Projekt UnicodeBuild

Analiza problemu

Jarosław Kuchta

9.06.2025

# Języki i systemy pisma

Wikipedia definiuje język (ang. language) jako ustrukturalizowany system komunikacji, na który składa się słownik i gramatyka. Język służy przede wszystkim jako środek do komunikacji między ludźmi. Komunikacja ta może być *werbalna* lub *niewerbalna*. Komunikacja werbalna polega na przekazywaniu informacji w formie języka mówionego lub pisanego, z użyciem zrozumiałego dla obu stron komunikacji zbioru słów (*słownika*) i reguł układania słów w sensowne zdania (*gramatyka*). Komunikacja niewerbalna nie używa słów, a informacja może być przekazywana poprzez gesty (tzw. *mowa ciała*), dotyk, mimikę twarzy. Komunikacja niewerbalna może być trudniejsza do wyrażenia i zrozumienia, a także bardziej niejednoznaczna niż komunikacja werbalna. Z drugiej strony komunikacja niewerbalna umożliwia porozumienie między stronami nie znającymi wspólnego języka. Unicode zajmuje się przede wszystkim komunikacją werbalną, pisemną, ale w niektórych przypadkach słowa mogą być zastąpione przez symbole graficzne, w tym reprezentujące np. gesty i mimikę.

Wyróżnia się języki naturalne lub sztuczne. *Języki naturalne* (ang. natural languages) są stosowane wyłącznie do komunikacji między ludźmi. Języki naturalne tworzyły się przez długi czas, w sposób naturalny, a ich reguły były wtórnie określone względem praktycznego użycia. Języki naturalne podlegały (i w dalszym ciągu podlegają) naturalnym zmianom, część *języków wymarła* (ang. extinct languages), tzn. nie są już używane przez żyjących ludzi, a ich poznanie jest możliwe dzięki *artefaktom pisemnym,* czyli inskrypcjom. Niektóre języki wymarłe pozostają nieodkodowane, tzn. znany jest zestaw symboli, ale nieznane jest ich znaczenie. Unicode zajmuje się przede wszystkim językami *żywymi* (tzn. niewymarłymi), ale część standardu dotyczy też języków wymarłych, w tym nieodkodowanych.

W przeciwieństwie do języków naturalnych *języki sztuczne* (ang. artificial languages) zostały opracowane w stosunkowo krótkim czasie, zazwyczaj przez jedną osobę lub niewielką grupę osób, w określonym celu. Języki sztuczne mogą być używane zarówno do komunikacji między ludźmi (np. esperanto, język Braille’a), jak i pomiędzy ludźmi a urządzeniami (np. język APL służący do programowania komputerów) albo też między samymi urządzeniami (bez pośrednictwa ludzi). Unicode zajmuje się językami sztucznymi o ile wymagają one specjalnie zdefiniowanego zestawu znaków.

Komunikacja werbalna może być mówiona lub pisana, tak więc język może być realizowany w *mowie* lub w *piśmie*. Większość języków mówionych ma swoje odpowiedniki w piśmie, ale istnieją też języki, których sposób zapisu nie istnieje lub jest nieznany. Unicode zajmuje się tylko pisemną formą języka, chociaż część systemów pisma stanowi wierne odwzorowanie mowy, a część systemów jest oparta o symbole reprezentujące określone znaczenie, a sposób ich wymowy niekoniecznie jest określony.

*Mowa* (ang. speech) polega na artykułowaniu głosu poprzez użycie ludzkich organów mowy, takich jak krtań, język (ang. tongue), podniebienie miękkie i twarde, dziąsła, wargi. System organów mowy, przez który przechodzi głos nazywa się kanałem głosowym (ang. vocal tract). Mowa może być utrwalona za pomocą środków technicznych (nagranie) do późniejszego odtworzenia. Współcześnie możliwe jest też *generowanie mowy* bez użycia ludzkich organów mowy, lecz środków technicznych dających uformowanie ciągu dźwięków tak, aby przypominały mowę ludzką.

Wypowiedź ustna (ang. oral statement) jest to uporządkowana w czasie sekwencja fonemów, przy czym *fonem* (ang. phoneme) jest abstrakcyjną jednostką mowy, realizowaną w mowie rzeczywistej przez głoski. *Głoska* (ang. phone) jest najmniejszym możliwym elementem wypowiedzi ustnej o stałych cechach fonetycznych (artykulacyjnych i akustycznych). Cechy artykulacyjne reprezentują sposób kształtowania dźwięku, a cechy akustyczne – sposób propagacji dźwięku. Wyróżnia się samogłoski, spółgłoski i półsamogłoski.

*Samogłoski* (ang. vowels) są artykułowane bez zamykania kanału głosowego. Rozkład energii akustycznej w czasie jest jednorodny. W różnych językach używa się różnej liczby samogłosek, od 2 do około 30 (w językach europejskich od 5 do 16). W części systemów pisma samogłoski są zapisywane jako osobne znaki, a w części jako dodatkowe symbole (kropki, kreski etc.) umieszczane ponad, pod lub obok znaków reprezentujących samogłoski.

*Spółgłoski* (ang. consonants) są artykułowane z przymknięciem lub całkowitym zamknięciem kanału głosowego (wyjątek stanowi spółgłoska *h*). Przy artykulacji spółgłoski często następuje faza *plozji* polegająca na nagłym otwarciu kanału głosowego i uwolnieniu przepływu powietrza do jamy nosowej lub jamy ustnej (np. po bokach języka). Większość samogłosek jest artykułowanych przy wydechu powietrza z płuc, ale część spółgłosek (tzw. mlaski lub kliki) powstaje przy nagłym otwarciu zwarcia warg lub języka bez przepływu powietrza z płuc, a niektóre przy krótkim wdechu (cmoki). W różnych językach liczba spółgłosek jest różna (od 6 do 84).

Pojedyncze samogłoski lub zespoły złożone z samogłoski i spółgłosek (jednej lub kilku) tworzą *sylaby* (ang. syllables). Samogłoskę nazywa się wówczas ośrodkiem lub jądrem sylaby (ang. nucleus). W niektórych językach ośrodkiem sylaby może być *spółgłoska sylabiczna* (ang. sonant). Są języki pisane, w których różne sylaby zapisuje się w postaci osobnych znaków. Są to tzw. *sylabariusze* (ang. syllabaries).

Czasami oprócz pojęcia sylaby stosowane jest pojęcie mory. Mora (ang. mora) to najmniejsza jednostka czasu, równa lub krótsza od sylaby, istniejąca w niektórych językach mówionych, w których długość fonetyczna ma znaczenie. Na przykład w języku japońskim nazwa miasta *Osaka* składa się z trzech sylab (*O-sa-ka*), ale czterech mor (*O-o-sa-ka*), ponieważ pierwsza sylaba, jest wymawiana z długą samogłoską.

W ramach poszczególnych języków wyróżnia się dialekty. W językoznawstwie przyjmuje się, że *dialekt* to odmiana języka o cechach fonetycznych, leksykalnych lub gramatycznych, które odróżniają ją od innych odmian tego języka. Granica pomiędzy językiem a dialektem bywa nieostra, tak że nie wiadomo dokładnie, czy modyfikacja słownika i zasad gramatycznych pewnego języka doprowadziła do powstania nowego języka, czy też w dalszym ciągu trzeba mówić o dialekcie języka standardowego. W Unicode kwestia dialektów ogranicza się do języków pisanych. Może być tak, że kilka języków pisanych stosuje ten sam zbiór znaków (np. cyrylicę), ale jego część jest wspólna, a niektóre z języków stosują różne rozłączne podzbiory znaków. Tak jest np. w przypadku języków rosyjskiego, ukraińskiego i białoruskiego. W takich przypadkach Unicode zaznacza to w nazwach opisowych poszczególnych znaków.

Oprócz języków mówionych (ang. spoken languages) istnieją też języki migowe (ang. sign languages) stosowane do komunikacji z ludźmi niesłyszącymi lub niemówiącymi. Istnieje wiele języków migowych, które rozwijały się naturalnie w różnych grupach ludzi, dlatego języki migowe zalicza się do języków naturalnych. Język migowy należy odróżnić od języka ciała (zaliczanego do języków niewerbalnych), gdyż każdy język migowy ma zarówno symbolikę opartą głównie na gestach rąk, jak i reguły stosowania tych gestów. Gesty mogą wyrażać pewne idee (podobnie jak w języku ideograficznym), głoski (jak w alfabecie) lub sylaby. Unicode ma w swoim zakresie blok znaków definiujących gesty (poprzez ułożenie dłoni, palców etc.) stosowane w języku migowym opracowanym w 1974 r. przez Valerie Sutton. Nie definiuje znaczenia gestów, a jedynie ilustruje ich wygląd. Zbiór znaków może służyć do graficznego zapisu różnych języków migowych.

Unicode odnosi się do języków nie bezpośrednio, lecz pośrednio – poprzez tzw. systemy pisma. W językoznawstwie pod pojęciem *systemu pisma* (ang. writing system) rozumie się zbiór symboli, zwany *pismem* (ang. script), służący do utrwalania języka mówionego oraz zestaw reguł opisujących poprawne użycie (i łączenie) tych symboli. Jeden język mówiony może mieć kilka różnych systemów pisma, a jeden system pisma może być stosowany w wielu językach.

Pismo zapisuje język mówiony tak, aby osoba czytająca pismo mogła odtworzyć wypowiedź ustną nawet po tysiącach lat. Do czytania pisma najczęściej stosuje się narządy wzroku, ale są też pisma odczytywane przez dotyk (np. pismo Braille’a). Utrwalenie może przyjmować różne formy. Symbole mogą być napisane na papierze lub pergaminie, wyryte w kamieniu lub na tabliczce glinianej (pismo klinowe), malowane pędzelkiem na tkaninie. Istnieje nawet pismo węzełkowe, które jest utrwalane przez wiązanie węzełków na sznurku, a które przypisuje różne znaczenia różnym rodzajom węzłów i ich kombinacji. W ten sposób symbolem staje się pojedynczy węzeł lub układ węzłów. Pismo Braille’a powstaje poprzez wytłoczenie określonej kombinacji wypukłości w podstawowej siatce 2 x 3 punktów. Pismo może być też zakodowane w postaci cyfrowej w nowoczesnej technologii elektromagnetycznej lub optycznej, tak że potrzeba odpowiednich urządzeń dla odtworzenia wypowiedzi. Tym ostatnim sposobem zapisu Unicode się nie zajmuje. Chociaż zbiory znaków są zakodowane z użyciem 24-bitowego zapisu cyfrowego, to definicje znaków dotyczą graficznej formy zapisu (nawet w przypadku języka migowego lub pisma Braille’a).

Większość punktów kodowych zdefiniowanych w Unicode jest przypisana do jednego z systemów pisma stosowanych w języku naturalnym lub sztucznym. Część jednak reprezentuje symbole graficzne niestosowane w żadnym z nich. Są tam symbole stosowane w matematyce, chemii, muzyce i innych dziedzinach. Trudno nazwać takie zbiory symboli *pismem*. Jeśli te zbiory symboli można skojarzyć z regułami ich stosowania, to taki system pisma będzie nazywany *notacją* (ang. notation). Są jednak takie symbole, dla których nie obowiązują żadne reguły. Takie symbole tworzą po prostu zbiory symboli (ang. symbol sets).

# Klasyfikacja znaków Unicode

Klasyfikacja znaków Unicode, czyli tzw. punktów kodowych (ang. code points), może być dokonywana na podstawie:

* nazwy opisowej znaku,
* kategorii ogólnej znaku,
* bloku Unicode, do którego należy.

Klasyfikacja nie ma charakteru rozłącznego, lecz uzupełniający.

## Jednostki klasyfikacyjne

Każdy znak Unicode, można przypisać do pewnej jednostki organizacyjnej, którą może być:

* *pismo* (ang. script) – zbiór znaków do zapisu języka naturalnego lub sztucznego; z pismem jest skojarzony zestaw reguł określających sposób zapisu i odczytu znaków; zestaw ten są definiowane poza standardem Unicode. Pismo alternatywnie z notacją jest nazywane systemem pisma (ang. writing system).
* *notacja* (ang. notation) – zbiór znaków stosowanych w określonej dziedzinie nauki lub techniki; z notacją jest skojarzony zestaw reguł określających sposób zapisu i odczytu znaków; zestaw ten są definiowane poza standardem Unicode.
* *zbiór symboli* (ang. symbol set) – zbiór znaków stosowanych w różnych dziedzinach; reguł sposobu użycia tych symboli nie ma lub są jedynie zaleceniami.
* *podzbiór symboli* (ang. symbol subset) – wybrany podzbiór ze zbioru symboli.
* *obszar* (ang. area) – taki zbiór znaków, w którym poszczególne znaki nie są definiowane indywidualnie; zamiast tego Unicode jedynie definiuje początek i koniec obszaru.
* *język* (ang. language) – zazwyczaj język, który używa danego znaku jest określany przez pismo, ale w pewnych przypadkach dodatkowo jest opisywany w nazwie znaku.
* *grupa* (ang. group) – zbiór nadrzędny obejmujący inne jednostki organizacyjne, które mają pewne wspólne cechy; np. wyróżnia się rodzinę pism brahmickich (ang. Brahmi scripts).

Pismo i notację nazywa się *systemami pisma* (ang. writing systems). Klasyfikacja znaków do systemów pisma jest ścisła wynika z danych UCD. Klasyfikacja do pozostałych jednostek jest bardziej uznaniowa i może zależeć od algorytmu tworzonego przez użytkownika.

## Rodzaje pisma

Wyróżnia się kilka rodzajów pisma. Są to:

* *alfabet* (ang. alphabet) – pismo, w którym znaki nazywane są literami, a litery reprezentują głoski (samogłoski lub spółgłoski). Niektóre głoski są reprezentowane przez dwuliterowe sekwencje.
* *sylabariusz* (ang. syllabary) – pismo, w którym znaki reprezentują sylaby (lub mory). Symbol w sylabariuszu zazwyczaj reprezentuje dźwięk spółgłoski, po którym następuje dźwięk samogłoski (jądro). Sylabariusze są stosowane m.in. w języku japońskim i w językach z rodziny Yi ze wschodnich obszarów Chin.
* *alfabet sylabiczny*, *abugida* (ang. abugida) – rodzaj alfabetu, w którym jedna litera reprezentuje najczęściej całą sylabę z domyślną samogłoską. Jeśli w sylabie znajduje się inna samogłoska niż domyślna, to zapisuje się ją za pomocą znaków diakrytycznych. Istnieją również odrębne znaki reprezentujące samogłoski. Alfabety sylabiczne są stosowane m.in. w językach bengalskim i tybetańskim.
* *pismo spółgłoskowe*, *abdżad* (ang. abjad) – rodzaj alfabetu posiadający tylko znaki do zapisu spółgłosek. Samogłoski są domyślane przez czytelnika. Przykłady: pismo arabskie, aramejskie, fenickie, hebrajskie, syryjski.
* *pismo logograficzne* (ang. logographic) – pismo, w którym pojedyncze znaki składają się z dwóch komponentów: komponentu reprezentującego znaczenie (ideę) oraz komponentu reprezentującego wymowę. Pismo chińskie jest przykładem pisma logograficznego.
* *pismo ideograficzne* (ang. ideographic) – pismo (lub raczej notacja), w którym znaki reprezentują pewne abstrakcyjne idee. Ideogramami są np. liczby i symbole matematyczne, symbole waluty, znak ‘&’.
* *pismo piktograficzne* (ang. pictographic) – pismo (lub zbiór symboli), w którym znaki reprezentują pewne znaczenia przez podobieństwo do fizycznych obiektów. Piktogramy były stosowane np. w starożytnych cywilizacjach sumeryjskiej, egipskiej i chińskiej. Z czasem piktogramy ewaluowały w logogramy. Współcześnie symbol koperty jest używany do oznaczenia wychodzącej lub przychodzącej poczty e-mail.
* pismo *hieroglificzne* (ang. hieroglyphs) – starożytne pismo egipskie łączące elementy ideograficzne, logograficzne, sylabiczne i alfabetyczne.
* *pismo logo-sylabiczne* (ang. logo-syllabary) – pismo zawierające dwa osobne zestawy znaków: jeden dla zapisu znaczenia, drugi do zapisu sylabicznego. Przykłady zawierają inskrypcje starożytnych kultur ameryki środkowej.
* *pismo klinowe* (ang. cuneiform) – rodzaj najstarszego pisma logo-sylabicznego utrwalonego na tabliczkach glinianych przez wyciskanie kształtów klinowych za pomocą trzciny.

## Kierunkowość pisma

Alfabety mogą być typu unikameralnego (ang. unicameral), zwanego również jednoszeregowym (ang. unicase). Taki typ alfabetu nie ma oddzielnych liter wielkich i małych. Pismo hebrajskie, arabskie, pisma brahmickie to przykłady alfabetów *unikameralnych*. Z kolei pismo łacińskie, greckie, cyrylica i ormiańskie są bikameralne.

Systemy pisma można podzielić pod względem podziału tekstu na linie i kolejności ich odczytywania. Wyróżnia się:

* pisma poziome i pionowe; w poziomych kolejne znaki tekstu pisane są w liniach poziomych, a pismach pionowych – w liniach poziomych.
* pisma pionowe dzieli się na te, w których pionowe linie tekstu czyta się od strony lewej do prawej lub od prawej do lewej.
* pisma poziome dzieli się na te, w których kolejne znaki w linii czyta się od strony lewej do prawej i te, w których czyta się od prawej do lewej.

Wiele pism z Azji wschodniej może być pisanych w układzie poziomym lub pionowym. Znaki chińskie, koreańskie znaki hangul i japońskie znaki kana mogą być pisane w obu układach, gdyż są symbolami logo-graficznymi lub logo-sylabicznymi na planie kwadratu.

W pismach o liniach układających się poziomo znaki są z reguły pisane i czytane od lewej do prawej (LTR). Są jednak takie (arabskie, hebrajskie), w których znaki są czytane z reguły od strony prawej do lewej (RTL). Ze względu na liczne wyjątki od tej reguły są one nazywane pismami *dwukierunkowymi* (ang. bidirectional scripts).

## Kategorie ogólne znaków

Unicode przypisuje każdy znak do jednej z 30 kategorii ogólnych. Każdą z nich oznacza dwuliterowym skrótem. I tak:

* Do kategorii Lu (Uppercase Letter) zalicza wszystkie wielkie litery alfabetu, a do kategorii Ll (Lowercase Letter) zalicza wszystkie małe litery alfabetu. Oczywiście dotyczy to alfabetów bikameralnych. Do kategorii Lu należy 1858 znaków, a do Ll – 2258 znaków. Wszystkie litery wielkie mają swoje odpowiedniki w literach mały i vice versa.
* Kategoria Lt (Titlecase Letter) obejmuje tylko 31 znaków Unicode. Zgodnie z nazwą należą do niej cztery dwuznaki składające się z wielkie i małej litery alfabetu łacińskiego ‘ǅ’, ‘ǈ’, ‘ǋ’, ‘ǲ’, ale też 27 liter wielkich alfabetu greckiego ze znakami diakrytycznymi, np. „GREEK CAPITAL LETTER ALPHA WITH PROSGEGRAMMENI”.
* Kategoria Lm (Modifier Letter) jest zarezerwowana dla liter (lub innych symboli), które modyfikują cechy fonetyczne innych liter. Zapisywane są zwykle za lub przed modyfikowaną literą). W tym sensie są podobne do znaków diakrytycznych, ale nie łączą się w jednej komórce znakowej z modyfikowanym znakiem, lecz mają swoje własne szerokości. Takie znaki są wykorzystywane w systemach transkrypcji fonetycznej dla określenia odpowiedniej wymowy głoski. Unicode zalicza do tej kategorii 404 znaki.
* Kategoria Lo (Other Letter) jest przeznaczona dla wszystkich innych liter (w tym liter alfabetów unikameralnych), sylab oraz symboli przypominających litery (ang. letterlike symbols) z wyglądu, ale o innym wykorzystaniu (jak np. „FEMININE ORDINAL INDICATOR”). W Unicode jest zdefiniowanych 21518 takich znaków.
* Kategoria Nd (Decimal Number) oznacza cyfry dziesiętne z różnych alfabetów. W sumie jest 760 takich cyfr. Część tych znaków to cyfry dziesiętne alfabetu łacińskiego o zmodyfikowanych kształtach (np. obrysowanych, wytłuszczonych, o kształtach przypominających wyświetlacze cyfrowe.
* Kategoria Nl (Letter Number) obejmuje inne cyfry i liczby, np. z rzymskiego systemu liczbowego, z pisma klinowego. Zdefiniowano 236 takich znaków.
* Kategoria No (Other Number) jest przeznaczona dla innych znaków numerycznych, np. zawiera cyfry dziesiętne w indeksie górnym, znaki ułamków zwykłych jak ‘½’, cyfry połówkowe alfabetu tybetańskiego. Wyodrębniono 915 takich znaków.
* Kategoria Mn (Nonspacing Mark) zawiera znaki diakrytyczne (combining marks) o zerowej szerokości. Modyfikują one cechy fonetyczne liter i sylab. Mogą powodować wymowę nosową, określać samogłoskę w sylabie. Same nie mają szerokości, lecz wyświetlane są wraz z modyfikowanym znakiem (zazwyczaj ponad lub pod). Jest w sumie 2020 znaków diakrytycznych.
* Kategoria Mc (Spacing Mark) zawiera znaki łączone wyświetlane po lub przed modyfikowaną literą lub sylabą, np. znak VISARGA oznaczający w fonologii sanskryckiej bezdźwięczną spółgłoskę szczelinową krtaniową, a wyświetlaną za znakiem jako dwie kropki jedna nad drugą (podobnie do znaku dwukropka). W Unicode jest 4068 takich znaków.
* Kategoria Me (Enclosing Mark) zawiera tylko 13 znaków diakrytycznych obejmujących (enclosing marks). Należą do nich znaki modyfikujące liczebniki w cyrylicy w formie okręgu złożonego z kropek, przecinków lub krzyżyków, wewnątrz których umieszcza się modyfikowany znak liczebnika. Należą „obejmujące” nawiasy, należą koło, romb, trójkąt wewnątrz których umieszcza się inne znaki.
* Kategoria Pc (Connector Punctuation) jest przeznaczona dla znaków interpunkcyjnych łączące, np. krawat (tie) w formie łuku łączącego dwa znaki. Unicode definiuje tylko 10 takich znaków.
* Kategoria Pd (Dash Punctuation) oznacza znak interpunkcyjny będący kreską lub myślnikiem. Wyróżnia się 27 takich znaków należących do różnych alfabetów, o różnych kształtach i przeznaczeniu.
* Kategoria Ps (Open Punctuation) i Pe (Close Punctuation) oznacza znak interpunkcyjny odpowiednio otwierający i zamykający (parę). Należą tu nawiasy okrągłe, kwadratowe i klamrowe, znaki „sufitu” i „podłogi”. Razem 79 i 77 takich znaków. Różnica wynika z różnych kształtów niektórych apostrofów.
* Analogicznie kategoria Pi (Initial Punctuation) i Pf (Final Punctuation) oznacza odpowiednio otwierający i zamykający cudzysłów. Jest odpowiednio 12 i 10 takich znaków.
* Kategoria Po (Other Punctuation) łączy inne znaki interpunkcyjne – nie należące do żadnej innej kategorii znaków interpunkcyjnych. Jest 640 takich znaków zarówno wykorzystywanych wspólnie, jak i należących do różnych alfabetów.
* Kategoria Sm (Math Symbol) grupuje symbole matematyczne – 970 symboli.
* Kategoria Sc (Currency Symbol) zawiera różne znaki waluty – 63 znaki.
* Kategoria Sk (Modifier Symbol) łączy symbole modyfikujące (o takim samym znaczeniu jak Lm), ale niebędące literami – 125 znaków.
* Kategoria So (Other Symbol) to wszystkie inne symbole – razem 7376 znaków.
* Kategoria Zs (Space Separator) oznacza spacje o różnych szerokościach i przeznaczeniu – razem 17 spacji.
* Kategoria Zl (Line Separator) zawiera tylko jeden znak o kodzie \x2028 – separator linii.
* Kategoria Zp (Paragraph Separator) zawiera tylko jeden znak o kodzie \x2029 – separator akapitów.
* Kategoria Cc (Control) zawiera nie znaki widoczne, lecz kody sterujące typu C0 (ASCII) lub C1 – 65 znaków.
* Kategoria Cf (Format) łączy znaki sterujące formatem – kierunkiem pisania dla pism dwukierunkowych, znaki zerowej szerokości (spacje i łączniki) – razem 170 znaków.
* Kategoria Cs (Surrogate) to obszary znaków zastępczych do kodowania znaków 24-bitowych w kodzie 16-bitowym. Zdefiniowanych jest tylko 6 kodów dla oznaczenia początków i końców obszarów.
* Kategoria Co (Private Use) to obszary znaków zastępczych do użytku prywatnego (do dowolnego stosowania). Zdefiniowanych jest tylko 6 kodów dla oznaczenia początków i końców obszarów.
* Kategoria Cn (Unassigned) oznacza nieprzydzielone kody znaków.

