All subset and indexing functions work on code points (the smallest binary unit to represent part of the abstract character when stored in memory, byte for String, word for WString) and not code units (the abstract character). You can usually get the right behavior if you use WString, but it is more of a hack/convenience.