look at some simple text strings.

This has three syllables "ki taa bu" . كِ تَابُ

The same string written without the vowel marks . ك ت ب

In practice, while writing the letters, the writer joins the letters with one another and hence the writing appears very continuous.

Kutib . كُتِبَ

Notice how the letters have been joined together (not the vowel marks) and the writing appears continuous. What happens in practice however is that some of the consonants (six of them

in Arabic) do not join with other letters and one will see breaks.

The writing system is reasonably close to a syllabic writing system where each syllable is of the "CVC", "CV" or "C" form. Doubling of consonants is also provided for.

We now make an important observation.

The shape assigned to a syllable varies depending on the position occupied by the syllable in the word. Thus, each consonant has three possible shapes when seen within a word, besides a fourth "standalone form". The three variations correspond to the initial, medial and final forms for a consonant depending on whether the consonant occurs as the first syllable, a middle syllable or the final syllable of a word. It must be kept in mind that this is a general rule and deviations will be seen in specific cases. The following lines show the four forms for the first four consonants in the list of letters.

Final	Medial	Initial	Standalone
ب	ب	ب	ب
ت	ت	ت	ت
ث	ث	ث	ث
ج	ج	ج	ج

One observes that the initial form has an ending suited for joining with the next letter while the final form has a beginning suited for joining with the previous letter. The middle form for most of the consonants will have ends suited for joining with letters on either side. However, the six non joining letters will introduce breaks in the otherwise continuous text, which is often a delight for the skillful calligrapher. The non joining letters will join with the preceding letter but not the one which follows. We must remember that logically the letters will have to be identified from the right end of the line.

The Arabic alphabet includes twentyseven consonants out of which six are non joining. A consonant called the "hamza" has the characteristics of a "Glottal stop" and is used as the twentyeighth. Besides these, the alphabet includes the "Alif" which is a shape that serves as a "carrier" or a "chair" for the hamza. In the case of Urdu, the Alif is also used with the vowel marks to identify standalone vowels.