## Indeksowane systemy pisma

Jednym ze sposobów klasyfikacji znaku jest sprawdzenie pliku Scripts.txt. W pliku tym określone są zbiory znaków przypisane dla poszczególnych systemów pisma. Ta klasyfikacja może być jednak zawodna. Dla przykładu do pisma Latin przypisany jest zestaw znaków o kodach od \x2160 do \x2182, czyli liczebniki rzymskie od „ROMAN NUMERAL ONE” do „ROMAN NUMERAL TEN THOUSAND”. Również do pisma Latin są przypisane znaki „KELVIN SIGN” i „ANGSTROM SIGN”.

Wśród systemów pisma Unicode jest pozycja o nazwie Common. Do tej klasy należą znaki, które mają zastosowanie w wielu systemach pisma. Jeśli istotne jest sprawdzenie, w których konkretnie systemach pisma stosuje się dany znak, to można to sprawdzić w pliku ScriptExtensions.txt. Można się stamtąd dowiedzieć, że znak diakrytyczny „CARON” jest stosowany w alfabetach opartych na piśmie łacińskim oraz w piśmie Bopomofo. W pliku ScriptExtensions.txt (w odróżnieniu od pliku Scripts.txt) nie stosuje się pełnych nazw pism, lecz ich 4-literowe nazw skrócone zdefiniowane w ISO 15924. W przypadku jakichś wątpliwości można odczytać znaczenie tych 4-literowych skróconych nazw z pliku PropertyValueAliases.txt.

## Bloki Unicode

Plik Blocks.txt definiuje bloki Unicode. Jest w sumie 338 bloków. Każdy blok jest definiowany przez początek i koniec i określa przeznaczenie punktów kodowych mieszczących się w zakresie bloku. Zdecydowana większość bloków przypisuje znaki do systemów pisma – większość do określonego alfabetu. Wówczas wszystkie znaki należące do danego bloku są poprzez plik Scipts.txt przypisane do tego samego pisma. W niektórych blokach znajdują się znaki, które są przypisane do dwóch lub kilku pism.

Są też bloki, które nie są przypisane do żadnego pisma przez Scripts.txt. W tych przypadkach przeznaczenie bloku jest opisane tylko tekstowo. Z pewnych opisów wynika, że bloki zawierają znaki wykorzystywane do pisma migowego, do pisma Braille’a, w dziedzinie matematyki, muzyki, chemii (a raczej alchemii), do transkrypcji fonetycznej Inne bloki zawierają symbole wykorzystywane w pismach ozdobnych, na mapach, ikony i emotki albo wreszcie symbole ogólnego przeznaczenia. Katalog przeznaczenia nie jest zamknięty.

## Klasyfikacja systemów pisma

Można stworzyć drzewo systemów pisma. Na najwyższym poziomie znajduje się pięć grup klasyfikujące typy systemów pisma:

* Obszary,
* Pisma,
* Notacje,
* Zbiory znaków,
* Podzbiory.

Obszary są dwa:

* Obszar prywatnego użycia (Private Use),
* Obszar kodów zastępczych (Surrogates).

Unicode odwołuje się do 170 pism. Można je dalej podzielić na grupy obejmujące:

* Pisma antyczne (Ancient scripts). Są to pisma stosowane w starożytności. Stanowią podstawę dla pism współczesnych. Należą tu hieroglify egipskie, hieroglify hetyckie (inaczej anatolijskie), pismo klinowe, pisma liniowe A i B.
* Pisma euro-kaukaskie (Euro-Caucasian scripts). Do tej grupy należą alfabety pochodzące od pisma fenickiego: łacińskie, greckie, koptyjskie, cyrylica, armeńskie, gruzińskie. Należą też pisma wymarłe: staroitalskie, cypryjskie, głagolica etc.
* Pisma bliskowschodnie (Near East scripts). Należą tu starożytne pismo aramejskie i inne pisma stosowane na Bliskim Wschodzie: hebrajskie, arabskie, syryjskie.
* Pisma azjatyckie (Asian scripts). Ta wyjątkowo duża grupa zawiera pisma stosowane w Azji środkowej i południowo-wschodniej:
  + rodzinę pism brahmickich (Brahmi scripts) - do której należy starożytne pismo brahmickie i inne pochodzące od niego alfabety sylabiczne (abugidy),
  + grupę pism wynalezionych współcześnie (w XX wieku).
* Pisma mongolskie (Mongolian scripts) – stanowią odrębną rodzinę pism stosowanych dla języka mongolskiego,
* Pisma CJK (Chinese-Japan-Korea), które się dzielą na:
  + pisma chińskie,
  + pisma japońskie,
  + pisma koreańskie.
* Pisma afrykańskie (African scripts). Zawiera głównie pisma wynalezione w XX wieku dla języków afrykańskich i jedno starożytne pismo berberyjskie.
* Pisma północnoamerykańskie (Nort American scripts). Ta grupa zawiera pisma wynalezione współcześnie dla zapisywania języków ludów zamieszkujących Amerykę Północną.
* Inne pisma (Other scripts) – należą tutaj trzy pisma wynalezione współcześnie o stosowaniu na całym świecie lub w różnych regionach świata oraz dwa określenia:
  + Common – na znaki stosowane w różnych pismach, najczęściej litery od A do Z, od a do z i cyfry od 0 do 9.
  + Inherited – na pismo określane przez kontekst.

Notacje są traktowane jak pisma. Podobnie jak pisma definiują zbiory znaków i mają reguły zapisu. Odróżnia je to, że nie wyrażają języków naturalnych, lecz pewne umowne sposoby zapisu w różnych dziedzinach. Wyróżnia się:

* notację matematyczną,
* grupę notacji numerycznych obejmującą systemy zapisu liczb i liczebników,
* grupę notacji muzycznych (np. Byzantine Musical Notation),
* grupę notacji metrycznych (do oznaczania metryki poezji lub śpiewu),
* grupę notacji fonetycznych (IPA, UPA),
* grupę notacji skrótowych (shorthand).
* grupę notacji programistycznych (APL, Z-notation),
* grupę notacji do zapisu języka gestów (signwriting).

Zbiory znaków (symbol sets) łączą znaki o pewnej wspólnej charakterystyce, ale bez określonych zasad stosowania. Zbiory znaków mogą być samodzielne, mogą być grupowane, mogą wchodzić w skład innych zbiorów znaków (w tym definiowanych przez pisma i notacje). Wyróżnia się:

* grupę strzałki i harpuny (harpun przypomina strzałkę, ale ma grot jednoostrzowy – zadzior). Grupa ta obejmuje:
  + strzałki (arrows),
  + groty strzałek (arrowheads),
  + brzechwy strzałek (arrow tails),
  + harpuny (harpoons),
* grupę symboli stosowanych w grach:
  + w szachach,
  + w Mahjong,
  + w grach karcianych,
  + w grze w kości,
  + w warcabach,
* grupę symboli stosowanych w językach CJK:
  + grupę znaków kompatybilnych z CJK (CJK Compatibility Symbols), zawierają one symbole kwadratowe (CJK i alfanumeryczne) zdefiniowane w celu zapewnienia zgodności z pismami z grupy CJK. Składa się ona z dwóch podzbiorów:
    - symbole kwadratowe CJK (CJK Squared Words),
    - kwadratowe skróty (Squared Abbreviations).
  + symbole i znaki interpunkcyjne CJK (CJK Symbols and Punctuation),
  + ideograficzne adnotacje fonetyczne (Kanbun),
  + znaki opisujące symbole ideograficzne (ideographic description characters),
  + telegraficzne symbole ideograficzne (Ideographic Telegraph Symbols),
  + elementy systemu ideograficznego Kangxi (Kangxi Radicals),
  + formy pionowe znaków (presentation forms for vertical ideographic),
  + formy szerokości połówkowej i pełnej znaków ideograficznych podzielone na dwa podzbiory,
* grupę symboli stosowanych w piśmie koreańskim Hangul,
  + Choseong – nagłos sylaby (syllable onset),
  + Jungseong – ośrodek sylaby (syllable nucleus),
  + Jongseong – wygłos sylaby (syllable coda),
* zbiór wielkich liter w alfabecie gruzińskim stosowanych w tytułach i inskrypcjach (Mtavruli),
* zbiór znaków wspólnych dla pism japońskich Katakana i Hiragana,
* grupę form liczbowych:
  + ułamki zwykłe (Vulgar Fractions),
  + ułamki północno-indyjskie (North Indic Fractions),
* symbole stosowane w systemach rozpoznawania pisma (OCR Symbols),
* grupę znaków diakrytycznych (Diacritical Symbols). Są to symbole, które dodaje się do podstawowych znaków zmieniając ich wymowę lub znaczenie. Grupa ta dzieli się na dwa podzbiory:
  + kombinatoryczne symbole diakrytyczne (Combining Diacritic Symbols), które dodaje się nad lub pod znakiem bazowym, niektóre symbole diakrytyczne otaczają symbol bazowy,
  + litery modyfikujące (Modifier Letters), czyli litery, które umieszcza się obok znaku podstawowego (najczęściej przed, ale czasami po), najczęściej mają formę indeksu górnego, a czasami dolnego,
  + symbole modyfikujące (Modifier Symbols), działające tak samo jak litery modyfikujące, ale nie są literami,
* znaki kontrolne (Control Chars), ten zbiór obejmuje znaki ASCII o kodach od 0 do 1F, 7F oraz od 80 do 9F,
* symbole waluty (Currency Symbols),

# Analiza nazw opisowych

Podstawę generowania nazw skróconych znaków będzie stanowiła analiza nazw opisowych znaków. Znajdują się one w pliku UnicodeData.txt w kolumnie drugiej, Name.

## Odfiltrowanie nazw specjalnych

Nazwa specjalna rozpoczyna się od znaku ‘<’ i kończy znakiem ‘>’. Jest 65 identycznych nazw „<control>”, które są przypisane do znaków sterujących należących do kategorii Cc. Specyficzne nazwy tych znaków są zebrane w tab. 1. Ze względu na wymaganie zwięzłości wyników trzeba zastosować nazwy skrócone z kolumny drugiej.

Wszystkie inne nazwy specjalne są stosowane do oznaczania początku i końca obszarów kodów, w których poszczególne znaki nie mają indywidualnych definicji. Obszary takie znajdują się na liście bloków Unicode. Znak rozpoczynający dany obszar ma format:

<*nazwa bloku*, First>

a znak kończący:

<*nazwa bloku*, Last>

Ponieważ w specyfikacji Unicode jest 19 takich obszarów, to jest 38 znaków o takich nazwach (tab. 2).

Tab. Nazwy znaków sterujących

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | Kod | Skrót | Nazwa | | 00 | NUL | Null | | 01 | SOH | Start of Heading | | 02 | STX | Start of Text | | 03 | ETX | End of Text | | 04 | EOT | End of Transmission | | 05 | ENQ | Enquiry | | 06 | ACK | Acknowledge | | 07 | BEL | Bell | | 08 | BS | Backspace | | 09 | HT | Horizontal Tabulation | | 0A | LF | Line Feed | | 0B | VT | Vertical Tabulation | | 0C | FF | Form Feed | | 0D | CR | Carriage Return | | 0E | SO | Shift Out | | 0F | SI | Shift In | | 10 | DLE | Data Link Escape | | 11 | DC1 | Device Control One | | 12 | DC | Device Control Two | | 13 | DC3 | Device Control Three | | 14 | DC4 | Device Control Four | | 15 | NAK | Negative Acknowledge | | 16 | SYN | Synchronous Idle | | 17 | ETB | End of Transmission Block | | 18 | CAN | Cancel | | 19 | EM | End of medium | | 1A | SUB | Substitute | | 1B | ESC | Escape | | 1C | FS | File Separator | | 1D | GS | Group Separator | | 1E | RS | Record Separator | | 1F | US | Unit Separator | | 7F | DEL | Delete | | |  |  |  | | --- | --- | --- | | Kod | Skrót | Nazwa | | 80 | PAD | Padding Character | | 81 | HOP | High Octet Preset | | 82 | BPH | Break Permitted Here | | 83 | NBH | No Break Here | | 84 | IND | Index | | 85 | NEL | Next Line | | 86 | SSA | Start of Selected Area | | 87 | ESA | End of Selected Area | | 88 | HTS | Character Tabulation Set | | 89 | HTJ | Character Tabulation With Justification | | 8A | VTS | Vertical Tabulation Set | | 8B | PLD | Partial Line Down | | 8C | PLU | Partial Line Up | | 8D | RI | Reverse Index | | 8E | SS2 | Single-Shift 2 | | 8F | SS3 | Single-Shift 3 | | 90 | DCS | Device Control String | | 91 | PU1 | Private Use 1 | | 92 | PU2 | Private Use 2 | | 93 | STS | Set Transmit State | | 94 | CCH | Cancel character | | 95 | MW | Message Waiting | | 96 | SPA | Start of Protected Area | | 97 | EPA | End of Protected Area | | 98 | SOS | Start of String | | 99 | SGCI | Single Graphic Character Introducer | | 9A | SCI | Single Character Introducer | | 9B | CSI | Control Sequence Introducer | | 9C | ST | String Terminator | | 9D | OSC | Operating System Command | | 9E | PM | Privacy Message | | 9F | APC | Application Program Command | |

Tab. . Nazwy specjalne punktów kodowych dla obszarów Unicode

| Kod | Nazwa |
| --- | --- |
| 3400 | <CJK Ideograph Extension A, First> |
| 4DBF | <CJK Ideograph Extension A, Last> |
| 4E00 | <CJK Ideograph, First> |
| 9FFF | <CJK Ideograph, Last> |
| AC00 | <Hangul Syllable, First> |
| D7A3 | <Hangul Syllable, Last> |
| D800 | <Non Private Use High Surrogate, First> |
| DB7F | <Non Private Use High Surrogate, Last> |
| DB80 | <Private Use High Surrogate, First> |
| DBFF | <Private Use High Surrogate, Last> |
| DC00 | <Low Surrogate, First> |
| DFFF | <Low Surrogate, Last> |
| E000 | <Private Use, First> |
| F8FF | <Private Use, Last> |
| 17000 | <Tangut Ideograph, First> |
| 187F7 | <Tangut Ideograph, Last> |
| 18D00 | <Tangut Ideograph Supplement, First> |
| 18D08 | <Tangut Ideograph Supplement, Last> |
| 20000 | <CJK Ideograph Extension B, First> |
| 2A6DF | <CJK Ideograph Extension B, Last> |
| 2A700 | <CJK Ideograph Extension C, First> |
| 2B739 | <CJK Ideograph Extension C, Last> |
| 2B740 | <CJK Ideograph Extension D, First> |
| 2B81D | <CJK Ideograph Extension D, Last> |
| 2B820 | <CJK Ideograph Extension E, First> |
| 2CEA1 | <CJK Ideograph Extension E, Last> |
| 2CEB0 | <CJK Ideograph Extension F, First> |
| 2EBE0 | <CJK Ideograph Extension F, Last> |
| 2EBF0 | <CJK Ideograph Extension I, First> |
| 2EE5D | <CJK Ideograph Extension I, Last> |
| 30000 | <CJK Ideograph Extension G, First> |
| 3134A | <CJK Ideograph Extension G, Last> |
| 31350 | <CJK Ideograph Extension H, First> |
| 323AF | <CJK Ideograph Extension H, Last> |
| F0000 | <Plane 15 Private Use, First> |
| FFFFD | <Plane 15 Private Use, Last> |
| 100000 | <Plane 16 Private Use, First> |
| 10FFFD | <Plane 16 Private Use, Last> |

Wszystkie pozostałe nazwy opisowe znaków podlegają dalszemu przetwarzaniu opisanemu poniżej.

## Rozpoznawanie fraz kluczowych

Bardzo często opisowa nazwa zawiera pewną frazę kluczową, której rozpoznanie jest istotne dla określenia systemu pisma, rodzaju znaku itp. Np. nazwa opisowa

LATIN CAPITAL LETTER A

ma dwie frazy:

* LETTER – na oznaczenie systemu pisma łacińskiego,
* CAPITAL LETTER – na oznaczenie wielkiej litery.

Fraza kluczowa może być poszukiwana:

* na początku nazwy opisowej, wówczas jest zakończona znakiem gwiazdki, np. „LATIN\*”;
* na końcu nazwy opisowej, wówczas jest poprzedzona znakiem gwiazdki, np. „\*CHESS”;
* w dowolnym miejscu, wówczas nie ma znaku gwiazdki albo ma znak gwiazdki zarówno na początku, jak i na końcu.

Przy rozpoznawaniu fraz kluczowych istotny jest podział nazwy opisowej na wyrazy w celu uniknięcia fałszywie pozytywnego rozpoznania dłuższych wyrazów nazwy opisowej, które mogą zawierać frazę (jednowyrazową), np. przy poszukiwaniu frazy „TONE” mogłyby fałszywie rozpoznane nazwy „GEM STONE” i „WHITE TOUCHTONE TELEPHONE”.

Z drugiej strony jednak frazy mogą być kilkuwyrazowe. Mogą też zawierać znaki ‘-’. Dlatego podział nazwy opisowej na wyrazy musi uwzględniać takie frazy złożone.

Trzecim problemem do rozwiązania jest możliwość zawierania jednej frazy w innej. Np. pismo arabskie jest rozpoznawane po frazie „ARABIC”, ale już fraza „ARABIC-INDIC” oznacza notację arabsko-indyjską.

## Rozpoznawanie systemów pisma

Bardzo często opisowa nazwa zawiera frazę (słowo lub ciąg słów) kluczową oznaczającą nazwę systemu pisma, np. „LATIN CAPITAL LETTER A”, „CYRILLIC CAPITAL LETTER A”. Zazwyczaj fraza ta jest umieszczona na początku nazwy opisowej znaku (jak w ostatnich przykładach), ale są wyjątki:

1. Nazwa systemu pisma może się w ogóle nie pojawiać w nazwie opisowej znaku. Przykłady: „FEMININE ORDINAL INDICATOR”, „MASCULINE ORDINAL INDICATOR”, „TURNED CAPITAL F”.
2. Nazwa systemu pisma może być poprzedzona modyfikatorem formy znaku, np. „SUPERSCRIPT LATIN SMALL LETTER N”, „PARENTHESIZED LATIN SMALL LETTER A”, „MODIFIER LETTER SMALL GREEK GAMMA”.
3. Słowo, które zazwyczaj oznacza nazwę systemu pisma, może być po prostu przymiotnikiem, np. „WHITE LATIN CROSS” oznacza „biały łaciński krzyż”.

W bazie danych UCD rozpoznano w sumie 163 pisma (ang. scripts) pokazanych w tab. 3. W pierwszej kolumnie podano oficjalną nazwę pisma, w drugiej – rodzaj pisma, w trzeciej – frazę kluczową, a w czwartej 4-literową nazwę skróconą zgodną ze standardem ISO 15924 i opcjonalny alias. W piątej kolumnie pokazano zakresy znaków Unicode, w których znajdują się znaki przypisane do danego pisma. Zakresy te są określone przez jeden lub wiele bloków Unicode.

Tab. . Rozpoznawane pisma Unicode wraz z frazami kluczowymi

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Pismo | Rodzaj | Fraza kluczowa | Skrót | RTL | Bikam. |
| Adlam | alphabet | ADLAM\* | adlm | ☑ | ☑ |
| Ahom | abugida | AHOM\* | ahom | ☐ | ☐ |
| Anatolian Hieroglyphs | logographic | ANATOLIAN HIEROGLYPH\* | hluw | ☐ | ☐ |
| Arabic | abjad | ARABIC\* | arab | ☑ | ☐ |
| Armenian | alphabet | ARMENIAN\* | armn | ☐ | ☑ |
| Avestan | alphabet | AVESTAN\* | avst | ☑ | ☐ |
| Balinese | abugida | BALINESE\* | bali | ☐ | ☐ |
| Bamum | pictographic | BAMUM\* | bamu | ☐ | ☐ |
| Bassa Vah | alphabet | BASSA VAH\* | bass | ☐ | ☐ |
| Batak | abugida | BATAK\* | batk | ☐ | ☐ |
| Bengali | abugida | BENGALI\* | beng | ☐ | ☐ |
| Bhaiksuki | abugida | BHAIKSUKI\* | bhks | ☐ | ☐ |
| Bopomofo | semi-syllabary | BOPOMOFO\* | bopo | ☐ | ☐ |
| Brahmi | abugida | BRAHMI\* | brah | ☐ | ☐ |
| Buginese (Lontara) | abugida | BUGINESE\* | bugi | ☐ | ☐ |
| Buhid | abugida | BUHID\* | buhd | ☐ | ☐ |
| Canadian Aboriginal | abugida | CANADIAN SYLLABICS\* | cans | ☐ | ☐ |
| Carian | alphabet | CARIAN\* | cari | ☐ | ☐ |
| Caucasian Albanian | alphabet | CAUCASIAN ALBANIAN\* | aghb | ☐ | ☐ |
| Chakma | abugida | CHAKMA\* | cakm | ☐ | ☐ |
| Cham | abugida | CHAM\* | cham | ☐ | ☐ |
| Cherokee | abugida | CHEROKEE\* | cher | ☐ | ☑ |
| Chorasmian | abjad | CHORASMIAN\* | chrs | ☑ | ☐ |
| CJK Compatibility | logo-syllabary | CJK\* | cjkc | ☐ | ☐ |
| CJK Compatibility Ideographs | ideograph | CJK COMPATIBILITY IDEOGRAPH\* | cjki | ☐ | ☐ |
| CJK Radicals | ideograph | CJK RADICAL\* | cjkr | ☐ | ☐ |
| CJK Strokes | syllabary | CJK STROKE\* | cjks | ☐ | ☐ |
| CJK Unified Ideographs | ideograph | \*CJK UNIFIED IDEOGRAPH\* | cjku | ☐ | ☐ |
| Clear Script | alphabet | MONGOLIAN \* TODO \* | todo | ☑ | ☐ |
| Coptic | alphabet | COPTIC\* | copt | ☐ | ☑ |
| Cuneiform | cuneiform | CUNEIFORM\* | xsux  (cune) | ☐ | ☐ |
| Cypriot | syllabary | CYPRIOT\* | cprt | ☐ | ☐ |
| Cypro-Minoan | syllabary | CYPRO-MINOAN\* | cpmn | ☐ | ☐ |
| Cyrillic | alphabet | CYRILLIC\* | cyrl | ☐ | ☑ |
| Deseret | alphabet | DESERET\* | dsrt | ☐ | ☑ |
| Devanagari | abugida | DEVANAGARI\* | deva | ☐ | ☐ |
| Dives Akuru | abugida | DIVES AKURU\* | diak | ☐ | ☐ |
| Dogra | abugida | DOGRA\* | dogr | ☐ | ☐ |
| Egyptian Hieroglyphs | hieroglyphic | EGYPTIAN HIEROGLYPH\* | egyp | ☐ | ☐ |
| Elbasan | alphabet | ELBASAN\* | elba | ☐ | ☐ |
| Elymaic | abjad | ELYMAIC\* | elym | ☑ | ☐ |
| Ethiopic | abugida | ETHIOPIC\* | ethi | ☐ | ☐ |
| Galik | alphabet | MONGOLIAN \* ALI GALI\* | gali | ☑ | ☐ |
| Garay | alphabet | GARAY\* | gara | ☑ | ☑ |
| Georgian | alphabet | GEORGIAN\* | geor | ☐ | ☑ |
| Glagolitic | alphabet | GLAGOLITIC\* | glag | ☐ | ☑ |
| Gothic | alphabet | GOTHIC\* | goth | ☐ | ☐ |
| Grantha | abugida | GRANTHA\* | gran | ☐ | ☐ |
| Greek | alphabet | GREEK\* | grek  (greek) | ☐ | ☑ |
| Gujarati | abugida | GUJARATI\* | gujr | ☐ | ☐ |
| Gunjala Gondi | abugida | GUNJALA GONDI\* | gong | ☐ | ☐ |
| Gurmukhi | abugida | GURMUKHI\* | guru | ☐ | ☐ |
| Gurung Khema | abugida | GURUNG KHEMA\* | gukh | ☐ | ☐ |
| Hangul | alphabet | HANGUL\* | hang | ☐ | ☐ |
| Hanifi Rohingya | alphabet | HANIFI ROHINGYA\* | rohg | ☑ | ☐ |
| Hanunoo | abugida | HANUNOO\* | hano | ☐ | ☐ |
| Hatran | abjad | HATRAN\* | hatr | ☑ | ☐ |
| Hebrew | abjad | HEBREW\* | hebr | ☑ | ☐ |
| Hentaigana | syllabary | HENTAIGANA\* | hent | ☐ | ☐ |
| Hiragana | syllabary | HIRAGANA\* | hira | ☐ | ☐ |
| Imperial Aramaic | abjad | IMPERIAL ARAMAIC\* | armi | ☑ | ☐ |
| Inscriptional Pahlavi | abjad | INSCRIPTIONAL PAHLAVI\* | phli | ☑ | ☐ |
| Inscriptional Parthian | abjad | INSCRIPTIONAL PARTHIAN\* | prti | ☑ | ☐ |
| Javanese | abugida | JAVANESE\* | java | ☐ | ☐ |
| Kaithi | abugida | KAITHI\* | kthi | ☐ | ☐ |
| Kannada | abugida | KANNADA\* | knda | ☐ | ☐ |
| Katakana | syllabary | KATAKANA\* | kana | ☐ | ☐ |
| Kawi | abugida | KAWI\* | kawi | ☐ | ☐ |
| Kayah Li | alphabet | KAYAH LI\* | kali | ☐ | ☐ |
| Kharoshthi | abugida | KHAROSHTHI\* | khar | ☑ | ☐ |
| Khitan Small | logographic | KHITAN SMALL SCRIPT\* | kits | ☐ | ☐ |
| Khmer | abugida | KHMER\* | khmr | ☐ | ☐ |
| Khojki | abugida | KHOJKI\* | khoj | ☐ | ☐ |
| Khudawadi | abugida | KHUDAWADI\* | sind | ☐ | ☐ |
| Kirat Rai | abugida | KIRAT RAI\* | krai | ☐ | ☐ |
| Lao | abugida | LAO\* | laoo  (lao) | ☐ | ☐ |
| Latin | alphabet | LATIN\* | latn  (latin) | ☐ | ☐ |
| Lepcha | abugida | LEPCHA\* | lepc | ☐ | ☐ |
| Limbu | abugida | LIMBU\* | limb | ☐ | ☐ |
| Linear A | logographic | LINEAR A\* | lina | ☐ | ☐ |
| Linear B | syllabary | LINEAR B\* | linb | ☐ | ☐ |
| Lisu | abugida | LISU\* | lisu | ☐ | ☐ |
| Lycian | alphabet | LYCIAN\* | lyci | ☐ | ☐ |
| Lydian | alphabet | LYDIAN\* | lydi | ☐ | ☐ |
| Mahajani | abugida | MAHAJANI\* | mahj | ☐ | ☐ |
| Makasar | abugida | MAKASAR\* | maka | ☐ | ☐ |
| Malayalam | abugida | MALAYALAM\* | mlym | ☐ | ☐ |
| Mandaic | alphabet | MANDAIC\* | mand | ☑ | ☐ |
| Manichaean | abjad | MANICHAEAN\* | mani | ☑ | ☐ |
| Marchen | abugida | MARCHEN\* | marc | ☐ | ☐ |
| Masaram Gondi | abugida | MASARAM GONDI\* | gonm | ☐ | ☐ |
| Medefaidrin | constructed | MEDEFAIDRIN\* | medf | ☐ | ☑ |
| Meetei Mayek | abugida | MEETEI MAYEK\* | mtei | ☐ | ☐ |
| Mende Kikakui | syllabary | MENDE KIKAKUI\* | mend | ☑ | ☐ |
| Meroitic Cursive | abugida | MEROITIC CURSIVE\* | merc | ☐ | ☐ |
| Meroitic Hieroglyphs | hieroglyphic | MEROITIC HIEROGLYPHIC\* | mero | ☑ | ☐ |
| Miao | abugida | MIAO\* | plrd (miao) | ☐ | ☐ |
| Modi | abugida | MODI\* | modi | ☐ | ☐ |
| Mongolian | alphabet | MONGOLIAN\* | mong | ☐ | ☐ |
| Mro (Mru) | alphabet | MRO\* | mroo (mro) | ☐ | ☐ |
| Multani | abugida | MULTANI\* | mult | ☐ | ☐ |
| Myanmar (Burmese) | abugida | MYANMAR\* | mymr | ☐ | ☐ |
| Nabataean | abjad | NABATAEAN\* | nbat | ☑ | ☐ |
| Nag Mundari | alphabet | NAG MUNDARI\* | nagm | ☐ | ☐ |
| Nandinagari | abugida | NANDINAGARI\* | nand | ☐ | ☐ |
| New Tai Lue | abugida | NEW TAI LUE\* | talu | ☐ | ☐ |
| Newa Prachalit | abugida | NEWA\* | newa | ☐ | ☐ |
| Nko | alphabet | NKO\* | nkoo  (nko) | ☑ | ☐ |
| Nushu | syllabary | NUSHU\* | nshu | ☐ | ☐ |
| Nyiakeng Puachue Hmong | alphabet | NYIAKENG PUACHUE HMONG\* | hmnp | ☐ | ☐ |
| Ogham | alphabet | OGHAM\* | ogam | ☐ | ☐ |
| Ol Chiki | alphabet | OL CHIKI\* | olck | ☐ | ☐ |
| Ol Onal | alphabet | OL ONAL\* | onao | ☐ | ☐ |
| Old Hungarian | alphabet | OLD HUNGARIAN\* | hung | ☐ | ☐ |
| Old Permic | alphabet | OLD PERMIC\* | perm | ☐ | ☐ |
| Old Persian | cuneiform | OLD PERSIAN\* | xpeo | ☐ | ☐ |
| Old Sogdian | abjad | OLD SOGDIAN\* | sogo | ☑ | ☐ |
| Old South Arabian | abjad | OLD SOUTH ARABIAN\* | sarb | ☑ | ☐ |
| Old Turkic | alphabet | OLD TURKIC\* | orkh | ☑ | ☐ |
| Old Uyghur | abjad | OLD UYGHUR\* | ougr | ☑ | ☐ |
| Oriya | abugida | ORIYA\* | orya | ☐ | ☐ |
| Osage | alphabet | OSAGE\* | osge | ☐ | ☑ |
| Osmanya | alphabet | OSMANYA\* | osma | ☐ | ☐ |
| Pahawh Hmong | abugida | PAHAWH HMONG\* | hmng | ☐ | ☐ |
| Palmyrene | alphabet | PALMYRENE\* | palm | ☐ | ☐ |
| Pau Cin Hau | alphabet | PAU CIN HAU\* | pauc | ☐ | ☐ |
| Phags-Pa | alphabet | PHAGS-PA\* | phag | ☐ | ☐ |
| Phoenician | abjad | PHOENICIAN\* | phnx | ☑ | ☐ |
| Psalter Pahlavi | abjad | PSALTER PAHLAVI\* | phlp | ☑ | ☐ |
| Rejang | abugida | REJANG\* | rjng | ☐ | ☐ |
| Runic | alphabet | RUNIC\* | runr | ☐ | ☐ |
| Samaritan | abjad | SAMARITAN\* | samr | ☐ | ☐ |
| Saurashtra | abugida | SAURASHTRA\* | saur | ☐ | ☐ |
| Sharada | abugida | SHARADA\* | shrd | ☐ | ☐ |
| Shavian | alphabet | SHAVIAN\* | shaw | ☐ | ☐ |
| Siddham | abugida | SIDDHAM\* | sidd | ☐ | ☐ |
| Sinhala | abugida | SINHALA\* | sinh | ☐ | ☐ |
| Sogdian | abjad | SOGDIAN\* | sogd | ☑ | ☐ |
| Sora Sompeng | alphabet | SORA SOMPENG\* | sora | ☐ | ☐ |
| Soyombo | abugida | SOYOMBO\* | soyo | ☐ | ☐ |
| Sundanese | abugida | SUNDANESE\* | sund | ☐ | ☐ |
| Sunuwar | alphabet | SUNUWAR\* | sunu | ☐ | ☐ |
| Syloti Nagri | abugida | SYLOTI NAGRI\* | sylo | ☐ | ☐ |
| Syriac | abjad | SYRIAC\* | syrc | ☑ | ☐ |
| Tagalog | abugida | TAGALOG\* | tglg | ☐ | ☐ |
| Tagbanwa | abugida | TAGBANWA\* | tagb | ☐ | ☐ |
| Tai Le | abugida | TAI LE\* | tale | ☐ | ☐ |
| Tai Tham | abugida | TAI THAM\* | lana | ☐ | ☐ |
| Tai Viet | abugida | TAI VIET\* | tavt | ☐ | ☐ |
| Takri | abugida | TAKRI\* | takr | ☐ | ☐ |
| Tamil | abugida | TAMIL\* | taml | ☐ | ☐ |
| Tangsa | alphabet | TANGSA\* | tnsa | ☐ | ☐ |
| Tangut | logographic | TANGUT\* | tang | ☐ | ☐ |
| Telugu | abugida | TELUGU\* | telu | ☐ | ☐ |
| Thaana | abugida | THAANA\* | thaa | ☑ | ☐ |
| Thai | abugida | THAI\* | thai | ☐ | ☐ |
| Tibetan | abugida | TIBETAN\* | tibt | ☐ | ☐ |
| Tifinagh | abjad | TIFINAGH\* | tfng | ☐ | ☐ |
| Tirhuta | abugida | TIRHUTA\* | tirh | ☐ | ☐ |
| Todhri | alphabet | TODHRI\* | todr | ☐ | ☐ |
| Toto | alphabet | TOTO\* | toto | ☐ | ☐ |
| Tulu-Tigalari | abugida | TULU-TIGALARI\* | tutg | ☐ | ☐ |
| Ugaritic | abjad | UGARITIC\* | ugar | ☐ | ☐ |
| Vai | syllabary | VAI\* | vaii | ☐ | ☐ |
| Vithkuqi | alphabet | VITHKUQI\* | vith | ☐ | ☑ |
| Wancho | alphabet | WANCHO\* | wcho | ☐ | ☐ |
| Warang Citi | alphabet | WARANG CITI\* | wara | ☐ | ☑ |
| Yezidi | alphabet | YEZIDI\* | yezi | ☑ | ☐ |
| Yi Syllable | abugida | YI SYLLABLE\* | yiii (yi) | ☐ | ☐ |
| Zanabazar Square | abugida | ZANABAZAR SQUARE\* | zanb | ☐ | ☐ |

Standardowe, 4-literowe nazw skrócone pisma nie zawsze są wygodne w stosowaniu. dlatego w niektórych przypadkach zapewniono aliasy do nich:

* Dla przykładu pismo greckie ma 4-literową nazwę standardową „grek”, a łacińskie „latn”. Dodanie po jeszcze jednej literze ułatwiłoby pisanie tych nazw przez człowieka. Dlatego dodano do nich aliasy (odpowiednio „greek”, i „latin”).
* Standardowa nazwa pisma klinowego „xsux” jest co prawda zgodna z regułami ISO, które nakazują nazywać języki i pisma wymarłe od litery ‘x”, ale mało czytelna. Dlatego do pisma klinowego dodano alias „cune”.
* Pismo o standardowej nazwie „prld” to pismo opracowane przez Sama Pollarda, misjonarza – metodysty dla języka A-Hmao, jednego z rodziny języków Miao stosowanych w południowo-wschodniej Azji. To pismo jest znane pod nazwami Pollard Miao lub Miao, więc dodano do niego alias „miao”.
* Kilka pism ma nazwy standardowe dłuższe niż ich oficjalne nazwy. Są to pisma: Lao, Mro, Nko, Yi. Dla tych pism oprócz standardowych nazw „laoo”, „mroo”, „nkoo” i „yiii” zdefiniowano aliasy, odpowiednio: „lao”, „mro”, „nko” i „yi”.

UWAGA: Przy rozpoznawaniu systemów pisma trzeba wziąć pod uwagę, że niektóre notacje mogą być powiązane z pismami. Dla przykładu znak o nazwie ARABIC-INDIC DIGIT ZERO należy zarówno do arabsko-indyjskiej notacji, jak i pisma arabskiego (a dodatkowo należy też do zbioru symboli liczbowych i podzbioru cyfr).

## Rodzaje pism

Pisma można podzielić na kilka rodzajów (tab. 4). Rodzaje pism nie są jawnie identyfikowane w standardzie Unicode, lecz opisane w Wikipedii. Mogą mieć znaczenie przy parsowaniu opisów punktów kodowych.

Tab. . Rodzaje pism

|  |  |
| --- | --- |
| Rodzaj pisma | Opis |
| alfabet  (alphabet) | System pisma oparty na standardowym zestawie liter zapisywanych w celu reprezentowania określonych dźwięków w języku mówionym. |
| sylabariusz (syllabary) | system pisma, w którym pojedyncze znaki reprezentują całe sylaby, a nie pojedyncze głoski. |
| spółgłoskowe (abjad) | System pisma, w którym reprezentowane są tylko spółgłoski, pozostawiające dźwięki samogłosek do wywnioskowania przez czytelnika. |
| alfabet sylabiczny (abugida) | Rodzaj alfabetu, w którym jeden znak reprezentuje zasadniczo całą sylabę wraz z „domyślną” samogłoską. Samogłoska domyślna najczęściej jest dość neutralna, typu szwa, jednak w zależności od języka może mieć różne zabarwienie. Podstawą znaku jest najczęściej spółgłoska, a występujące w sylabie samogłoski inne niż domyślna zapisuje się za pomocą dodania różnorodnych znaków diakrytycznych, choć istnieją również odrębne znaki do zapisu samogłosek. |
| semi-sylabiczne  (semi-syllabary) | System pisma, który zachowuje się częściowo jak alfabet, a częściowo jak sylabariusz. |
| sztuczne  (constructed) | Pismo opracowane na potrzeby języka, którego fonologia, gramatyka, ortografia i słownictwo nie rozwinęły się w sposób naturalny, ale zostały świadomie wymyślone w jakimś celu |
| klinowe (cuneiform) | Logo-sylabiczny system pisma, który był używany do zapisu kilku języków starożytnego Bliskiego Wschodu, nazwany tak od charakterystycznych odcisków w kształcie klina, które tworzą ich znaki. |
| hieroglificzne (hieroglyphic) | Starożytny system pisma, który łączy elementy ideograficzne, logograficzne, sylabiczne i alfabetyczne z ponad 1000 różnych znaków. |
| ideograficzne (ideographic) | System pisma składający się z symboli, które reprezentują idee lub koncepcje niezależne od konkretnego języka. |
| logograficzne (logographic) | System pisma, w którym każda jednostka reprezentuje semantyczny składnik języka, taki jak słowo lub morfem. |
| piktograficzne (pictographic) | System pisma składający się z zestawu symboli graficznych, który przekazuje znaczenie poprzez wizualne podobieństwo do obiektów fizycznych. |
| piktofonetyczne  (pictophonetic) | System pisma zawierający jednostki składające się z elementów ideograficznych i fonetycznych. |
| mieszane (mixed) | System pisma, który ma mieszane cechy pisma logograficznego, sylabariusza i notacji fonetycznej. |

Szczególne rodzaje pism to:

* pisma RTL (Right-To-Left) i dwukierunkowe (bidirectional) – większość pism jest czytanych wierszami od góry do dołu, a znaki w wierszach od lewej do prawej (LTR). Część ma znaki czytane od prawej do lewej (RTL). Spośród nich jest też wiele takich, których zwykły tekst jest czytany od prawej do lewej, ale niektóre wyjątki (np. liczby) od lewej do prawej. Pisma zawierające zarówno tekst RTL, jak i LTR są zwane dwukierunkowymi.
* pisma bikameralne (bicameral) – są to alfabety, które mają inne znaki dla liter wielkich (capital letters) i małych (small letters). Można je rozpoznać po tym, że mają niepuste pola UpperCase i LowerCase w specyfikacji punktów kodowych

Pisma RTL (dwukierunkowe) można rozpoznać wg pola Bidir specyfikacji punktów kodowych. Pole to zawiera jedno-, dwu- lub trójliterowe określenie dwukierunkowości znaku (tab. 5). Do rozpoznania pisma RTL nadają się znaki o kierunkowości R, AL i AN. Jeśli w danym piśmie stosuje się takie znaki, to pismo jest RTL. Jeśli nie – to jest pismem LTR. Kierunkowość L, EN, ES i ET jest stosowana w pismach LTR oraz we fragmentach LTR osadzonych w pismach RTL. Znaki o kierunkowości CS, NSM, BN, WS, ON, B i S są stosowane w obu rodzajach pisma. Kierunkowość LTR, LRO, RLE, RLO, PDF, LRI, RLI, FSI i PDI jest deklarowana tylko dla pojedynczych znaków Unicode sterujących formatowaniem dwukierunkowym (tab. 6).

Tab. . Rodzaje dwukierunkowości znaku

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Bidir | Rodzaj znaku | Opis | LTR | RTL |
| L | Left-to-Right | znaki wyświetlane od lewej do prawej | ☑ | ☐ |
| R | Right-to-Left | znaki wyświetlane od prawej do lewej | ☐ | ☑ |
| AL | Arabic Right-to-Left | znaki wyświetlane od prawej do lewej w alfabecie arabskim i podobnych | ☐ | ☑ |
| EN | European Number | cyfry europejskie | ☑ | ☐ |
| ES | European Number Separator | europejskie separatory liczbowe (plus, minus) | ☑ | ☐ |
| ET | European Number Terminator | europejskie terminatory liczbowe (znak hash, symbole waluty, itp.) | ☑ | ☐ |
| AN | Arabic Number | cyfry arabsko-indyjskie | ☐ | ☑ |
| CS | Common Number Separator | wspólne separatory liczbowe | ☑ | ☑ |
| NSM | Nonspacing mark | znaki diakrytyczne bez szerokości | ☑ | ☑ |
| BN | Boundary Neutral | znaki kontrole i formatujące | ☑ | ☑ |
| WS | Whitespace | spacje | ☑ | ☑ |
| ON | Other Neutrals | inne znaki neutralne | ☑ | ☑ |
| B | Paragraph Separator | separator akapitów | ☑ | ☑ |
| S | Segment Separator | separator linii | ☑ | ☑ |
| LRE | Left-to-Right Embedding | osadzone formatowanie od lewej do prawej | ☐ | ☐ |
| LRO | Left-To-Rigth Override | przesłonięte formatowanie od lewej do prawej | ☐ | ☐ |
| RLE | Right-to-Left Embedding | osadzone formatowanie od prawej do lewej | ☐ | ☐ |
| RLO | Right-To-Left Override | przesłonięte formatowanie od prawej do lewej | ☐ | ☐ |
| PDF | Pop Directional Format | przywrócenie poprzedniego formatowania kierunkowego | ☐ | ☐ |
| LRI | Left-to-Right Isolate | izolowane formatowanie od lewej do prawej | ☐ | ☐ |
| RLI | Right-to-Left Isolate | izolowane formatowanie od prawej do lewej | ☐ | ☐ |
| FSI | First Strong Isolate | izolowane formatowanie pierwszego od lewej do prawej | ☐ | ☐ |
| PDI | Pop Directional Isolate | przywrócenie kierunkowości izolowanej | ☐ | ☐ |

## Obsługa dwukierunkowości

W standardzie Unicode znaki są uporządkowane "logicznie", tj. w kolejności, w jakiej mają być interpretowane, a nie "wizualnie" (w kolejności, w której są prezentowane). Wszystkie znaki Unicode dzielone są pod względem obsługi dwukierunkowości na cztery kategorie:

* Znaki o dwukierunkowości silnej – to te, które mają określony kierunek. Przykładami tego typu znaków są: większość znaków alfabetycznych, znaków sylabicznych, ideogramów Han, cyfry spoza języków europejskich lub arabskich oraz znaki interpunkcyjne, które są specyficzne tylko dla tych skryptów.
* Znaki o dwukierunkowości słabej – to te, które mają niejasny kierunek. Przykładami tego typu znaków są cyfry europejskie, cyfry wschodnioarabsko-indyjskie, symbole arytmetyczne i symbole walut.
* Znaki o kierunkowości neutralnej nie pozwalają określić kierunku określić bez kontekstu. Przykłady obejmują separatory akapitów, tabulatory i większość innych białych znaków. Do tej kategorii należą również symbole interpunkcyjne, które są wspólne dla wielu skryptów, takie jak dwukropek, przecinek, kropka i spacja nieprzerywalna.
* Znaki jawnego formatowania, nazywane również "kierunkowymi znakami formatowania", to specjalne sekwencje Unicode, które kierują algorytmem do modyfikowania jego domyślnego zachowania.

Znaki jawnego formatowania dzielą się na "znaczniki", znaki "osadzania", "izolaty" i znaki "przesłonięcia". Ich efekty trwają do momentu wystąpienia separatora akapitu lub znaku "pop".

* Znaczniki - jeśli po "słabym" znaku następuje kolejny "słaby" znak, algorytm przygląda się pierwszemu sąsiedniemu "silnemu" znakowi. Czasami prowadzi to do niezamierzonych błędów wyświetlania. Błędy te są poprawiane lub eliminowane za pomocą znaków "pseudo-silnych". Są dwa takie znaczniki: Znak o kodzie LEFT-TO-RIGHT MARK (LRM) oraz RIGHT-TO-LEFT MARK (RLM). Są one wstawiane w miejscu, w którym zamknięty słaby znak odziedziczy swój kierunek pisania.
* Osadzanie kierunkowych znaków formatowania jest klasyczną metodą Unicode jawnego formatowania i od wersji Unicode 6.3 jest odradzane na rzecz "izolatów". "Osadzanie" sygnalizuje, że fragment tekstu ma być traktowany jako kierunkowo odrębny. Tekst znajdujący się w zakresie osadzanych znaków formatowania nie jest niezależny od otaczającego go tekstu. Ponadto znaki w osadzeniu mogą wpływać na kolejność znaków na zewnątrz. Unicode 6.3 uznał, że osadzenia kierunkowe zwykle mają zbyt silny wpływ na otoczenie i dlatego są niepotrzebnie trudne w użyciu.
* Znaki izolowanego formatowania kierunkowego sygnalizują, że fragment tekstu ma być traktowany jako kierunkowo odizolowany od otoczenia. Począwszy od Unicode 6.3, są to znaki formatowania, które są zachęcane w nowych dokumentach - gdy platformy docelowe są znane z ich obsługi. Te znaki formatujące zostały wprowadzone po tym, jak stało się jasne, że osadzanie kierunkowe zwykle ma zbyt silny wpływ na ich otoczenie, a tym samym jest niepotrzebnie trudne w użyciu. W przeciwieństwie do starszych znaków formatowania kierunkowego "osadzania", "izolaty" nie mają wpływu na kolejność tekstu poza ich zakresem. Izolaty mogą być zagnieżdżane i mogą być umieszczane w osadzeniach i przesłonięciach.
* Znaki przesłonięcia formatowania kierunkowego dopuszczają szczególne przypadki, takie jak numery części (np. aby wymusić pisanie numeru części składającego się z mieszanego języka angielskiego, cyfr i liter hebrajskich od prawej do lewej) i zaleca się ich unikanie, gdy tylko jest to możliwe. Podobnie jak w przypadku innych znaków formatowania kierunkowego, "przesłonięcia" mogą być zagnieżdżone jeden w drugim oraz w osadzaniu i izolatach.

Tab. . Kody znaków sterujących formatowaniem dwukierunkowym

|  |  |  |  |
| --- | --- | --- | --- |
| Kod | Symbol | Nazwa znaku | Znaczenie |
| 202A | LRE | LEFT-TO-RIGHT EMBEDDING | osadzone formatowanie od lewej do prawej |
| 202D | LRO | LEFT-TO-RIGHT OVERRIDE | przesłonięte formatowanie od lewej do prawej |
| 202B | RLE | RIGHT-TO-LEFT EMBEDDING | osadzone formatowanie od prawej do lewej |
| 202E | RLO | RIGHT-TO-LEFT OVERRIDE | przesłonięte formatowanie od prawej do lewej |
| 202C | PDF | POP DIRECTIONAL FORMATTING | przywrócenie poprzedniego formatowania kierunkowego |
| 2066 | LRI | LEFT-TO-RIGHT ISOLATE | izolowane formatowanie od lewej do prawej |
| 2067 | RLI | RIGHT-TO-LEFT ISOLATE | izolowane formatowanie od prawej do lewej |
| 2068 | FSI | FIRST STRONG ISOLATE | izolowane formatowanie pierwszego od lewej do prawej |
| 2069 | PDI | POP DIRECTIONAL ISOLATE | przywrócenie kierunkowości izolowanej |

## Rozpoznawanie języków

Unicode rzadko odwołuje się do języków bezpośrednio, najczęściej robi to pośrednio – przez systemy pisma. W niektórych przypadkach jednak w opisach znaków znajdują się zarówno frazy kluczowe pism, jak i języków. Przypadki te pokazano w tab. 7.

Tab. . Identyfikowane języki w nazwach znaków Unicode

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Pismo | Język(i) | Fraza kluczowa języka | Skrót języka | Uwagi |
| Cyrillic | byelorussian +ukrainian | BYELORUSSIAN-UKRAINIAN | beu (beukr) | a) |
| ukrainian | UKRAINIAN | uk (ukr) |
| Canadian Aboriginal | Aivilik | AIVILIK | aiv |  |
| Athabascan | ATHAPASCAN | ath |  |
| Blackfoot | BLACKFOOT | bla |  |
| Carrier | CARRIER | crx |  |
| Naskapi | NASKAPI | nsk |  |
| Nattilik | NATTILIK | ntm |  |
| Nunavik | NUNAVIK | iu (iku) |  |
| Nunavut | NUNAVUT | ikt |  |
| Ojibway | OJIBWAY | oji |  |
| Sayisi | SAYISI | chp |  |
| South-slavey | SOUTH-SLAVEY | xsl |  |
| Bible Cree  (macrolanguage) | BIBLE-CREE | cre | b) |
| Moose Cree | MOOSE-CREE | crm |  |
| Northern Cree | N-CREE | crl |  |
| Thopmpson Cree | TH-CREE | crt |  |
| West Cree | WEST-CREE | crk |  |
| Woods Cree | WOODS-CREE | cwd |  |
| Y Cree | Y-CREE | cry | c) |
| Mongolian | Manchu | MANCHU | mnc | d), e) |
| Xibe | SIBE | sjo |
| Galik | ALI GALI | gal |
| Myanmar | Aiton | AITON | aio |  |
| Shan | SHAN | shn |  |
| Arabic | Sindhi | SINDHI | snd |  |

Uwagi:

1. Unicode nie odnosi się do samego języka białoruskiego, ale do dwóch wariantów cyrylicy (oprócz głównego). Jeden wariant jest stosowany w języku ukraińskim, a drugi w ukraińskim i białoruskim
2. Alfabet sylabiczny rdzennych Kanadyjczyków (ang. Canadian Aboriginal Syllabics) ma kilka wariantów stosowanych w poszczególnych językach autochtonicznych Kanady. Wariant Bible Cree odnosi się do makrojęzyka Cree.
3. Nie jest jasne, co oznacza wariant Y-Cree. Nie istnieje oficjalny język o nazwie Y-Cree w standardzie ISO 639-3. Możliwe, że tak jest przywoływany jeden z dialektów języka Cree. Skrót nazwy „cry” został utworzony sztucznie.
4. Oprócz wymienionych tutaj języków mongolskich w standardzie Unicode zdefiniowane są też znaki dla pisma Clear Script i Galik. Wszystkie one mają opisy zaczynające się od „MONGOLIAN”.
5. Niektóre znaki Unicode są wykorzystywane zarówno w języku Manchu, jak i Galik. Składnia opisu jest podobna do „MONGOLIAN LETTER MANCHU ALI GALI \*”. Litery, które należą tylko do języka Galik mają składnię podobną do „MONGOLIAN LETTER ALI GALI \*”, a do Manchu – „MONGOLIAN LETTER MANCHU \*”. Są jeszcze znaki nieliterowe języka Manchu o składni „MONGOLIAN MANCHU \*”.

## Rozpoznawanie systemów notacji

Dalsza analiza przeznaczenia znaków i bloków Unicode umożliwiła rozpoznanie szeregu *notacji,* czyli systemów pisma niepowiązanych (przynajmniej bezpośrednio) z żadnym konkretnym językiem. Są one wymienione w tab. 8. Wyróżnia się wśród nich notacje typu:

* liczbowego (ang. numeric) – do zapisu liczb w różnych systemach,
* matematycznego (ang. mathematical) – do zapisu wzorów i wyrażeń matematycznych,
* muzycznego (ang. musical) – do zapisu muzyki,
* metrycznego (ang. metrical) – do oznaczania rytmu (długości sylab) w poezji (zwłaszcza łacińskiej i greckiej),
* fonetycznego (ang. phonetic) – do zapisu wskazówek dotyczących wymowy,
* stenograficznego (ang. stenographic) – do zapisywania mówionego tekstu w sposób skrócony,
* gestykulacyjnego (ang. gestural) – do zapisywania gestów języka migowego.

Tab. . Notacje Unicode wyróżnione w UCD

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Notacja | Rodzaj | Fraza kluczowa | Skrót | RTL | Uwagi |
| Aegean Numbers | numeric | AEGEAN | AEGN | ☐ | a) |
| APL | symbolic | APL FUNCTIONAL SYMBOL | APL |  | b) |
| Arabic-Indic Numeric Notation | numeric | ARABIC-INDIC | ARIN | ☑ | c) |
| Balinese Musical Symbols | musical | BALINESE MUSICAL SYMBOL | BLMN |  |  |
| Byzantine Musical Symbols | musical | BYZANTINE MUSICAL SYMBOL | BZMN |  |  |
| Coptic Epact Numbers | numeric | COPTIC EPACT | COPN | ☐ |  |
| Counting Rod Numerals | numeric | COUNTING ROD | CNTR | ☐ |  |
| Duployan Shorthand | stenographic | DUPLOYAN | DUPL |  |  |
| Extended Arabic-Indic Numeric Notation | numeric | EXTENDED ARABIC-INDIC | AREN | ☐ | c) |
| Garay Digits | numeric | GARAY DIGIT | GARA | ☑ |  |
| Greek Acrophonic Numbers | numeric | GREEK ACROPHONIC | GRAN | ☐ |  |
| Greek Ancient Instrumental Notation | musical | GREEK INSTRUMENTAL NOTATION | GRIN |  |  |
| Greek Ancient Vocal Notation | musical | GREEK VOCAL NOTATION | GRVN |  |  |
| Greek Metrical System | metrical | METRICAL | GRMS |  |  |
| Greek Musical Symbols | musical | GREEK MUSICAL | GRMN |  |  |
| Hanifi Rohingya Digits | numeric | HANIFI ROHINGYA DIGIT | RIGH | ☑ |  |
| Indic Siyaq Numbers | numeric | INDIC SIYAQ | ISN | ☐ | d) |
| Kaktovik Numerals | numeric | KAKTOVIK NUMERAL | KTVN | ☐ |  |
| Kanbun | phonetic | IDEOGRAPHIC ANNOTATION | KANB |  |  |
| Kievan Musical Symbols | musical | MUSICAL SYMBOL KIEVAN | KVMN |  |  |
| Mathematical Notation | mathematical | MATHEMATICAL | MATH |  | e) |
| Mayan Numerals | numeric | MAYAN NUMERAL | MAYN | ☐ |  |
| Modern Arabic Mathematical | mathematical | ARABIC MATHEMATICAL | AMATH |  |  |
| Musical Notation | musical | MUSICAL SYMBOL | MMN |  |  |
| Ottoman Siyaq Numbers | numeric | OTTOMAN SIYAQ | SIYN |  | f) |
| Persian Musical Symbols | musical | MUSICAL SYMBOL \* | PSMN |  |  |
| Phaistos Disk | symbolic | PHAISTOS DISK | PHSD |  |  |
| Roman Numbers | numeric | ROMAN NUMERAL | ROMN | ☐ |  |
| Rumi Numeral Symbols | numeric | RUMI | RUMI | ☑ |  |
| Signwriting | gesture | SIGNWRITING | SWRT |  |  |
| Suzhou Numbers | numeric | HANGZHOU | SZHN (HZHN) | ☐ |  |
| Z Notation | mathematical | Z NOTATION | ZNOT |  |  |
| Znamenny Chant | musical | ZNAMENNY | ZNMN |  |  |

Uwagi:

1. Notacja Aegean Numbers zawiera nie tylko liczby, lecz też symbole jednostek wagi i innych miar.
2. APL jest notacją używaną w języku opracowanym w 1960 r. przez Kennetha E. Iversona, stosowanym przez IBM jako dodatek do języka Basic.
3. Cyfry wschodnioarabskie arabskie, zwane również cyframi indoarabskimi lub cyframi arabsko-indyjskimi, są symbolami używanymi do reprezentowania cyfr w połączeniu z alfabetem arabskim w krajach Maszrik (wschód świata arabskiego), Półwyspu Arabskiego i jego odmianie w innych krajach, które używają cyfr perskich na płaskowyżu irańskim i w Azji. W Unicode są dwa zakresy cyfr pochodzących z tego zbioru symboli o wartościach od 0 do 9 (te same wartości, ale inne kształty).
4. Indic Siyaq Numbers to podzbiór [pisma arabskiego](https://en.wikipedia.org/wiki/Arabic_script), który był używany w księgowości w [Indiach](https://en.wikipedia.org/wiki/India) pod rządami [Mogołów](https://en.wikipedia.org/wiki/Mughal_Empire) od XVII wieku do połowy XX wieku. Obejmuje on 69 symboli (nie tylko cyfr).
5. Notacja matematyczna (ang. Mathematical Notation) zawiera cyfry dziesiętne symbole (operatory) z kategorii Sm i jest też używana dla liter alfabetu łacińskiego, greckiego i hebrajskiego do zaznaczenia, że są one stosowane w zapisie matematycznym, zazwyczaj do oznaczania zmiennych. Ten zbiór znaków zawiera też podzbiory liter łacińskich pisanych różnymi krojami pisma.
6. Ottoman Siyaq Numbers to podzbiór pisma arabskiego używany do księgowania w dokumentach z czasów Imperium Osmańskiego.
7. Nowoczesna arabska notacja matematyczna (ang. Modern Arabic Mathematical notation) to notacja matematyczna oparta na piśmie arabskim, używana zwłaszcza na poziomie edukacji przed-uniwersyteckiej. Jej forma w większości wywodzi się notacji zachodniej, ale jest zapisywana od prawej do lewej zgodnie z normalnym kierunkiem pisma arabskiego. Inne różnice obejmują zastąpienie liter alfabetu greckiego i łacińskiego dla symboli literami arabskimi oraz używanie nazw arabskich dla funkcji i relacji.
8. Counting Rod Numbers to blok Unicode zawierający tradycyjne chińskie symbole prętów liczących, których matematycy używali do obliczeń w starożytnych Chinach, Japonii, Korei i Wietnamie. Blok zawiera również pięć ideograficznych znaczników liczbowych, które są szeroko stosowane w Azji Wschodniej.
9. System liczbowy Suzhou (oznaczany w Unicode przez frazę kluczową HANGZHOU) to zbiór 9 cyfr dziesiętnych (od 1 do 9) używanych na chińskich targowiskach. Jest jedyną zachowaną odmianą system prętowego. W większości wyparty przez system cyfr hinduskich (indyjsko-arabskich).
10. Roman Numbers to zbiór symbol pochodzących z alfabetu łacińskiego używanych w liczbowej notacji rzymskiej. Mimo iż symbole te wyglądają jak litery łacińskie (z wyjątkiem tylko jednego), to mają osobne kody Unicode.
11. Rumi Numeral Symbols to blok Unicode zawierający znaki numeryczne używane w Fezie, Maroku i innych miejscach w Afryce Północnej i na Półwyspie Iberyjskim między X a XVII wiekiem.
12. System liczbowy Majów był systemem do reprezentowania liczb i dat kalendarzowych w cywilizacji Majów. Był to system liczbowy pozycyjny dwudziestkowy (podstawa 20). Cyfry składają się z trzech symboli: zera (muszli), jedności (kropki) i piątki (kreski). Na przykład trzynaście jest zapisywane jako trzy kropki w poziomym rzędzie nad dwiema poziomymi kreskami; czasami jest również zapisywane jako trzy pionowe kropki na lewo od dwóch pionowych kresek. Za pomocą tych trzech symboli można zapisać każdą z dwudziestu cyfr dwudziestkowych.
13. System liczbowy Kaktovik to system o podstawie 20 stworzony przez alaskańskich Inuitów, używany też w językach Aleutów i ludów Yupik.
14. Coptic Epact Numbers to specjalny system znaków numerycznych używany w niektórych regionach Egiptu w czasach późnego antyku, głównie w języku koptyjskim. Zamiast używać liter alfabetu koptyjskiego do zapisu liczb (jak to robiono w wielu kulturach, np. w Grecji czy Rzymie), system ten wykorzystywał oddzielne znaki dla każdej cyfry, dziesiątki i setki
15. Greckie numeryczne symbole akrofoniczne reprezentują wartości liczbowe używane w starożytnym greckim systemie numeracji, który opierał się na zasadach akrofonicznych – co oznacza, że symbole pochodzą od pierwszej litery odpowiadającego greckiemu słowu dla liczby. Były one używane w różnych wariantach, które w nazwie opisowej znaku Unicode rozróżniane są przez dodatkowe słowa kluczowe (tab. 9). Część symboli wyrażała jednostki monetarne i inne jednostki miary rozróżniane przez słowa kluczowe pokazane w tab. 10.

Tab. . Warianty greckich symboli akrofonicznych

|  |  |  |
| --- | --- | --- |
| Słowo kluczowe | Wariant | Opis |
| ATTIC | attycki | odnosi się do Attyki – starożytnego regionu Grecji obejmującego miasto Ateny i okolice |
| HERAEUM | herański | odnosi się do świątyni bogini Hery |
| THESPIAN | tespiański | związany teatrem lub dramatem (od imienia Thespis, legendarnego aktora z VI wieku p.n.e.) |
| CYRENAIC | cyrenajski | odnosi się do regionie Cyreny, starożytnej greckiej kolonii w północnej Afryce |
| EPIDAUREAN | epidaurejski | odnosi się do Epidauru – starożytnego greckiego miasta w Argolidzie, znanego przede wszystkim jako miejsce kultu boga Asklepiosa, patrona medycyny |
| TROEZENIAN | trojzeński | odnosi się do mieszkańca lub czegoś związanego ze starożytnym miastem Troizena. położonego w regionie Argolidy na Peloponezie w Grecji. |
| CARYSTIAN | karystyjski | odnosi się do Carystos, starożytnego miasta-państwa położonego na południowym wybrzeżu wyspy Eubea w Grecji. |
| NAXIAN | naksoski | odnosi się do wyspy Naksos – największej z Cyklad na Morzu Egejskim |
| DELPHIC | delficki | odnosi się do starożytnego greckiego sanktuarium w Delfach lub do wyroczni delfickiej, czyli kapłanki boga Apollina (Pytii), która udzielała odpowiedzi na pytania pielgrzymów |
| STRATIAN | stratyjski | odnosi się do regionu Stratos, miasta w starożytnej Achai |

Tab. . Jednostki monetarne i inne jednostki miary starożytnej Grecji

|  |  |  |
| --- | --- | --- |
| Słowo kluczowe | Jednostka | Opis |
| DRACHMA(S) | drachma | Starożytna waluta emitowana przez wiele miast-państw w okresie dziesięciu stuleci, od okresu archaicznego, przez okres klasyczny, okres hellenistyczny aż do okresu rzymskiego. Moneta, zwykle wykonana ze srebra, a czasami ze złota, miała swoje początki w systemie barterowym, w którym drachmę odnoszono do garści drewnianych strzał. Drachma była unikalna dla każdego miasta-państwa, które ją biło, i czasami były w obiegu w całym basenie Morza Śródziemnego. Drachma ateńska była uważana za najsilniejszą i stała się najpopularniejszym środkiem płatniczym. |
| QUARTER | jedna czwarta | zapewne jedna czwarta część drachmy |
| TALENT(S) | talent | duża jednostka wagowo-walutową, używana m.in. w kontekście danin, podatków i skarbu państwa. Jeden talent odpowiadał około 26–36 kg srebra. |
| STATER(S) | stater | starożytna srebrna moneta o dużej wartości używana w różnych regionach Grecji |
| MNA(S) | mina | jednostka wagi i waluty. Jedna mina odpowiadała około 100 drachmom lub prawie 0,5 kg srebra. |
| PLETHRON | plethron | starożytna grecka jednostka miary, używana zarówno jako jednostka długości, jak i powierzchni. Jako jednostka długości odpowiadała około 100 stóp greckich, czyli mniej więcej 30 metrów. Jako jednostka powierzchni oznaczała kwadrat o bokach jednego plethronu, czyli około 900 m². |

1. Grecka starożytna notacja muzyczna składa się dwóch osobnych systemów: notacji wokalnej i notacji instrumentalnej. Symbole z tych systemów nie są nazwane, lecz numerowane. Oba systemy mają po 54 symboli.
2. Grecki system metryczny był używany w analizie metrycznej poezji starogreckiej, w której metrum nie opierało się na akcencie, lecz na długości sylab. Każda sylaba mogła być krótka (1 mora) lub długa (2 mory).
3. Kanbun to klasyczny system zapisu i czytania chińskich tekstów literackich w Japonii, stosowany od okresu Nara (VIII w.) aż do XX wieku. Choć teksty były pisane po chińsku, Japończycy opracowali specjalny sposób ich czytania i interpretacji, dzięki któremu przy zachowaniu oryginalnego brzmienia chińskiego można było rozumieć ja na sposób japoński.
4. Unicode obsługuje kilka notacje muzycznych: balijską, bizantyjską, kijowską, perską, współczesną notację muzyczną oraz greckie symbole długości muzycznej – potrójnej (triseme), poczwórnej (tetraseme) i pięciokrotnej (pentaseme) i pauzy (leimma). Greckie symbole długości muzycznej są znakami diakrytycznymi (są kombinowane z innymi symbolami). Perska notacja muzyczna ma tylko dwa symbole, które można rozpoznać przez początkową frazę MUSICAL SYMBOL, która jednak nie rozwija się w żadną inną frazę notacji muzycznych.
5. Notacja Znamenny Chant odnosi się do tradycji śpiewu stosowanej w Rosyjskim Kościele Prawosławnym. Pieśni w tej notacji nie są zapisywane za pomocą nut, lecz specjalnych znaków zwanych proporcami (*znamëna*) lub hakami (*kryuki*).
6. System zapisu znaków (ang. signwriting) opisuje kształty, lecz nie opisuje znaczenia gestów. Dla przykładu SIGNWRITING HAND-FIST INDEX oznacza znak dłoni – pięść z wysuniętym palcem wskazującym.
7. Notacja Duployan Shorthand zawiera symbole stosowane w różnych systemach stenografii, w tym francuskiego Duployéana, pisma Chinook, stenograficznego zapisu rumuńskiego oraz angielskich stenograficznych systemów Sloan- Duployéana, Pernina i Perraulta.

Nie wszystkie notacje zostały w ten sposób wyróżnione. W szczególności notacje fonetyczne IPA (ang. International Phonetic Notation) i UPA (ang. Uralic Phonetic Notation) są oparte odpowiednio o alfabet łaciński i fiński (pismo łacińskie z rozszerzeniami) i nie są odróżniane przez żadną frazę kluczową.

Notacje numeryczne mogą być ukierunkowane LTR lub RTL.

## Typy znaków

Wiele znaków Unicode ma w opisie słowo lub frazę kluczową oznaczającą typ znaku.

|  |  |
| --- | --- |
| Fraza kluczowa | Znaczenie |
| ACCENT | Akcent |
| ARROW | Strzałka |
| ARROWHEAD | Grot strzałki |
| BARB | Asymetryczny grot harpuna |
| CHARACTER | Znak |
| COMBINING | Znak diakrytyczny bez szerokości |
| CONSONANT | Spółgłoska |
| DENOMINATOR | Mianownik |
| DIGIT | Cyfra |
| FRACTION | Ułamek |
| FULL STOP | Kropka (na zakończenie zdania) |
| HARPOON | Harpun (strzałka z grotem asymetrycznym) |
| HIEROGLYPH | Hieroglif |
| IDEOGRAPH | Ideogram |
| LETTER | Litera |
| LIGATURE | Ligatura |
| MARK | Marker |
| MODIFIER | Znak modyfikujący |
| NUMBER | Liczba |
| NUMERAL | Liczebnik |
| NUMERATOR | Licznik |
| POINT | Punkt (dotyczy języka hebrajskiego) |
| PUNCTUATION | Znak interpunkcyjny |
| SEPARATOR | Separator |
| SIGN | Znak |
| SYLLABLE | Sylaba |
| SYMBOL | Symbol |
| TAG | Znacznik |
| TAIL | Zakończenie (zwłaszcza strzałki) |
| TONE | Intonacja |
| VARIATION SELECTOR | Selektor odmiany |
| VOWEL | Samogłoska |

## Znaki diakrytyczne

Znaki diakrytyczne to symbole, które nie mają własnej szerokości i są dodawane ponad lub pod modyfikowanym znakiem (kategoria Mn), a czasami go obejmują (kategoria Me). W nazwach opisowych znaków Unicode mają słowo kluczowe COMBINING, dlatego będą tu nazywane symbolami „kombinatorycznymi”.

Niektóre symbole kombinatoryczne mają swoje identycznie wyglądające odpowiedniki jako osobne punkty kodowe z własną szerokością tak, że mogą być używane jako osobne symbole (tab. 11).

Tab. . Zestawienie symboli bazowych i kombinatorycznych

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Symbol bazowy | | | Symbol kombinatoryczny | | |
| 005E | ^ | CIRCUMFLEX ACCENT | 0302 | ◌̂ | COMBINING CIRCUMFLEX ACCENT |
| 005F | \_ | LOW LINE | 0332 | ◌̲ | COMBINING LOW LINE |
| 0060 | ` | GRAVE ACCENT | 0300 | ◌̀ | COMBINING GRAVE ACCENT |
| 007E | ~ | TILDE | 0303 | ◌̃ | COMBINING TILDE |
| 00A8 | ¨ | DIAERESIS | 0308 | ◌̈ | COMBINING DIAERESIS |
| 00AF | ¯ | MACRON | 0304 | ◌̄ | COMBINING MACRON |
| 00B4 | ´ | ACUTE ACCENT | 0301 | ◌́ | COMBINING ACUTE ACCENT |
| 00B8 | ¸ | CEDILLA | 0327 | ◌̧ | COMBINING CEDILLA |
| 02C7 | ˇ | CARON | 030C | ◌̌ | COMBINING CARON |
| 02D8 | ˘ | BREVE | 0306 | ◌̆ | COMBINING BREVE |
| 02D9 | ˙ | DOT ABOVE | 0307 | ◌̇ | COMBINING DOT ABOVE |
| 02DA | ˚ | RING ABOVE | 030A | ◌̊ | COMBINING RING ABOVE |
| 02DB | ˛ | OGONEK | 0328 | ◌̨ | COMBINING OGONEK |
| 02DD | ˝ | DOUBLE ACUTE ACCENT | 030B | ◌̋ | COMBINING DOUBLE ACUTE ACCENT |
| 037A | ͺ | GREEK YPOGEGRAMMENI | 0345 | ◌ͅ | COMBINING GREEK YPOGEGRAMMENI |
| 0385 | ΅ | GREEK DIALYTIKA TONOS | 0344 | ◌̈́ | COMBINING GREEK DIALYTIKA TONOS |
| 093D | ऽ | DEVANAGARI SIGN AVAGRAHA | A8F1 | ꣱ | COMBINING DEVANAGARI SIGN AVAGRAHA |
| 1FBD | ᾽ | GREEK KORONIS | 0343 | ◌̓ | COMBINING GREEK KORONIS |
| 2017 | ‗ | DOUBLE LOW LINE | 0333 | ◌̳ | COMBINING DOUBLE LOW LINE |
| 203E | ‾ | OVERLINE | 0305 | ◌̅ | COMBINING OVERLINE |
| 2193 | ↓ | DOWNWARDS ARROW | 1AB3 | ◌᪳ | COMBINING DOWNWARDS ARROW |
| 21BD | ↽ | LEFTWARDS HARPOON WITH BARB DOWNWARDS | 20ED | ◌⃭ | COMBINING LEFTWARDS HARPOON WITH BARB DOWNWARDS |
| 21C1 | ⇁ | RIGHTWARDS HARPOON WITH BARB DOWNWARDS | 20EC | ◌⃬ | COMBINING RIGHTWARDS HARPOON WITH BARB DOWNWARDS |
| 221E | ∞ | INFINITY | 1AB2 | ◌᪲ | COMBINING INFINITY |
| 2E2F | ⸯ | VERTICAL TILDE | 033E | ◌̾ | COMBINING VERTICAL TILDE |
| 309B | ゛ | KATAKANA-HIRAGANA VOICED SOUND MARK | 3099 | ◌゙ | COMBINING KATAKANA-HIRAGANA VOICED SOUND MARK |
| 309C | ゜ | KATAKANA-HIRAGANA SEMI-VOICED SOUND MARK | 309A | ◌゚ | COMBINING KATAKANA-HIRAGANA SEMI-VOICED SOUND MARK |
| A67E | ꙾ | CYRILLIC KAVYKA | A67C | ◌꙼ | COMBINING CYRILLIC KAVYKA |
| A67F | ꙿ | CYRILLIC PAYEROK | A67D | ◌꙽ | COMBINING CYRILLIC PAYEROK |

Część symboli bazowych tworzy symbole kombinatoryczne po zmianie położenia (tab. 12).

Tab. . Zestawienie symboli bazowych i kombinatorycznych

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Symbol bazowy | | | Symbol kombinatoryczny | | | Położenie |
| 00B4 | ´ | ACUTE ACCENT | 0317 | ◌̗ | COMBINING ACUTE ACCENT BELOW | BELOW |
| 2248 | ≈ | ALMOST EQUAL TO | 034C | ◌͌ | COMBINING ALMOST EQUAL TO ABOVE | ABOVE |
| 2248 | ≈ | ALMOST EQUAL TO | 1DFD | ◌᷽ | COMBINING ALMOST EQUAL TO BELOW | BELOW |
| 002A | \* | ASTERISK | 0359 | ◌͙ | COMBINING ASTERISK BELOW | BELOW |
| 002A | \* | ASTERISK | 20F0 | ◌⃰ | COMBINING ASTERISK ABOVE | ABOVE |
| 02D8 | ˘ | BREVE | 032E | ◌̮ | COMBINING BREVE BELOW | BELOW |
| 02C7 | ˇ | CARON | 032C | ◌̬ | COMBINING CARON BELOW | BELOW |
| 005E | ^ | CIRCUMFLEX ACCENT | 032D | ◌̭ | COMBINING CIRCUMFLEX ACCENT BELOW | BELOW |
| 002C | , | COMMA | 0313 | ◌̓ | COMBINING COMMA ABOVE | ABOVE |
| 002C | , | COMMA | 0326 | ◌̦ | COMBINING COMMA BELOW | BELOW |
| 00A8 | ¨ | DIAERESIS | 0324 | ◌̤ | COMBINING DIAERESIS BELOW | BELOW |
| 02D9 | ˙ | DOT ABOVE | 0307 | ◌̇ | COMBINING DOT ABOVE | ABOVE |
| 02D9 | ˙ | DOT ABOVE | 0323 | ◌̣ | COMBINING DOT BELOW | BELOW |
| 2016 | ‖ | DOUBLE VERTICAL LINE | 0348 | ◌͈ | COMBINING DOUBLE VERTICAL LINE BELOW | BELOW |
| 2016 | ‖ | DOUBLE VERTICAL LINE | 030E | ◌̎ | COMBINING DOUBLE VERTICAL LINE ABOVE | ABOVE |
| 22A4 | ⊤ | DOWN TACK | 031E | ◌̞ | COMBINING DOWN TACK BELOW | BELOW |
| 003D | = | EQUALS SIGN | 0347 | ◌͇ | COMBINING EQUALS SIGN BELOW | BELOW |
| 0060 | ` | GRAVE ACCENT | 0316 | ◌̖ | COMBINING GRAVE ACCENT BELOW | BELOW |
| 223B | ∻ | HOMOTHETIC | 034B | ◌͋ | COMBINING HOMOTHETIC ABOVE | ABOVE |
| 0072 | r | LATIN SMALL LETTER R | 1DCA | ◌᷊ | COMBINING LATIN SMALL LETTER R BELOW | BELOW |
| 028D | ʍ | LATIN SMALL LETTER TURNED W | 1AC0 | ◌ᫀ | COMBINING LATIN SMALL LETTER TURNED W BELOW | BELOW |
| 0077 | w | LATIN SMALL LETTER W | 1ABF | ◌ᪿ | COMBINING LATIN SMALL LETTER W BELOW | BELOW |
| 2194 | ↔ | LEFT RIGHT ARROW | 20E1 | ◌⃡ | COMBINING LEFT RIGHT ARROW ABOVE | ABOVE |
| 2194 | ↔ | LEFT RIGHT ARROW | 034D | ◌͍ | COMBINING LEFT RIGHT ARROW BELOW | BELOW |
| 22A3 | ⊣ | LEFT TACK | 0318 | ◌̘ | COMBINING LEFT TACK BELOW | BELOW |
| 2190 | ← | LEFTWARDS ARROW | 20EA | ◌⃪ | COMBINING LEFTWARDS ARROW OVERLAY | OVERLAY |
| 2190 | ← | LEFWARDS ARROW | 20D6 | ◌⃖ | COMBINING LEFT ARROW ABOVE | ABOVE |
| 00AF | ¯ | MACRON | 0331 | ◌̱ | COMBINING MACRON BELOW | BELOW |
| 2212 | − | MINUS SIGN | 0320 | ◌̠ | COMBINING MINUS SIGN BELOW | BELOW |
| 00D7 | × | MULTIPLICATION SIGN | 0353 | ◌͓ | COMBINING X BELOW | BELOW |
| 00D7 | × | MULTIPLICATION SIGN | 033D | ◌̽ | COMBINING X ABOVE | ABOVE |
| 2241 | ≁ | NOT TILDE | 034A | ◌͊ | COMBINING NOT TILDE ABOVE | ABOVE |
| 0023 | # | NUMBER SIGN | 1AC6 | ◌᫆ | COMBINING NUMBER SIGN ABOVE | ABOVE |
| 02DB | ˛ | OGONEK | 1DCE | ◌᷎ | COMBINING OGONEK ABOVE | ABOVE |
| 002B | + | PLUS SIGN | 031F | ◌̟ | COMBINING PLUS SIGN BELOW | BELOW |
| 002B | + | PLUS SIGN | 1AC8 | ◌᫈ | COMBINING PLUS SIGN ABOVE | ABOVE |
| 005C | \ | REVERSE SOLIDUS | 20E5 | ◌⃥ | COMBINING REVERSE SOLIDUS OVERLAY | OVERLAY |
| 2E41 | ⹁ | REVERSED COMMA | 0314 | ◌̔ | COMBINING REVERSED COMMA ABOVE | ABOVE |
| 22A2 | ⊢ | RIGHT TACK | 0319 | ◌̙ | COMBINING RIGHT TACK BELOW | BELOW |
| 2192 | → | RIGHTWARDS ARROW | 20D7 | ◌⃗ | COMBINING RIGHT ARROW ABOVE | ABOVE |
| 02DA | ˚ | RING ABOVE | 20D8 | ◌⃘ | COMBINING RING OVERLAY | OVERLAY |
| 02DA | ˚ | RING ABOVE | 030A | ◌̊ | COMBINING RING ABOVE | ABOVE |
| 02DA | ˚ | RING ABOVE | 0325 | ◌̥ | COMBINING RING BELOW | BELOW |
| 007E | ~ | TILDE | 0334 | ◌̴ | COMBINING TILDE OVERLAY | OVERLAY |
| 007E | ~ | TILDE | 0330 | ◌̰ | COMBINING TILDE BELOW | BELOW |
| 2E32 | ⸲ | TURNED COMMA | 0312 | ◌̒ | COMBINING TURNED COMMA ABOVE | ABOVE |
| 22A5 | ⊥ | UP TACK | 031D | ◌̝ | COMBINING UP TACK BELOW | BELOW |
| 22A5 | ⊥ | UP TACK | 1DF5 | ◌᷵ | COMBINING UP TACK ABOVE | ABOVE |
| 2191 | ↑ | UPWARDS ARROW | 034E | ◌͎ | COMBINING UPWARDS ARROW BELOW | BELOW |
| 007C | | | VERTICAL LINE | 030D | ◌̍ | COMBINING VERTICAL LINE ABOVE | ABOVE |
| 007C | | | VERTICAL LINE | 0329 | ◌̩ | COMBINING VERTICAL LINE BELOW | BELOW |

Oprócz tego wiele liter i cyfr ma swoje odpowiedniki w symbolach kombinatorycznych, jak pokazano w tab. 13.

Tab. . Symbole alfanumeryczne i ich odpowiedniki kombinatoryczne

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Litera lub cyfra | | | Symbol kombinatoryczny | | |
| 0061 | a | LATIN SMALL LETTER A | 0363 | ◌ͣ | COMBINING LATIN SMALL LETTER A |
| 0062 | b | LATIN SMALL LETTER B | 1DE8 | ◌ᷨ | COMBINING LATIN SMALL LETTER B |
| 0063 | c | LATIN SMALL LETTER C | 0368 | ◌ͨ | COMBINING LATIN SMALL LETTER C |
| 0064 | d | LATIN SMALL LETTER D | 0369 | ◌ͩ | COMBINING LATIN SMALL LETTER D |
| 0065 | e | LATIN SMALL LETTER E | 0364 | ◌ͤ | COMBINING LATIN SMALL LETTER E |
| 0066 | f | LATIN SMALL LETTER F | 1DEB | ◌ᷫ | COMBINING LATIN SMALL LETTER F |
| 0067 | g | LATIN SMALL LETTER G | 1DDA | ◌ᷚ | COMBINING LATIN SMALL LETTER G |
| 0068 | h | LATIN SMALL LETTER H | 036A | ◌ͪ | COMBINING LATIN SMALL LETTER H |
| 0069 | i | LATIN SMALL LETTER I | 0365 | ◌ͥ | COMBINING LATIN SMALL LETTER I |
| 006B | k | LATIN SMALL LETTER K | 1DDC | ◌ᷜ | COMBINING LATIN SMALL LETTER K |
| 006C | l | LATIN SMALL LETTER L | 1DDD | ◌ᷝ | COMBINING LATIN SMALL LETTER L |
| 006D | m | LATIN SMALL LETTER M | 036B | ◌ͫ | COMBINING LATIN SMALL LETTER M |
| 006E | n | LATIN SMALL LETTER N | 1DE0 | ◌ᷠ | COMBINING LATIN SMALL LETTER N |
| 006F | o | LATIN SMALL LETTER O | 0366 | ◌ͦ | COMBINING LATIN SMALL LETTER O |
| 0070 | p | LATIN SMALL LETTER P | 1DEE | ◌ᷮ | COMBINING LATIN SMALL LETTER P |
| 0072 | r | LATIN SMALL LETTER R | 036C | ◌ͬ | COMBINING LATIN SMALL LETTER R |
| 0073 | s | LATIN SMALL LETTER S | 1DE4 | ◌ᷤ | COMBINING LATIN SMALL LETTER S |
| 0074 | t | LATIN SMALL LETTER T | 036D | ◌ͭ | COMBINING LATIN SMALL LETTER T |
| 0075 | u | LATIN SMALL LETTER U | 0367 | ◌ͧ | COMBINING LATIN SMALL LETTER U |
| 0076 | v | LATIN SMALL LETTER V | 036E | ◌ͮ | COMBINING LATIN SMALL LETTER V |
| 0077 | w | LATIN SMALL LETTER W | 1DF1 | ◌ᷱ | COMBINING LATIN SMALL LETTER W |
| 0078 | x | LATIN SMALL LETTER X | 036F | ◌ͯ | COMBINING LATIN SMALL LETTER X |
| 007A | z | LATIN SMALL LETTER Z | 1DE6 | ◌ᷦ | COMBINING LATIN SMALL LETTER Z |
| 00E4 | ä | LATIN SMALL LETTER A WITH DIAERESIS | 1DF2 | ◌ᷲ | COMBINING LATIN SMALL LETTER A WITH DIAERESIS |
| 00E6 | æ | LATIN SMALL LETTER AE | 1DD4 | ◌ᷔ | COMBINING LATIN SMALL LETTER AE |
| 00F0 | ð | LATIN SMALL LETTER ETH | 1DD9 | ◌ᷙ | COMBINING LATIN SMALL LETTER ETH |
| 00F6 | ö | LATIN SMALL LETTER O WITH DIAERESIS | 1DF3 | ◌ᷳ | COMBINING LATIN SMALL LETTER O WITH DIAERESIS |
| 00FC | ü | LATIN SMALL LETTER U WITH DIAERESIS | 1DF4 | ◌ᷴ | COMBINING LATIN SMALL LETTER U WITH DIAERESIS |
| 017F | ſ | LATIN SMALL LETTER LONG S | 1DE5 | ◌ᷥ | COMBINING LATIN SMALL LETTER LONG S |
| 0251 | ɑ | LATIN SMALL LETTER ALPHA | 1DE7 | ◌ᷧ | COMBINING LATIN SMALL LETTER ALPHA |
| 0259 | ə | LATIN SMALL LETTER SCHWA | 1DEA | ◌ᷪ | COMBINING LATIN SMALL LETTER SCHWA |
| 0262 | ɢ | LATIN LETTER SMALL CAPITAL G | 1DDB | ◌ᷛ | COMBINING LATIN LETTER SMALL CAPITAL G |
| 0274 | ɴ | LATIN LETTER SMALL CAPITAL N | 1DE1 | ◌ᷡ | COMBINING LATIN LETTER SMALL CAPITAL N |
| 0280 | ʀ | LATIN LETTER SMALL CAPITAL R | 1DE2 | ◌ᷢ | COMBINING LATIN LETTER SMALL CAPITAL R |
| 0283 | ʃ | LATIN SMALL LETTER ESH | 1DEF | ◌ᷯ | COMBINING LATIN SMALL LETTER ESH |
| 029F | ʟ | LATIN LETTER SMALL CAPITAL L | 1DDE | ◌ᷞ | COMBINING LATIN LETTER SMALL CAPITAL L |
| 0456 | і | CYRILLIC SMALL LETTER BYELORUSSIAN-UKRAINIAN I | 1E08F | ◌ | COMBINING CYRILLIC SMALL LETTER BYELORUSSIAN-UKRAINIAN I |
| 0905 | अ | DEVANAGARI LETTER A | A8EA | ꣪ | COMBINING DEVANAGARI LETTER A |
| 0909 | उ | DEVANAGARI LETTER U | A8EB | ꣫ | COMBINING DEVANAGARI LETTER U |
| 0915 | क | DEVANAGARI LETTER KA | A8EC | ꣬ | COMBINING DEVANAGARI LETTER KA |
| 0928 | न | DEVANAGARI LETTER NA | A8ED | ꣭ | COMBINING DEVANAGARI LETTER NA |
| 092A | प | DEVANAGARI LETTER PA | A8EE | ꣮ | COMBINING DEVANAGARI LETTER PA |
| 0930 | र | DEVANAGARI LETTER RA | A8EF | ꣯ | COMBINING DEVANAGARI LETTER RA |
| 0966 | ० | DEVANAGARI DIGIT ZERO | A8E0 | ꣠ | COMBINING DEVANAGARI DIGIT ZERO |
| 0967 | १ | DEVANAGARI DIGIT ONE | A8E1 | ꣡ | COMBINING DEVANAGARI DIGIT ONE |
| 0968 | २ | DEVANAGARI DIGIT TWO | A8E2 | ꣢ | COMBINING DEVANAGARI DIGIT TWO |
| 0969 | ३ | DEVANAGARI DIGIT THREE | A8E3 | ꣣ | COMBINING DEVANAGARI DIGIT THREE |
| 096A | ४ | DEVANAGARI DIGIT FOUR | A8E4 | ꣤ | COMBINING DEVANAGARI DIGIT FOUR |
| 096B | ५ | DEVANAGARI DIGIT FIVE | A8E5 | ꣥ | COMBINING DEVANAGARI DIGIT FIVE |
| 096C | ६ | DEVANAGARI DIGIT SIX | A8E6 | ꣦ | COMBINING DEVANAGARI DIGIT SIX |
| 096D | ७ | DEVANAGARI DIGIT SEVEN | A8E7 | ꣧ | COMBINING DEVANAGARI DIGIT SEVEN |
| 096E | ८ | DEVANAGARI DIGIT EIGHT | A8E8 | ꣨ | COMBINING DEVANAGARI DIGIT EIGHT |
| 096F | ९ | DEVANAGARI DIGIT NINE | A8E9 | ꣩ | COMBINING DEVANAGARI DIGIT NINE |
| 1D0D | ᴍ | LATIN LETTER SMALL CAPITAL M | 1DDF | ◌ᷟ | COMBINING LATIN LETTER SMALL CAPITAL M |
| 1D79 | ᵹ | LATIN SMALL LETTER INSULAR G | 1ACC | ◌ᫌ | COMBINING LATIN SMALL LETTER INSULAR G |
| A735 | ꜵ | LATIN SMALL LETTER AO | 1DD5 | ◌ᷕ | COMBINING LATIN SMALL LETTER AO |
| A739 | ꜹ | LATIN SMALL LETTER AV | 1DD6 | ◌ᷖ | COMBINING LATIN SMALL LETTER AV |
| A75B | ꝛ | LATIN SMALL LETTER R ROTUNDA | 1DE3 | ◌ᷣ | COMBINING LATIN SMALL LETTER R ROTUNDA |
| A77A | ꝺ | LATIN SMALL LETTER INSULAR D | 1DD8 | ◌ᷘ | COMBINING LATIN SMALL LETTER INSULAR D |
| A783 | ꞃ | LATIN SMALL LETTER INSULAR R | 1ACD | ◌ᫍ | COMBINING LATIN SMALL LETTER INSULAR R |
| A787 | ꞇ | LATIN SMALL LETTER INSULAR T | 1ACE | ◌ᫎ | COMBINING LATIN SMALL LETTER INSULAR T |
| A7B5 | ꞵ | LATIN SMALL LETTER BETA | 1DE9 | ◌ᷩ | COMBINING LATIN SMALL LETTER BETA |
| AB38 | ꬸ | LATIN SMALL LETTER L WITH DOUBLE MIDDLE TILDE | 1DEC | ◌ᷬ | COMBINING LATIN SMALL LETTER L WITH DOUBLE MIDDLE TILDE |
| 10350 | 𐍐 | OLD PERMIC LETTER AN | 10376 | ◌ | COMBINING OLD PERMIC LETTER AN |
| 10353 | 𐍓 | OLD PERMIC LETTER DOI | 10377 | ◌ | COMBINING OLD PERMIC LETTER DOI |
| 10357 | 𐍗 | OLD PERMIC LETTER ZATA | 10378 | ◌ | COMBINING OLD PERMIC LETTER ZATA |
| 1035D | 𐍝 | OLD PERMIC LETTER NENOE | 10379 | ◌ | COMBINING OLD PERMIC LETTER NENOE |
| 10361 | 𐍡 | OLD PERMIC LETTER SII | 1037A | ◌ | COMBINING OLD PERMIC LETTER SII |
| 11305 | 𑌅 | GRANTHA LETTER A | 11370 | ◌ | COMBINING GRANTHA LETTER A |
| 11315 | 𑌕 | GRANTHA LETTER KA | 11371 | ◌ | COMBINING GRANTHA LETTER KA |
| 11328 | 𑌨 | GRANTHA LETTER NA | 11372 | ◌ | COMBINING GRANTHA LETTER NA |
| 1132A | 𑌪 | GRANTHA LETTER PA | 11374 | ◌ | COMBINING GRANTHA LETTER PA |

Jest wreszcie duża grupa symboli kombinatorycznych, które nie mają swoich odpowiedników w symbolach regularnych (tab. 14).

Tab. . Symbole kombinatoryczne bez symboli bazowych

| Kod | Glif | Symbol kombinatoryczny |
| --- | --- | --- |
| 0300 | ◌̀ | COMBINING GRAVE ACCENT |
| 0301 | ◌́ | COMBINING ACUTE ACCENT |
| 0302 | ◌̂ | COMBINING CIRCUMFLEX ACCENT |
| 0303 | ◌̃ | COMBINING TILDE |
| 0304 | ◌̄ | COMBINING MACRON |
| 0305 | ◌̅ | COMBINING OVERLINE |
| 0306 | ◌̆ | COMBINING BREVE |
| 0307 | ◌̇ | COMBINING DOT ABOVE |
| 0308 | ◌̈ | COMBINING DIAERESIS |
| 0309 | ◌̉ | COMBINING HOOK ABOVE |
| 030A | ◌̊ | COMBINING RING ABOVE |
| 030B | ◌̋ | COMBINING DOUBLE ACUTE ACCENT |
| 030C | ◌̌ | COMBINING CARON |
| 030D | ◌̍ | COMBINING VERTICAL LINE ABOVE |
| 030E | ◌̎ | COMBINING DOUBLE VERTICAL LINE ABOVE |
| 030F | ◌̏ | COMBINING DOUBLE GRAVE ACCENT |
| 0310 | ◌̐ | COMBINING CANDRABINDU |
| 0311 | ◌̑ | COMBINING INVERTED BREVE |
| 0312 | ◌̒ | COMBINING TURNED COMMA ABOVE |
| 0313 | ◌̓ | COMBINING COMMA ABOVE |
| 0314 | ◌̔ | COMBINING REVERSED COMMA ABOVE |
| 0315 | ◌̕ | COMBINING COMMA ABOVE RIGHT |
| 0316 | ◌̖ | COMBINING GRAVE ACCENT BELOW |
| 0317 | ◌̗ | COMBINING ACUTE ACCENT BELOW |
| 0318 | ◌̘ | COMBINING LEFT TACK BELOW |
| 0319 | ◌̙ | COMBINING RIGHT TACK BELOW |
| 031A | ◌̚ | COMBINING LEFT ANGLE ABOVE |
| 031B | ◌̛ | COMBINING HORN |
| 031C | ◌̜ | COMBINING LEFT HALF RING BELOW |
| 031D | ◌̝ | COMBINING UP TACK BELOW |
| 031E | ◌̞ | COMBINING DOWN TACK BELOW |
| 031F | ◌̟ | COMBINING PLUS SIGN BELOW |
| 0320 | ◌̠ | COMBINING MINUS SIGN BELOW |
| 0321 | ◌̡ | COMBINING PALATALIZED HOOK BELOW |
| 0322 | ◌̢ | COMBINING RETROFLEX HOOK BELOW |
| 0323 | ◌̣ | COMBINING DOT BELOW |
| 0324 | ◌̤ | COMBINING DIAERESIS BELOW |
| 0325 | ◌̥ | COMBINING RING BELOW |
| 0326 | ◌̦ | COMBINING COMMA BELOW |
| 0327 | ◌̧ | COMBINING CEDILLA |
| 0328 | ◌̨ | COMBINING OGONEK |
| 0329 | ◌̩ | COMBINING VERTICAL LINE BELOW |
| 032A | ◌̪ | COMBINING BRIDGE BELOW |
| 032B | ◌̫ | COMBINING INVERTED DOUBLE ARCH BELOW |
| 032C | ◌̬ | COMBINING CARON BELOW |
| 032D | ◌̭ | COMBINING CIRCUMFLEX ACCENT BELOW |
| 032E | ◌̮ | COMBINING BREVE BELOW |
| 032F | ◌̯ | COMBINING INVERTED BREVE BELOW |
| 0330 | ◌̰ | COMBINING TILDE BELOW |
| 0331 | ◌̱ | COMBINING MACRON BELOW |
| 0332 | ◌̲ | COMBINING LOW LINE |
| 0333 | ◌̳ | COMBINING DOUBLE LOW LINE |
| 0334 | ◌̴ | COMBINING TILDE OVERLAY |
| 0335 | ◌̵ | COMBINING SHORT STROKE OVERLAY |
| 0336 | ◌̶ | COMBINING LONG STROKE OVERLAY |
| 0337 | ◌̷ | COMBINING SHORT SOLIDUS OVERLAY |
| 0338 | ◌̸ | COMBINING LONG SOLIDUS OVERLAY |
| 0339 | ◌̹ | COMBINING RIGHT HALF RING BELOW |
| 033A | ◌̺ | COMBINING INVERTED BRIDGE BELOW |
| 033B | ◌̻ | COMBINING SQUARE BELOW |
| 033C | ◌̼ | COMBINING SEAGULL BELOW |
| 033D | ◌̽ | COMBINING X ABOVE |
| 033E | ◌̾ | COMBINING VERTICAL TILDE |
| 033F | ◌̿ | COMBINING DOUBLE OVERLINE |
| 0340 | ◌̀ | COMBINING GRAVE TONE MARK |
| 0341 | ◌́ | COMBINING ACUTE TONE MARK |
| 0343 | ◌̓ | COMBINING GREEK KORONIS |
| 0344 | ◌̈́ | COMBINING GREEK DIALYTIKA TONOS |
| 0345 | ◌ͅ | COMBINING GREEK YPOGEGRAMMENI |
| 0346 | ◌͆ | COMBINING BRIDGE ABOVE |
| 0347 | ◌͇ | COMBINING EQUALS SIGN BELOW |
| 0348 | ◌͈ | COMBINING DOUBLE VERTICAL LINE BELOW |
| 0349 | ◌͉ | COMBINING LEFT ANGLE BELOW |
| 034A | ◌͊ | COMBINING NOT TILDE ABOVE |
| 034B | ◌͋ | COMBINING HOMOTHETIC ABOVE |
| 034C | ◌͌ | COMBINING ALMOST EQUAL TO ABOVE |
| 034D | ◌͍ | COMBINING LEFT RIGHT ARROW BELOW |
| 034E | ◌͎ | COMBINING UPWARDS ARROW BELOW |
| 034F | ◌͏ | COMBINING GRAPHEME JOINER |
| 0350 | ◌͐ | COMBINING RIGHT ARROWHEAD ABOVE |
| 0351 | ◌͑ | COMBINING LEFT HALF RING ABOVE |
| 0352 | ◌͒ | COMBINING FERMATA |
| 0353 | ◌͓ | COMBINING X BELOW |
| 0354 | ◌͔ | COMBINING LEFT ARROWHEAD BELOW |
| 0355 | ◌͕ | COMBINING RIGHT ARROWHEAD BELOW |
| 0356 | ◌͖ | COMBINING RIGHT ARROWHEAD AND UP ARROWHEAD BELOW |
| 0357 | ◌͗ | COMBINING RIGHT HALF RING ABOVE |
| 0358 | ◌͘ | COMBINING DOT ABOVE RIGHT |
| 0359 | ◌͙ | COMBINING ASTERISK BELOW |
| 035A | ◌͚ | COMBINING DOUBLE RING BELOW |
| 035B | ◌͛ | COMBINING ZIGZAG ABOVE |
| 035C | ◌͜ | COMBINING DOUBLE BREVE BELOW |
| 035D | ◌͝ | COMBINING DOUBLE BREVE |
| 035E | ◌͞ | COMBINING DOUBLE MACRON |
| 035F | ◌͟ | COMBINING DOUBLE MACRON BELOW |
| 0360 | ◌͠ | COMBINING DOUBLE TILDE |
| 0361 | ◌͡ | COMBINING DOUBLE INVERTED BREVE |
| 0362 | ◌͢ | COMBINING DOUBLE RIGHTWARDS ARROW BELOW |
| 0363 | ◌ͣ | COMBINING LATIN SMALL LETTER A |
| 0364 | ◌ͤ | COMBINING LATIN SMALL LETTER E |
| 0365 | ◌ͥ | COMBINING LATIN SMALL LETTER I |
| 0366 | ◌ͦ | COMBINING LATIN SMALL LETTER O |
| 0367 | ◌ͧ | COMBINING LATIN SMALL LETTER U |
| 0368 | ◌ͨ | COMBINING LATIN SMALL LETTER C |
| 0369 | ◌ͩ | COMBINING LATIN SMALL LETTER D |
| 036A | ◌ͪ | COMBINING LATIN SMALL LETTER H |
| 036B | ◌ͫ | COMBINING LATIN SMALL LETTER M |
| 036C | ◌ͬ | COMBINING LATIN SMALL LETTER R |
| 036D | ◌ͭ | COMBINING LATIN SMALL LETTER T |
| 036E | ◌ͮ | COMBINING LATIN SMALL LETTER V |
| 036F | ◌ͯ | COMBINING LATIN SMALL LETTER X |
| 0483 | ◌҃ | COMBINING CYRILLIC TITLO |
| 0484 | ◌҄ | COMBINING CYRILLIC PALATALIZATION |
| 0485 | ◌҅ | COMBINING CYRILLIC DASIA PNEUMATA |
| 0486 | ◌҆ | COMBINING CYRILLIC PSILI PNEUMATA |
| 0487 | ◌҇ | COMBINING CYRILLIC POKRYTIE |
| 0488 | ҈ | COMBINING CYRILLIC HUNDRED THOUSANDS SIGN |
| 0489 | ҉ | COMBINING CYRILLIC MILLIONS SIGN |
| 1AB0 | ◌᪰ | COMBINING DOUBLED CIRCUMFLEX ACCENT |
| 1AB1 | ◌᪱ | COMBINING DIAERESIS-RING |
| 1AB2 | ◌᪲ | COMBINING INFINITY |
| 1AB3 | ◌᪳ | COMBINING DOWNWARDS ARROW |
| 1AB4 | ◌᪴ | COMBINING TRIPLE DOT |
| 1AB5 | ◌᪵ | COMBINING X-X BELOW |
| 1AB6 | ◌᪶ | COMBINING WIGGLY LINE BELOW |
| 1AB7 | ◌᪷ | COMBINING OPEN MARK BELOW |
| 1AB8 | ◌᪸ | COMBINING DOUBLE OPEN MARK BELOW |
| 1AB9 | ◌᪹ | COMBINING LIGHT CENTRALIZATION STROKE BELOW |
| 1ABA | ◌᪺ | COMBINING STRONG CENTRALIZATION STROKE BELOW |
| 1ABB | ◌᪻ | COMBINING PARENTHESES ABOVE |
| 1ABC | ◌᪼ | COMBINING DOUBLE PARENTHESES ABOVE |
| 1ABD | ◌᪽ | COMBINING PARENTHESES BELOW |
| 1ABE | ᪾ | COMBINING PARENTHESES OVERLAY |
| 1ABF | ◌ᪿ | COMBINING LATIN SMALL LETTER W BELOW |
| 1AC0 | ◌ᫀ | COMBINING LATIN SMALL LETTER TURNED W BELOW |
| 1AC1 | ◌᫁ | COMBINING LEFT PARENTHESIS ABOVE LEFT |
| 1AC2 | ◌᫂ | COMBINING RIGHT PARENTHESIS ABOVE RIGHT |
| 1AC3 | ◌᫃ | COMBINING LEFT PARENTHESIS BELOW LEFT |
| 1AC4 | ◌᫄ | COMBINING RIGHT PARENTHESIS BELOW RIGHT |
| 1AC5 | ◌᫅ | COMBINING SQUARE BRACKETS ABOVE |
| 1AC6 | ◌᫆ | COMBINING NUMBER SIGN ABOVE |
| 1AC7 | ◌᫇ | COMBINING INVERTED DOUBLE ARCH ABOVE |
| 1AC8 | ◌᫈ | COMBINING PLUS SIGN ABOVE |
| 1AC9 | ◌᫉ | COMBINING DOUBLE PLUS SIGN ABOVE |
| 1ACA | ◌᫊ | COMBINING DOUBLE PLUS SIGN BELOW |
| 1ACB | ◌᫋ | COMBINING TRIPLE ACUTE ACCENT |
| 1ACC | ◌ᫌ | COMBINING LATIN SMALL LETTER INSULAR G |
| 1ACD | ◌ᫍ | COMBINING LATIN SMALL LETTER INSULAR R |
| 1ACE | ◌ᫎ | COMBINING LATIN SMALL LETTER INSULAR T |
| 1DC0 | ◌᷀ | COMBINING DOTTED GRAVE ACCENT |
| 1DC1 | ◌᷁ | COMBINING DOTTED ACUTE ACCENT |
| 1DC2 | ◌᷂ | COMBINING SNAKE BELOW |
| 1DC3 | ◌᷃ | COMBINING SUSPENSION MARK |
| 1DC4 | ◌᷄ | COMBINING MACRON-ACUTE |
| 1DC5 | ◌᷅ | COMBINING GRAVE-MACRON |
| 1DC6 | ◌᷆ | COMBINING MACRON-GRAVE |
| 1DC7 | ◌᷇ | COMBINING ACUTE-MACRON |
| 1DC8 | ◌᷈ | COMBINING GRAVE-ACUTE-GRAVE |
| 1DC9 | ◌᷉ | COMBINING ACUTE-GRAVE-ACUTE |
| 1DCA | ◌᷊ | COMBINING LATIN SMALL LETTER R BELOW |
| 1DCB | ◌᷋ | COMBINING BREVE-MACRON |
| 1DCC | ◌᷌ | COMBINING MACRON-BREVE |
| 1DCD | ◌᷍ | COMBINING DOUBLE CIRCUMFLEX ABOVE |
| 1DCE | ◌᷎ | COMBINING OGONEK ABOVE |
| 1DCF | ◌᷏ | COMBINING ZIGZAG BELOW |
| 1DD0 | ◌᷐ | COMBINING IS BELOW |
| 1DD1 | ◌᷑ | COMBINING UR ABOVE |
| 1DD2 | ◌᷒ | COMBINING US ABOVE |
| 1DD3 | ◌ᷓ | COMBINING LATIN SMALL LETTER FLATTENED OPEN A ABOVE |
| 1DD4 | ◌ᷔ | COMBINING LATIN SMALL LETTER AE |
| 1DD5 | ◌ᷕ | COMBINING LATIN SMALL LETTER AO |
| 1DD6 | ◌ᷖ | COMBINING LATIN SMALL LETTER AV |
| 1DD7 | ◌ᷗ | COMBINING LATIN SMALL LETTER C CEDILLA |
| 1DD8 | ◌ᷘ | COMBINING LATIN SMALL LETTER INSULAR D |
| 1DD9 | ◌ᷙ | COMBINING LATIN SMALL LETTER ETH |
| 1DDA | ◌ᷚ | COMBINING LATIN SMALL LETTER G |
| 1DDB | ◌ᷛ | COMBINING LATIN LETTER SMALL CAPITAL G |
| 1DDC | ◌ᷜ | COMBINING LATIN SMALL LETTER K |
| 1DDD | ◌ᷝ | COMBINING LATIN SMALL LETTER L |
| 1DDE | ◌ᷞ | COMBINING LATIN LETTER SMALL CAPITAL L |
| 1DDF | ◌ᷟ | COMBINING LATIN LETTER SMALL CAPITAL M |
| 1DE0 | ◌ᷠ | COMBINING LATIN SMALL LETTER N |
| 1DE1 | ◌ᷡ | COMBINING LATIN LETTER SMALL CAPITAL N |
| 1DE2 | ◌ᷢ | COMBINING LATIN LETTER SMALL CAPITAL R |
| 1DE3 | ◌ᷣ | COMBINING LATIN SMALL LETTER R ROTUNDA |
| 1DE4 | ◌ᷤ | COMBINING LATIN SMALL LETTER S |
| 1DE5 | ◌ᷥ | COMBINING LATIN SMALL LETTER LONG S |
| 1DE6 | ◌ᷦ | COMBINING LATIN SMALL LETTER Z |
| 1DE7 | ◌ᷧ | COMBINING LATIN SMALL LETTER ALPHA |
| 1DE8 | ◌ᷨ | COMBINING LATIN SMALL LETTER B |
| 1DE9 | ◌ᷩ | COMBINING LATIN SMALL LETTER BETA |
| 1DEA | ◌ᷪ | COMBINING LATIN SMALL LETTER SCHWA |
| 1DEB | ◌ᷫ | COMBINING LATIN SMALL LETTER F |
| 1DEC | ◌ᷬ | COMBINING LATIN SMALL LETTER L WITH DOUBLE MIDDLE TILDE |
| 1DED | ◌ᷭ | COMBINING LATIN SMALL LETTER O WITH LIGHT CENTRALIZATION STROKE |
| 1DEE | ◌ᷮ | COMBINING LATIN SMALL LETTER P |
| 1DEF | ◌ᷯ | COMBINING LATIN SMALL LETTER ESH |
| 1DF0 | ◌ᷰ | COMBINING LATIN SMALL LETTER U WITH LIGHT CENTRALIZATION STROKE |
| 1DF1 | ◌ᷱ | COMBINING LATIN SMALL LETTER W |
| 1DF2 | ◌ᷲ | COMBINING LATIN SMALL LETTER A WITH DIAERESIS |
| 1DF3 | ◌ᷳ | COMBINING LATIN SMALL LETTER O WITH DIAERESIS |
| 1DF4 | ◌ᷴ | COMBINING LATIN SMALL LETTER U WITH DIAERESIS |
| 1DF5 | ◌᷵ | COMBINING UP TACK ABOVE |
| 1DF6 | ◌᷶ | COMBINING KAVYKA ABOVE RIGHT |
| 1DF7 | ◌᷷ | COMBINING KAVYKA ABOVE LEFT |
| 1DF8 | ◌᷸ | COMBINING DOT ABOVE LEFT |
| 1DF9 | ◌᷹ | COMBINING WIDE INVERTED BRIDGE BELOW |
| 1DFA | ◌᷺ | COMBINING DOT BELOW LEFT |
| 1DFB | ◌᷻ | COMBINING DELETION MARK |
| 1DFC | ◌᷼ | COMBINING DOUBLE INVERTED BREVE BELOW |
| 1DFD | ◌᷽ | COMBINING ALMOST EQUAL TO BELOW |
| 1DFE | ◌᷾ | COMBINING LEFT ARROWHEAD ABOVE |
| 1DFF | ◌᷿ | COMBINING RIGHT ARROWHEAD AND DOWN ARROWHEAD BELOW |
| 20D0 | ◌⃐ | COMBINING LEFT HARPOON ABOVE |
| 20D1 | ◌⃑ | COMBINING RIGHT HARPOON ABOVE |
| 20D2 | ◌⃒ | COMBINING LONG VERTICAL LINE OVERLAY |
| 20D3 | ◌⃓ | COMBINING SHORT VERTICAL LINE OVERLAY |
| 20D4 | ◌⃔ | COMBINING ANTICLOCKWISE ARROW ABOVE |
| 20D5 | ◌⃕ | COMBINING CLOCKWISE ARROW ABOVE |
| 20D6 | ◌⃖ | COMBINING LEFT ARROW ABOVE |
| 20D7 | ◌⃗ | COMBINING RIGHT ARROW ABOVE |
| 20D8 | ◌⃘ | COMBINING RING OVERLAY |
| 20D9 | ◌⃙ | COMBINING CLOCKWISE RING OVERLAY |
| 20DA | ◌⃚ | COMBINING ANTICLOCKWISE RING OVERLAY |
| 20DB | ◌⃛ | COMBINING THREE DOTS ABOVE |
| 20DC | ◌⃜ | COMBINING FOUR DOTS ABOVE |
| 20DD | ⃝ | COMBINING ENCLOSING CIRCLE |
| 20DE |  | COMBINING ENCLOSING SQUARE |
| 20DF |  | COMBINING ENCLOSING DIAMOND |
| 20E0 | ⃠ | COMBINING ENCLOSING CIRCLE BACKSLASH |
| 20E1 | ◌⃡ | COMBINING LEFT RIGHT ARROW ABOVE |
| 20E2 | ⃢ | COMBINING ENCLOSING SCREEN |
| 20E3 | ⃣ | COMBINING ENCLOSING KEYCAP |
| 20E4 | ⃤ | COMBINING ENCLOSING UPWARD POINTING TRIANGLE |
| 20E5 | ◌⃥ | COMBINING REVERSE SOLIDUS OVERLAY |
| 20E6 | ◌⃦ | COMBINING DOUBLE VERTICAL STROKE OVERLAY |
| 20E7 | ◌⃧ | COMBINING ANNUITY SYMBOL |
| 20E8 | ◌⃨ | COMBINING TRIPLE UNDERDOT |
| 20E9 | ◌⃩ | COMBINING WIDE BRIDGE ABOVE |
| 20EA | ◌⃪ | COMBINING LEFTWARDS ARROW OVERLAY |
| 20EB | ◌⃫ | COMBINING LONG DOUBLE SOLIDUS OVERLAY |
| 20EC | ◌⃬ | COMBINING RIGHTWARDS HARPOON WITH BARB DOWNWARDS |
| 20ED | ◌⃭ | COMBINING LEFTWARDS HARPOON WITH BARB DOWNWARDS |
| 20EE | ◌⃮ | COMBINING LEFT ARROW BELOW |
| 20EF | ◌⃯ | COMBINING RIGHT ARROW BELOW |
| 20F0 | ◌⃰ | COMBINING ASTERISK ABOVE |
| 2DE0 | ◌ⷠ | COMBINING CYRILLIC LETTER BE |
| 2DE1 | ◌ⷡ | COMBINING CYRILLIC LETTER VE |
| 2DE2 | ◌ⷢ | COMBINING CYRILLIC LETTER GHE |
| 2DE3 | ◌ⷣ | COMBINING CYRILLIC LETTER DE |
| 2DE4 | ◌ⷤ | COMBINING CYRILLIC LETTER ZHE |
| 2DE5 | ◌ⷥ | COMBINING CYRILLIC LETTER ZE |
| 2DE6 | ◌ⷦ | COMBINING CYRILLIC LETTER KA |
| 2DE7 | ◌ⷧ | COMBINING CYRILLIC LETTER EL |
| 2DE8 | ◌ⷨ | COMBINING CYRILLIC LETTER EM |
| 2DE9 | ◌ⷩ | COMBINING CYRILLIC LETTER EN |
| 2DEA | ◌ⷪ | COMBINING CYRILLIC LETTER O |
| 2DEB | ◌ⷫ | COMBINING CYRILLIC LETTER PE |
| 2DEC | ◌ⷬ | COMBINING CYRILLIC LETTER ER |
| 2DED | ◌ⷭ | COMBINING CYRILLIC LETTER ES |
| 2DEE | ◌ⷮ | COMBINING CYRILLIC LETTER TE |
| 2DEF | ◌ⷯ | COMBINING CYRILLIC LETTER HA |
| 2DF0 | ◌ⷰ | COMBINING CYRILLIC LETTER TSE |
| 2DF1 | ◌ⷱ | COMBINING CYRILLIC LETTER CHE |
| 2DF2 | ◌ⷲ | COMBINING CYRILLIC LETTER SHA |
| 2DF3 | ◌ⷳ | COMBINING CYRILLIC LETTER SHCHA |
| 2DF4 | ◌ⷴ | COMBINING CYRILLIC LETTER FITA |
| 2DF5 | ◌ⷵ | COMBINING CYRILLIC LETTER ES-TE |
| 2DF6 | ◌ⷶ | COMBINING CYRILLIC LETTER A |
| 2DF7 | ◌ⷷ | COMBINING CYRILLIC LETTER IE |
| 2DF8 | ◌ⷸ | COMBINING CYRILLIC LETTER DJERV |
| 2DF9 | ◌ⷹ | COMBINING CYRILLIC LETTER MONOGRAPH UK |
| 2DFA | ◌ⷺ | COMBINING CYRILLIC LETTER YAT |
| 2DFB | ◌ⷻ | COMBINING CYRILLIC LETTER YU |
| 2DFC | ◌ⷼ | COMBINING CYRILLIC LETTER IOTIFIED A |
| 2DFD | ◌ⷽ | COMBINING CYRILLIC LETTER LITTLE YUS |
| 2DFE | ◌ⷾ | COMBINING CYRILLIC LETTER BIG YUS |
| 2DFF | ◌ⷿ | COMBINING CYRILLIC LETTER IOTIFIED BIG YUS |
| 3099 | ◌゙ | COMBINING KATAKANA-HIRAGANA VOICED SOUND MARK |
| 309A | ◌゚ | COMBINING KATAKANA-HIRAGANA SEMI-VOICED SOUND MARK |
| A66F | ◌꙯ | COMBINING CYRILLIC VZMET |
| A670 | ꙰ | COMBINING CYRILLIC TEN MILLIONS SIGN |
| A671 | ꙱ | COMBINING CYRILLIC HUNDRED MILLIONS SIGN |
| A672 | ꙲ | COMBINING CYRILLIC THOUSAND MILLIONS SIGN |
| A674 | ◌ꙴ | COMBINING CYRILLIC LETTER UKRAINIAN IE |
| A675 | ◌ꙵ | COMBINING CYRILLIC LETTER I |
| A676 | ◌ꙶ | COMBINING CYRILLIC LETTER YI |
| A677 | ◌ꙷ | COMBINING CYRILLIC LETTER U |
| A678 | ◌ꙸ | COMBINING CYRILLIC LETTER HARD SIGN |
| A679 | ◌ꙹ | COMBINING CYRILLIC LETTER YERU |
| A67A | ◌ꙺ | COMBINING CYRILLIC LETTER SOFT SIGN |
| A67B | ◌ꙻ | COMBINING CYRILLIC LETTER OMEGA |
| A67C | ◌꙼ | COMBINING CYRILLIC KAVYKA |
| A67D | ◌꙽ | COMBINING CYRILLIC PAYEROK |
| A69E | ◌ꚞ | COMBINING CYRILLIC LETTER EF |
| A69F | ◌ꚟ | COMBINING CYRILLIC LETTER IOTIFIED E |
| A8E0 | ꣠ | COMBINING DEVANAGARI DIGIT ZERO |
| A8E1 | ꣡ | COMBINING DEVANAGARI DIGIT ONE |
| A8E2 | ꣢ | COMBINING DEVANAGARI DIGIT TWO |
| A8E3 | ꣣ | COMBINING DEVANAGARI DIGIT THREE |
| A8E4 | ꣤ | COMBINING DEVANAGARI DIGIT FOUR |
| A8E5 | ꣥ | COMBINING DEVANAGARI DIGIT FIVE |
| A8E6 | ꣦ | COMBINING DEVANAGARI DIGIT SIX |
| A8E7 | ꣧ | COMBINING DEVANAGARI DIGIT SEVEN |
| A8E8 | ꣨ | COMBINING DEVANAGARI DIGIT EIGHT |
| A8E9 | ꣩ | COMBINING DEVANAGARI DIGIT NINE |
| A8EA | ꣪ | COMBINING DEVANAGARI LETTER A |
| A8EB | ꣫ | COMBINING DEVANAGARI LETTER U |
| A8EC | ꣬ | COMBINING DEVANAGARI LETTER KA |
| A8ED | ꣭ | COMBINING DEVANAGARI LETTER NA |
| A8EE | ꣮ | COMBINING DEVANAGARI LETTER PA |
| A8EF | ꣯ | COMBINING DEVANAGARI LETTER RA |
| A8F0 | ꣰ | COMBINING DEVANAGARI LETTER VI |
| A8F1 | ꣱ | COMBINING DEVANAGARI SIGN AVAGRAHA |
| FE20 | ◌︠ | COMBINING LIGATURE LEFT HALF |
| FE21 | ◌︡ | COMBINING LIGATURE RIGHT HALF |
| FE22 | ◌︢ | COMBINING DOUBLE TILDE LEFT HALF |
| FE23 | ◌︣ | COMBINING DOUBLE TILDE RIGHT HALF |
| FE24 | ◌︤ | COMBINING MACRON LEFT HALF |
| FE25 | ◌︥ | COMBINING MACRON RIGHT HALF |
| FE26 | ◌︦ | COMBINING CONJOINING MACRON |
| FE27 | ◌︧ | COMBINING LIGATURE LEFT HALF BELOW |
| FE28 | ◌︨ | COMBINING LIGATURE RIGHT HALF BELOW |
| FE29 | ◌︩ | COMBINING TILDE LEFT HALF BELOW |
| FE2A | ◌︪ | COMBINING TILDE RIGHT HALF BELOW |
| FE2B | ◌︫ | COMBINING MACRON LEFT HALF BELOW |
| FE2C | ◌︬ | COMBINING MACRON RIGHT HALF BELOW |
| FE2D | ◌︭ | COMBINING CONJOINING MACRON BELOW |
| FE2E | ◌︮ | COMBINING CYRILLIC TITLO LEFT HALF |
| FE2F | ◌︯ | COMBINING CYRILLIC TITLO RIGHT HALF |
| 10376 | ◌ | COMBINING OLD PERMIC LETTER AN |
| 10377 | ◌ | COMBINING OLD PERMIC LETTER DOI |
| 10378 | ◌ | COMBINING OLD PERMIC LETTER ZATA |
| 10379 | ◌ | COMBINING OLD PERMIC LETTER NENOE |
| 1037A | ◌ | COMBINING OLD PERMIC LETTER SII |
| 1133B | ◌ | COMBINING BINDU BELOW |
| 11366 | ◌ | COMBINING GRANTHA DIGIT ZERO |
| 11367 | ◌ | COMBINING GRANTHA DIGIT ONE |
| 11368 | ◌ | COMBINING GRANTHA DIGIT TWO |
| 11369 | ◌ | COMBINING GRANTHA DIGIT THREE |
| 1136A | ◌ | COMBINING GRANTHA DIGIT FOUR |
| 1136B | ◌ | COMBINING GRANTHA DIGIT FIVE |
| 1136C | ◌ | COMBINING GRANTHA DIGIT SIX |
| 11370 | ◌ | COMBINING GRANTHA LETTER A |
| 11371 | ◌ | COMBINING GRANTHA LETTER KA |
| 11372 | ◌ | COMBINING GRANTHA LETTER NA |
| 11373 | ◌ | COMBINING GRANTHA LETTER VI |
| 11374 | ◌ | COMBINING GRANTHA LETTER PA |
| 1D242 | ◌ | COMBINING GREEK MUSICAL TRISEME |
| 1D243 | ◌ | COMBINING GREEK MUSICAL TETRASEME |
| 1D244 | ◌ | COMBINING GREEK MUSICAL PENTASEME |
| 1E000 | ◌ | COMBINING GLAGOLITIC LETTER AZU |
| 1E001 | ◌ | COMBINING GLAGOLITIC LETTER BUKY |
| 1E002 | ◌ | COMBINING GLAGOLITIC LETTER VEDE |
| 1E003 | ◌ | COMBINING GLAGOLITIC LETTER GLAGOLI |
| 1E004 | ◌ | COMBINING GLAGOLITIC LETTER DOBRO |
| 1E005 | ◌ | COMBINING GLAGOLITIC LETTER YESTU |
| 1E006 | ◌ | COMBINING GLAGOLITIC LETTER ZHIVETE |
| 1E008 | ◌ | COMBINING GLAGOLITIC LETTER ZEMLJA |
| 1E009 | ◌ | COMBINING GLAGOLITIC LETTER IZHE |
| 1E00A | ◌ | COMBINING GLAGOLITIC LETTER INITIAL IZHE |
| 1E00B | ◌ | COMBINING GLAGOLITIC LETTER I |
| 1E00C | ◌ | COMBINING GLAGOLITIC LETTER DJERVI |
| 1E00D | ◌ | COMBINING GLAGOLITIC LETTER KAKO |
| 1E00E | ◌ | COMBINING GLAGOLITIC LETTER LJUDIJE |
| 1E00F | ◌ | COMBINING GLAGOLITIC LETTER MYSLITE |
| 1E010 | ◌ | COMBINING GLAGOLITIC LETTER NASHI |
| 1E011 | ◌ | COMBINING GLAGOLITIC LETTER ONU |
| 1E012 | ◌ | COMBINING GLAGOLITIC LETTER POKOJI |
| 1E013 | ◌ | COMBINING GLAGOLITIC LETTER RITSI |
| 1E014 | ◌ | COMBINING GLAGOLITIC LETTER SLOVO |
| 1E015 | ◌ | COMBINING GLAGOLITIC LETTER TVRIDO |
| 1E016 | ◌ | COMBINING GLAGOLITIC LETTER UKU |
| 1E017 | ◌ | COMBINING GLAGOLITIC LETTER FRITU |
| 1E018 | ◌ | COMBINING GLAGOLITIC LETTER HERU |
| 1E01B | ◌ | COMBINING GLAGOLITIC LETTER SHTA |
| 1E01C | ◌ | COMBINING GLAGOLITIC LETTER TSI |
| 1E01D | ◌ | COMBINING GLAGOLITIC LETTER CHRIVI |
| 1E01E | ◌ | COMBINING GLAGOLITIC LETTER SHA |
| 1E01F | ◌ | COMBINING GLAGOLITIC LETTER YERU |
| 1E020 | ◌ | COMBINING GLAGOLITIC LETTER YERI |
| 1E021 | ◌ | COMBINING GLAGOLITIC LETTER YATI |
| 1E023 | ◌ | COMBINING GLAGOLITIC LETTER YU |
| 1E024 | ◌ | COMBINING GLAGOLITIC LETTER SMALL YUS |
| 1E026 | ◌ | COMBINING GLAGOLITIC LETTER YO |
| 1E027 | ◌ | COMBINING GLAGOLITIC LETTER IOTATED SMALL YUS |
| 1E028 | ◌ | COMBINING GLAGOLITIC LETTER BIG YUS |
| 1E029 | ◌ | COMBINING GLAGOLITIC LETTER IOTATED BIG YUS |
| 1E02A | ◌ | COMBINING GLAGOLITIC LETTER FITA |
| 1E08F | ◌ | COMBINING CYRILLIC SMALL LETTER BYELORUSSIAN-UKRAINIAN I |

## Litery i symbole modyfikujące

Litery modyfikujące (ang. modifying letters) to takie, które mają własną szerokość i są dodawane po lub przed modyfikowanym znakiem, z reguły jako pomniejszone wersje bazowych liter (kategorie Lm). Litery modyfikujące mają nazwy opisowe zaczynające się od „MODIFIER LETTER”. Część znaków modyfikujących inne znaki nie pochodzą od liter. Należą do kategorii Sk.

Zestawienie regularnych liter z alfabetu łacińskiego (i nie tylko) z odpowiednikami w znakach modyfikujących przedstawia tab. 15.

Tab. . Zestawienie liter bazowych i odpowiadających im znaków modyfikujących

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Litera bazowa | | | Znak modyfikujący | | |
| 0041 | A | LATIN CAPITAL LETTER A | 1D2C | ᴬ | MODIFIER LETTER CAPITAL A |
| 0042 | B | LATIN CAPITAL LETTER B | 1D2E | ᴮ | MODIFIER LETTER CAPITAL B |
| 0043 | C | LATIN CAPITAL LETTER C | A7F2 | ꟲ | MODIFIER LETTER CAPITAL C |
| 0044 | D | LATIN CAPITAL LETTER D | 1D30 | ᴰ | MODIFIER LETTER CAPITAL D |
| 0045 | E | LATIN CAPITAL LETTER E | 1D31 | ᴱ | MODIFIER LETTER CAPITAL E |
| 0046 | F | LATIN CAPITAL LETTER F | A7F3 | ꟳ | MODIFIER LETTER CAPITAL F |
| 0047 | G | LATIN CAPITAL LETTER G | 1D33 | ᴳ | MODIFIER LETTER CAPITAL G |
| 0048 | H | LATIN CAPITAL LETTER H | 1D34 | ᴴ | MODIFIER LETTER CAPITAL H |
| 0049 | I | LATIN CAPITAL LETTER I | 1D35 | ᴵ | MODIFIER LETTER CAPITAL I |
| 004A | J | LATIN CAPITAL LETTER J | 1D36 | ᴶ | MODIFIER LETTER CAPITAL J |
| 004B | K | LATIN CAPITAL LETTER K | 1D37 | ᴷ | MODIFIER LETTER CAPITAL K |
| 004C | L | LATIN CAPITAL LETTER L | 1D38 | ᴸ | MODIFIER LETTER CAPITAL L |
| 004D | M | LATIN CAPITAL LETTER M | 1D39 | ᴹ | MODIFIER LETTER CAPITAL M |
| 004E | N | LATIN CAPITAL LETTER N | 1D3A | ᴺ | MODIFIER LETTER CAPITAL N |
| 004F | O | LATIN CAPITAL LETTER O | 1D3C | ᴼ | MODIFIER LETTER CAPITAL O |
| 0050 | P | LATIN CAPITAL LETTER P | 1D3E | ᴾ | MODIFIER LETTER CAPITAL P |
| 0051 | Q | LATIN CAPITAL LETTER Q | A7F4 | ꟴ | MODIFIER LETTER CAPITAL Q |
| 0052 | R | LATIN CAPITAL LETTER R | 1D3F | ᴿ | MODIFIER LETTER CAPITAL R |
| 0054 | T | LATIN CAPITAL LETTER T | 1D40 | ᵀ | MODIFIER LETTER CAPITAL T |
| 0055 | U | LATIN CAPITAL LETTER U | 1D41 | ᵁ | MODIFIER LETTER CAPITAL U |
| 0056 | V | LATIN CAPITAL LETTER V | 2C7D | ⱽ | MODIFIER LETTER CAPITAL V |
| 0057 | W | LATIN CAPITAL LETTER W | 1D42 | ᵂ | MODIFIER LETTER CAPITAL W |
| 0061 | a | LATIN SMALL LETTER A | 1D43 | ᵃ | MODIFIER LETTER SMALL A |
| 0062 | b | LATIN SMALL LETTER B | 1D47 | ᵇ | MODIFIER LETTER SMALL B |
| 0063 | c | LATIN SMALL LETTER C | 1D9C | ᶜ | MODIFIER LETTER SMALL C |
| 0064 | d | LATIN SMALL LETTER D | 1D48 | ᵈ | MODIFIER LETTER SMALL D |
| 0065 | e | LATIN SMALL LETTER E | 1D49 | ᵉ | MODIFIER LETTER SMALL E |
| 0066 | f | LATIN SMALL LETTER F | 1DA0 | ᶠ | MODIFIER LETTER SMALL F |
| 0067 | g | LATIN SMALL LETTER G | 1D4D | ᵍ | MODIFIER LETTER SMALL G |
| 0068 | h | LATIN SMALL LETTER H | 02B0 | ʰ | MODIFIER LETTER SMALL H |
| 006A | j | LATIN SMALL LETTER J | 02B2 | ʲ | MODIFIER LETTER SMALL J |
| 006B | k | LATIN SMALL LETTER K | 1D4F | ᵏ | MODIFIER LETTER SMALL K |
| 006C | l | LATIN SMALL LETTER L | 02E1 | ˡ | MODIFIER LETTER SMALL L |
| 006D | m | LATIN SMALL LETTER M | 1D50 | ᵐ | MODIFIER LETTER SMALL M |
| 006F | o | LATIN SMALL LETTER O | 1D52 | ᵒ | MODIFIER LETTER SMALL O |
| 0070 | p | LATIN SMALL LETTER P | 1D56 | ᵖ | MODIFIER LETTER SMALL P |
| 0071 | q | LATIN SMALL LETTER Q | 107A5 | 𐞥 | MODIFIER LETTER SMALL Q |
| 0072 | r | LATIN SMALL LETTER R | 02B3 | ʳ | MODIFIER LETTER SMALL R |
| 0073 | s | LATIN SMALL LETTER S | 02E2 | ˢ | MODIFIER LETTER SMALL S |
| 0074 | t | LATIN SMALL LETTER T | 1D57 | ᵗ | MODIFIER LETTER SMALL T |
| 0075 | u | LATIN SMALL LETTER U | 1D58 | ᵘ | MODIFIER LETTER SMALL U |
| 0076 | v | LATIN SMALL LETTER V | 1D5B | ᵛ | MODIFIER LETTER SMALL V |
| 0077 | w | LATIN SMALL LETTER W | 02B7 | ʷ | MODIFIER LETTER SMALL W |
| 0078 | x | LATIN SMALL LETTER X | 02E3 | ˣ | MODIFIER LETTER SMALL X |
| 0079 | y | LATIN SMALL LETTER Y | 02B8 | ʸ | MODIFIER LETTER SMALL Y |
| 007A | z | LATIN SMALL LETTER Z | 1DBB | ᶻ | MODIFIER LETTER SMALL Z |
| 00C6 | Æ | LATIN CAPITAL LETTER AE | 1D2D | ᴭ | MODIFIER LETTER CAPITAL AE |
| 00E6 | æ | LATIN SMALL LETTER AE | 10783 | 𐞃 | MODIFIER LETTER SMALL AE |
| 00F0 | ð | LATIN SMALL LETTER ETH | 1D9E | ᶞ | MODIFIER LETTER SMALL ETH |
| 00F8 | ø | LATIN SMALL LETTER O WITH STROKE | 107A2 | 𐞢 | MODIFIER LETTER SMALL O WITH STROKE |
| 0126 | Ħ | LATIN CAPITAL LETTER H WITH STROKE | A7F8 | ꟸ | MODIFIER LETTER CAPITAL H WITH STROKE |
| 0127 | ħ | LATIN SMALL LETTER H WITH STROKE | 10795 | 𐞕 | MODIFIER LETTER SMALL H WITH STROKE |
| 014B | ŋ | LATIN SMALL LETTER ENG | 1D51 | ᵑ | MODIFIER LETTER SMALL ENG |
| 0153 | œ | LATIN SMALL LIGATURE OE | A7F9 | ꟹ | MODIFIER LETTER SMALL LIGATURE OE |
| 018E | Ǝ | LATIN CAPITAL LETTER REVERSED E | 1D32 | ᴲ | MODIFIER LETTER CAPITAL REVERSED E |
| 01AB | ƫ | LATIN SMALL LETTER T WITH PALATAL HOOK | 1DB5 | ᶵ | MODIFIER LETTER SMALL T WITH PALATAL HOOK |
| 01C0 | ǀ | LATIN LETTER DENTAL CLICK | 107B6 | 𐞶 | MODIFIER LETTER DENTAL CLICK |
| 01C1 | ǁ | LATIN LETTER LATERAL CLICK | 107B7 | 𐞷 | MODIFIER LETTER LATERAL CLICK |
| 01C2 | ǂ | LATIN LETTER ALVEOLAR CLICK | 107B8 | 𐞸 | MODIFIER LETTER ALVEOLAR CLICK |
| 0222 | Ȣ | LATIN CAPITAL LETTER OU | 1D3D | ᴽ | MODIFIER LETTER CAPITAL OU |
| 0250 | ɐ | LATIN SMALL LETTER TURNED A | 1D44 | ᵄ | MODIFIER LETTER SMALL TURNED A |
| 0251 | ɑ | LATIN SMALL LETTER ALPHA | 1D45 | ᵅ | MODIFIER LETTER SMALL ALPHA |
| 0252 | ɒ | LATIN SMALL LETTER TURNED ALPHA | 1D9B | ᶛ | MODIFIER LETTER SMALL TURNED ALPHA |
| 0253 | ɓ | LATIN SMALL LETTER B WITH HOOK | 10785 | 𐞅 | MODIFIER LETTER SMALL B WITH HOOK |
| 0254 | ɔ | LATIN SMALL LETTER OPEN O | 1D53 | ᵓ | MODIFIER LETTER SMALL OPEN O |
| 0255 | ɕ | LATIN SMALL LETTER C WITH CURL | 1D9D | ᶝ | MODIFIER LETTER SMALL C WITH CURL |
| 0256 | ɖ | LATIN SMALL LETTER D WITH TAIL | 1078B | 𐞋 | MODIFIER LETTER SMALL D WITH TAIL |
| 0257 | ɗ | LATIN SMALL LETTER D WITH HOOK | 1078C | 𐞌 | MODIFIER LETTER SMALL D WITH HOOK |
| 0258 | ɘ | LATIN SMALL LETTER REVERSED E | 1078E | 𐞎 | MODIFIER LETTER SMALL REVERSED E |
| 0259 | ə | LATIN SMALL LETTER SCHWA | 1D4A | ᵊ | MODIFIER LETTER SMALL SCHWA |
| 025B | ɛ | LATIN SMALL LETTER OPEN E | 1D4B | ᵋ | MODIFIER LETTER SMALL OPEN E |
| 025C | ɜ | LATIN SMALL LETTER REVERSED OPEN E | 1D9F | ᶟ | MODIFIER LETTER SMALL REVERSED OPEN E |
| 025E | ɞ | LATIN SMALL LETTER CLOSED REVERSED OPEN E | 1078F | 𐞏 | MODIFIER LETTER SMALL CLOSED REVERSED OPEN E |
| 025F | ɟ | LATIN SMALL LETTER DOTLESS J WITH STROKE | 1DA1 | ᶡ | MODIFIER LETTER SMALL DOTLESS J WITH STROKE |
| 0260 | ɠ | LATIN SMALL LETTER G WITH HOOK | 10793 | 𐞓 | MODIFIER LETTER SMALL G WITH HOOK |
| 0261 | ɡ | LATIN SMALL LETTER SCRIPT G | 1DA2 | ᶢ | MODIFIER LETTER SMALL SCRIPT G |
| 0262 | ɢ | LATIN LETTER SMALL CAPITAL G | 10792 | 𐞒 | MODIFIER LETTER SMALL CAPITAL G |
| 0263 | ɣ | LATIN SMALL LETTER GAMMA | 02E0 | ˠ | MODIFIER LETTER SMALL GAMMA |
| 0264 | ɤ | LATIN SMALL LETTER RAMS HORN | 10791 | 𐞑 | MODIFIER LETTER SMALL RAMS HORN |
| 0265 | ɥ | LATIN SMALL LETTER TURNED H | 1DA3 | ᶣ | MODIFIER LETTER SMALL TURNED H |
| 0266 | ɦ | LATIN SMALL LETTER H WITH HOOK | 02B1 | ʱ | MODIFIER LETTER SMALL H WITH HOOK |
| 0267 | ɧ | LATIN SMALL LETTER HENG WITH HOOK | 10797 | 𐞗 | MODIFIER LETTER SMALL HENG WITH HOOK |
| 0268 | ɨ | LATIN SMALL LETTER I WITH STROKE | 1DA4 | ᶤ | MODIFIER LETTER SMALL I WITH STROKE |
| 0269 | ɩ | LATIN SMALL LETTER IOTA | 1DA5 | ᶥ | MODIFIER LETTER SMALL IOTA |
| 026A | ɪ | LATIN LETTER SMALL CAPITAL I | 1DA6 | ᶦ | MODIFIER LETTER SMALL CAPITAL I |
| 026B | ɫ | LATIN SMALL LETTER L WITH MIDDLE TILDE | AB5E | ꭞ | MODIFIER LETTER SMALL L WITH MIDDLE TILDE |
| 026C | ɬ | LATIN SMALL LETTER L WITH BELT | 1079B | 𐞛 | MODIFIER LETTER SMALL L WITH BELT |
| 026D | ɭ | LATIN SMALL LETTER L WITH RETROFLEX HOOK | 1DA9 | ᶩ | MODIFIER LETTER SMALL L WITH RETROFLEX HOOK |
| 026E | ɮ | LATIN SMALL LETTER LEZH | 1079E | 𐞞 | MODIFIER LETTER SMALL LEZH |
| 026F | ɯ | LATIN SMALL LETTER TURNED M | 1D5A | ᵚ | MODIFIER LETTER SMALL TURNED M |
| 0270 | ɰ | LATIN SMALL LETTER TURNED M WITH LONG LEG | 1DAD | ᶭ | MODIFIER LETTER SMALL TURNED M WITH LONG LEG |
| 0271 | ɱ | LATIN SMALL LETTER M WITH HOOK | 1DAC | ᶬ | MODIFIER LETTER SMALL M WITH HOOK |
| 0272 | ɲ | LATIN SMALL LETTER N WITH LEFT HOOK | 1DAE | ᶮ | MODIFIER LETTER SMALL N WITH LEFT HOOK |
| 0273 | ɳ | LATIN SMALL LETTER N WITH RETROFLEX HOOK | 1DAF | ᶯ | MODIFIER LETTER SMALL N WITH RETROFLEX HOOK |
| 0274 | ɴ | LATIN LETTER SMALL CAPITAL N | 1DB0 | ᶰ | MODIFIER LETTER SMALL CAPITAL N |
| 0275 | ɵ | LATIN SMALL LETTER BARRED O | 1DB1 | ᶱ | MODIFIER LETTER SMALL BARRED O |
| 0276 | ɶ | LATIN LETTER SMALL CAPITAL OE | 107A3 | 𐞣 | MODIFIER LETTER SMALL CAPITAL OE |
| 0277 | ɷ | LATIN SMALL LETTER CLOSED OMEGA | 107A4 | 𐞤 | MODIFIER LETTER SMALL CLOSED OMEGA |
| 0278 | ɸ | LATIN SMALL LETTER PHI | 1DB2 | ᶲ | MODIFIER LETTER SMALL PHI |
| 0279 | ɹ | LATIN SMALL LETTER TURNED R | 02B4 | ʴ | MODIFIER LETTER SMALL TURNED R |
| 027A | ɺ | LATIN SMALL LETTER TURNED R WITH LONG LEG | 107A6 | 𐞦 | MODIFIER LETTER SMALL TURNED R WITH LONG LEG |
| 027B | ɻ | LATIN SMALL LETTER TURNED R WITH HOOK | 02B5 | ʵ | MODIFIER LETTER SMALL TURNED R WITH HOOK |
| 027D | ɽ | LATIN SMALL LETTER R WITH TAIL | 107A8 | 𐞨 | MODIFIER LETTER SMALL R WITH TAIL |
| 027E | ɾ | LATIN SMALL LETTER R WITH FISHHOOK | 107A9 | 𐞩 | MODIFIER LETTER SMALL R WITH FISHHOOK |
| 0280 | ʀ | LATIN LETTER SMALL CAPITAL R | 107AA | 𐞪 | MODIFIER LETTER SMALL CAPITAL R |
| 0281 | ʁ | LATIN LETTER SMALL CAPITAL INVERTED R | 02B6 | ʶ | MODIFIER LETTER SMALL CAPITAL INVERTED R |
| 0282 | ʂ | LATIN SMALL LETTER S WITH HOOK | 1DB3 | ᶳ | MODIFIER LETTER SMALL S WITH HOOK |
| 0283 | ʃ | LATIN SMALL LETTER ESH | 1DB4 | ᶴ | MODIFIER LETTER SMALL ESH |
| 0284 | ʄ | LATIN SMALL LETTER DOTLESS J WITH STROKE AND HOOK | 10798 | 𐞘 | MODIFIER LETTER SMALL DOTLESS J WITH STROKE AND HOOK |
| 0288 | ʈ | LATIN SMALL LETTER T WITH RETROFLEX HOOK | 107AF | 𐞯 | MODIFIER LETTER SMALL T WITH RETROFLEX HOOK |
| 0289 | ʉ | LATIN SMALL LETTER U BAR | 1DB6 | ᶶ | MODIFIER LETTER SMALL U BAR |
| 028A | ʊ | LATIN SMALL LETTER UPSILON | 1DB7 | ᶷ | MODIFIER LETTER SMALL UPSILON |
| 028B | ʋ | LATIN SMALL LETTER V WITH HOOK | 1DB9 | ᶹ | MODIFIER LETTER SMALL V WITH HOOK |
| 028C | ʌ | LATIN SMALL LETTER TURNED V | 1DBA | ᶺ | MODIFIER LETTER SMALL TURNED V |
| 028D | ʍ | LATIN SMALL LETTER TURNED W | AB69 | ꭩ | MODIFIER LETTER SMALL TURNED W |
| 028E | ʎ | LATIN SMALL LETTER TURNED Y | 107A0 | 𐞠 | MODIFIER LETTER SMALL TURNED Y |
| 028F | ʏ | LATIN LETTER SMALL CAPITAL Y | 107B2 | 𐞲 | MODIFIER LETTER SMALL CAPITAL Y |
| 0290 | ʐ | LATIN SMALL LETTER Z WITH RETROFLEX HOOK | 1DBC | ᶼ | MODIFIER LETTER SMALL Z WITH RETROFLEX HOOK |
| 0291 | ʑ | LATIN SMALL LETTER Z WITH CURL | 1DBD | ᶽ | MODIFIER LETTER SMALL Z WITH CURL |
| 0292 | ʒ | LATIN SMALL LETTER EZH | 1DBE | ᶾ | MODIFIER LETTER SMALL EZH |
| 0294 | ʔ | LATIN LETTER GLOTTAL STOP | 02C0 | ˀ | MODIFIER LETTER GLOTTAL STOP |
| 0298 | ʘ | LATIN LETTER BILABIAL CLICK | 107B5 | 𐞵 | MODIFIER LETTER BILABIAL CLICK |
| 0299 | ʙ | LATIN LETTER SMALL CAPITAL B | 10784 | 𐞄 | MODIFIER LETTER SMALL CAPITAL B |
| 029B | ʛ | LATIN LETTER SMALL CAPITAL G WITH HOOK | 10794 | 𐞔 | MODIFIER LETTER SMALL CAPITAL G WITH HOOK |
| 029C | ʜ | LATIN LETTER SMALL CAPITAL H | 10796 | 𐞖 | MODIFIER LETTER SMALL CAPITAL H |
| 029D | ʝ | LATIN SMALL LETTER J WITH CROSSED-TAIL | 1DA8 | ᶨ | MODIFIER LETTER SMALL J WITH CROSSED-TAIL |
| 029F | ʟ | LATIN LETTER SMALL CAPITAL L | 1DAB | ᶫ | MODIFIER LETTER SMALL CAPITAL L |
| 02A1 | ʡ | LATIN LETTER GLOTTAL STOP WITH STROKE | 107B3 | 𐞳 | MODIFIER LETTER GLOTTAL STOP WITH STROKE |
| 02A2 | ʢ | LATIN LETTER REVERSED GLOTTAL STOP WITH STROKE | 107B4 | 𐞴 | MODIFIER LETTER REVERSED GLOTTAL STOP WITH STROKE |
| 02A3 | ʣ | LATIN SMALL LETTER DZ DIGRAPH | 10787 | 𐞇 | MODIFIER LETTER SMALL DZ DIGRAPH |
| 02A4 | ʤ | LATIN SMALL LETTER DEZH DIGRAPH | 1078A | 𐞊 | MODIFIER LETTER SMALL DEZH DIGRAPH |
| 02A5 | ʥ | LATIN SMALL LETTER DZ DIGRAPH WITH CURL | 10789 | 𐞉 | MODIFIER LETTER SMALL DZ DIGRAPH WITH CURL |
| 02A6 | ʦ | LATIN SMALL LETTER TS DIGRAPH | 107AC | 𐞬 | MODIFIER LETTER SMALL TS DIGRAPH |
| 02A7 | ʧ | LATIN SMALL LETTER TESH DIGRAPH | 107AE | 𐞮 | MODIFIER LETTER SMALL TESH DIGRAPH |
| 02A8 | ʨ | LATIN SMALL LETTER TC DIGRAPH WITH CURL | 107AB | 𐞫 | MODIFIER LETTER SMALL TC DIGRAPH WITH CURL |
| 02A9 | ʩ | LATIN SMALL LETTER FENG DIGRAPH | 10790 | 𐞐 | MODIFIER LETTER SMALL FENG DIGRAPH |
| 02AA | ʪ | LATIN SMALL LETTER LS DIGRAPH | 10799 | 𐞙 | MODIFIER LETTER SMALL LS DIGRAPH |
| 02AB | ʫ | LATIN SMALL LETTER LZ DIGRAPH | 1079A | 𐞚 | MODIFIER LETTER SMALL LZ DIGRAPH |
| 03B1 | α | GREEK SMALL LETTER ALPHA | 1D45 | ᵅ | MODIFIER LETTER SMALL ALPHA |
| 03B2 | β | GREEK SMALL LETTER BETA | 1D5D | ᵝ | MODIFIER LETTER SMALL BETA |
| 03B3 | γ | GREEK SMALL LETTER GAMMA | 02E0 | ˠ | MODIFIER LETTER SMALL GAMMA |
| 03B4 | δ | GREEK SMALL LETTER DELTA | 1D5F | ᵟ | MODIFIER LETTER SMALL DELTA |
| 03B8 | θ | GREEK SMALL LETTER THETA | 1DBF | ᶿ | MODIFIER LETTER SMALL THETA |
| 03B9 | ι | GREEK SMALL LETTER IOTA | 1DA5 | ᶥ | MODIFIER LETTER SMALL IOTA |
| 03C5 | υ | GREEK SMALL LETTER UPSILON | 1DB7 | ᶷ | MODIFIER LETTER SMALL UPSILON |
| 03C6 | φ | GREEK SMALL LETTER PHI | 1DB2 | ᶲ | MODIFIER LETTER SMALL PHI |
| 03C7 | χ | GREEK SMALL LETTER CHI | 1D61 | ᵡ | MODIFIER LETTER SMALL CHI |
| 0410 | А | CYRILLIC CAPITAL LETTER A | 1D2C | ᴬ | MODIFIER LETTER CAPITAL A |
| 0418 | И | CYRILLIC CAPITAL LETTER I | 1D35 | ᴵ | MODIFIER LETTER CAPITAL I |
| 041E | О | CYRILLIC CAPITAL LETTER O | 1D3C | ᴼ | MODIFIER LETTER CAPITAL O |
| 0423 | У | CYRILLIC CAPITAL LETTER U | 1D41 | ᵁ | MODIFIER LETTER CAPITAL U |
| 042D | Э | CYRILLIC CAPITAL LETTER E | 1D31 | ᴱ | MODIFIER LETTER CAPITAL E |
| 0430 | а | CYRILLIC SMALL LETTER A | 1D43 | ᵃ | MODIFIER LETTER SMALL A |
| 043E | о | CYRILLIC SMALL LETTER O | 1D52 | ᵒ | MODIFIER LETTER SMALL O |
| 0443 | у | CYRILLIC SMALL LETTER U | 1D58 | ᵘ | MODIFIER LETTER SMALL U |
| 044D | э | CYRILLIC SMALL LETTER E | 1D49 | ᵉ | MODIFIER LETTER SMALL E |
| 04D9 | ә | CYRILLIC SMALL LETTER SCHWA | 1D4A | ᵊ | MODIFIER LETTER SMALL SCHWA |
| 04E9 | ө | CYRILLIC SMALL LETTER BARRED O | 1DB1 | ᶱ | MODIFIER LETTER SMALL BARRED O |
| 1D02 | ᴂ | LATIN SMALL LETTER TURNED AE | 1D46 | ᵆ | MODIFIER LETTER SMALL TURNED AE |
| 1D08 | ᴈ | LATIN SMALL LETTER TURNED OPEN E | 1D4C | ᵌ | MODIFIER LETTER SMALL TURNED OPEN E |
| 1D09 | ᴉ | LATIN SMALL LETTER TURNED I | 1D4E | ᵎ | MODIFIER LETTER SMALL TURNED I |
| 1D16 | ᴖ | LATIN SMALL LETTER TOP HALF O | 1D54 | ᵔ | MODIFIER LETTER SMALL TOP HALF O |
| 1D17 | ᴗ | LATIN SMALL LETTER BOTTOM HALF O | 1D55 | ᵕ | MODIFIER LETTER SMALL BOTTOM HALF O |
| 1D1C | ᴜ | LATIN LETTER SMALL CAPITAL U | 1DB8 | ᶸ | MODIFIER LETTER SMALL CAPITAL U |
| 1D1D | ᴝ | LATIN SMALL LETTER SIDEWAYS U | 1D59 | ᵙ | MODIFIER LETTER SMALL SIDEWAYS U |
| 1D7B | ᵻ | LATIN SMALL CAPITAL LETTER I WITH STROKE | 1DA7 | ᶧ | MODIFIER LETTER SMALL CAPITAL I WITH STROKE |
| 1D85 | ᶅ | LATIN SMALL LETTER L WITH PALATAL HOOK | 1DAA | ᶪ | MODIFIER LETTER SMALL L WITH PALATAL HOOK |
| 1D91 | ᶑ | LATIN SMALL LETTER D WITH HOOK AND TAIL | 1078D | 𐞍 | MODIFIER LETTER SMALL D WITH HOOK AND TAIL |
| 1E9F | ẟ | LATIN SMALL LETTER DELTA | 1D5F | ᵟ | MODIFIER LETTER SMALL DELTA |
| 2C71 | ⱱ | LATIN SMALL LETTER V WITH RIGHT HOOK | 107B0 | 𐞰 | MODIFIER LETTER SMALL V WITH RIGHT HOOK |
| 3041 | ぁ | HIRAGANA LETTER SMALL A | 1D43 | ᵃ | MODIFIER LETTER SMALL A |
| 3045 | ぅ | HIRAGANA LETTER SMALL U | 1D58 | ᵘ | MODIFIER LETTER SMALL U |
| 3047 | ぇ | HIRAGANA LETTER SMALL E | 1D49 | ᵉ | MODIFIER LETTER SMALL E |
| 3049 | ぉ | HIRAGANA LETTER SMALL O | 1D52 | ᵒ | MODIFIER LETTER SMALL O |
| 30A1 | ァ | KATAKANA LETTER SMALL A | 1D43 | ᵃ | MODIFIER LETTER SMALL A |
| 30A5 | ゥ | KATAKANA LETTER SMALL U | 1D58 | ᵘ | MODIFIER LETTER SMALL U |
| 30A7 | ェ | KATAKANA LETTER SMALL E | 1D49 | ᵉ | MODIFIER LETTER SMALL E |
| 30A9 | ォ | KATAKANA LETTER SMALL O | 1D52 | ᵒ | MODIFIER LETTER SMALL O |
| A727 | ꜧ | LATIN SMALL LETTER HENG | AB5C | ꭜ | MODIFIER LETTER SMALL HENG |
| A78E | ꞎ | LATIN SMALL LETTER L WITH RETROFLEX HOOK AND BELT | 1079D | 𐞝 | MODIFIER LETTER SMALL L WITH RETROFLEX HOOK AND BELT |
| A7B5 | ꞵ | LATIN SMALL LETTER BETA | 1D5D | ᵝ | MODIFIER LETTER SMALL BETA |
| A856 | ꡖ | PHAGS-PA LETTER SMALL A | 1D43 | ᵃ | MODIFIER LETTER SMALL A |
| AB37 | ꬷ | LATIN SMALL LETTER L WITH INVERTED LAZY S | AB5D | ꭝ | MODIFIER LETTER SMALL L WITH INVERTED LAZY S |
| AB52 | ꭒ | LATIN SMALL LETTER U WITH LEFT HOOK | AB5F | ꭟ | MODIFIER LETTER SMALL U WITH LEFT HOOK |
| AB53 | ꭓ | LATIN SMALL LETTER CHI | 1D61 | ᵡ | MODIFIER LETTER SMALL CHI |
| AB66 | ꭦ | LATIN SMALL LETTER DZ DIGRAPH WITH RETROFLEX HOOK | 10788 | 𐞈 | MODIFIER LETTER SMALL DZ DIGRAPH WITH RETROFLEX HOOK |
| AB67 | ꭧ | LATIN SMALL LETTER TS DIGRAPH WITH RETROFLEX HOOK | 107AD | 𐞭 | MODIFIER LETTER SMALL TS DIGRAPH WITH RETROFLEX HOOK |

Niektóre modyfikatory pochodzą od symboli (tab. 16).

Tab. . Zestawienie symboli bazowych i ich modyfikujących odpowiedników

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Symbol bazowy | | | Modyfikatory | | | Pozycja |
| 0021 | ! | EXCLAMATION MARK | A71D | ꜝ | MODIFIER LETTER RAISED EXCLAMATION MARK | RAISED |
| 0027 | ' | APOSTROPHE | 02BC | ʼ | MODIFIER LETTER APOSTROPHE |  |
| 002B | + | PLUS SIGN | 02D6 | ˖ | MODIFIER LETTER PLUS SIGN |  |
| 003A | : | COLON | 02F8 | ˸ | MODIFIER LETTER RAISED COLON | RAISED |
| 003A | : | COLON | A789 | ꞉ | MODIFIER LETTER COLON |  |
| 005E | ^ | CIRCUMFLEX ACCENT | 02C6 | ˆ | MODIFIER LETTER CIRCUMFLEX ACCENT |  |
| 005E | ^ | CIRCUMFLEX ACCENT | A788 | ꞈ | MODIFIER LETTER LOW CIRCUMFLEX ACCENT | LOW |
| 0060 | ` | GRAVE ACCENT | 02CE | ˎ | MODIFIER LETTER LOW GRAVE ACCENT | LOW |
| 0060 | ` | GRAVE ACCENT | 02F4 | ˴ | MODIFIER LETTER MIDDLE GRAVE ACCENT | MIDDLE |
| 007C | | | VERTICAL LINE | 02C8 | ˈ | MODIFIER LETTER VERTICAL LINE |  |
| 007C | | | VERTICAL LINE | 02CC | ˌ | MODIFIER LETTER LOW VERTICAL LINE | LOW |
| 007E | ~ | TILDE | 02F7 | ˷ | MODIFIER LETTER LOW TILDE | LOW |
| 00A1 | ¡ | INVERTED EXCLAMATION MARK | A71E | ꜞ | MODIFIER LETTER RAISED INVERTED EXCLAMATION MARK | RAISED |
| 00A1 | ¡ | INVERTED EXCLAMATION MARK | A71F | ꜟ | MODIFIER LETTER LOW INVERTED EXCLAMATION MARK | LOW |
| 00AF | ¯ | MACRON | 02C9 | ˉ | MODIFIER LETTER MACRON |  |
| 00AF | ¯ | MACRON | 02CD | ˍ | MODIFIER LETTER LOW MACRON | LOW |
| 00B4 | ´ | ACUTE ACCENT | 02CF | ˏ | MODIFIER LETTER LOW ACUTE ACCENT | LOW |
| 02DD | ˝ | DOUBLE ACUTE ACCENT | 02F6 | ˶ | MODIFIER LETTER MIDDLE DOUBLE ACUTE ACCENT | MIDDLE |
| 2032 | ′ | PRIME | 02B9 | ʹ | MODIFIER LETTER PRIME |  |
| 2033 | ″ | DOUBLE PRIME | 02BA | ʺ | MODIFIER LETTER DOUBLE PRIME |  |
| 2190 | ← | LEFTWARDS ARROW | 02FF | ˿ | MODIFIER LETTER LOW LEFT ARROW | LOW |
| 2212 | − | MINUS SIGN | 02D7 | ˗ | MODIFIER LETTER MINUS SIGN |  |
| 22A2 | ⊢ | RIGHT TACK | AB6B | ꭫ | MODIFIER LETTER RIGHT TACK |  |
| 22A3 | ⊣ | LEFT TACK | AB6A | ꭪ | MODIFIER LETTER LEFT TACK |  |
| 22A4 | ⊤ | DOWN TACK | 02D5 | ˕ | MODIFIER LETTER DOWN TACK |  |
| 22A5 | ⊥ | UP TACK | 02D4 | ˔ | MODIFIER LETTER UP TACK |  |
| 2303 | ⌃ | UP ARROWHEAD | 02C4 | ˄ | MODIFIER LETTER UP ARROWHEAD |  |
| 2303 | ⌃ | UP ARROWHEAD | 02F0 | ˰ | MODIFIER LETTER LOW UP ARROWHEAD |  |
| 2304 | ⌄ | DOWN ARROWHEAD | 02C5 | ˅ | MODIFIER LETTER DOWN ARROWHEAD |  |
| 2304 | ⌄ | DOWN ARROWHEAD | 02EF | ˯ | MODIFIER LETTER LOW DOWN ARROWHEAD | LOW |
| 2E32 | ⸲ | TURNED COMMA | 02BB | ʻ | MODIFIER LETTER TURNED COMMA |  |
| 2E41 | ⹁ | REVERSED COMMA | 02BD | ʽ | MODIFIER LETTER REVERSED COMMA |  |

Pozostałe modyfikatory nie mają symboli bazowych, ich nazwy niekoniecznie rozpoczynają się od MODIFIER LETTER, ale zawsze zawierają słowo kluczowe MODIFIER (tab. 17).

Tab. . Modyfikatory bez symboli bazowych

|  |  |  |
| --- | --- | --- |
| Kod | Glif | Modyfikator |
| 02BE | ʾ | MODIFIER LETTER RIGHT HALF RING |
| 02BF | ʿ | MODIFIER LETTER LEFT HALF RING |
| 02C1 | ˁ | MODIFIER LETTER REVERSED GLOTTAL STOP |
| 02C2 | ˂ | MODIFIER LETTER LEFT ARROWHEAD |
| 02C3 | ˃ | MODIFIER LETTER RIGHT ARROWHEAD |
| 02CA | ˊ | MODIFIER LETTER ACUTE ACCENT |
| 02CB | ˋ | MODIFIER LETTER GRAVE ACCENT |
| 02D0 | ː | MODIFIER LETTER TRIANGULAR COLON |
| 02D1 | ˑ | MODIFIER LETTER HALF TRIANGULAR COLON |
| 02D2 | ˒ | MODIFIER LETTER CENTRED RIGHT HALF RING |
| 02D3 | ˓ | MODIFIER LETTER CENTRED LEFT HALF RING |
| 02DE | ˞ | MODIFIER LETTER RHOTIC HOOK |
| 02DF | ˟ | MODIFIER LETTER CROSS ACCENT |
| 02E4 | ˤ | MODIFIER LETTER SMALL REVERSED GLOTTAL STOP |
| 02E5 | ˥ | MODIFIER LETTER EXTRA-HIGH TONE BAR |
| 02E6 | ˦ | MODIFIER LETTER HIGH TONE BAR |
| 02E7 | ˧ | MODIFIER LETTER MID TONE BAR |
| 02E8 | ˨ | MODIFIER LETTER LOW TONE BAR |
| 02E9 | ˩ | MODIFIER LETTER EXTRA-LOW TONE BAR |
| 02EA | ˪ | MODIFIER LETTER YIN DEPARTING TONE MARK |
| 02EB | ˫ | MODIFIER LETTER YANG DEPARTING TONE MARK |
| 02EC | ˬ | MODIFIER LETTER VOICING |
| 02ED | ˭ | MODIFIER LETTER UNASPIRATED |
| 02EE | ˮ | MODIFIER LETTER DOUBLE APOSTROPHE |
| 02F1 | ˱ | MODIFIER LETTER LOW LEFT ARROWHEAD |
| 02F2 | ˲ | MODIFIER LETTER LOW RIGHT ARROWHEAD |
| 02F3 | ˳ | MODIFIER LETTER LOW RING |
| 02F5 | ˵ | MODIFIER LETTER MIDDLE DOUBLE GRAVE ACCENT |
| 02F9 | ˹ | MODIFIER LETTER BEGIN HIGH TONE |
| 02FA | ˺ | MODIFIER LETTER END HIGH TONE |
| 02FB | ˻ | MODIFIER LETTER BEGIN LOW TONE |
| 02FC | ˼ | MODIFIER LETTER END LOW TONE |
| 02FD | ˽ | MODIFIER LETTER SHELF |
| 02FE | ˾ | MODIFIER LETTER OPEN SHELF |
| 0559 | ՙ | ARMENIAN MODIFIER LETTER LEFT HALF RING |
| 081A | ࠚ | SAMARITAN MODIFIER LETTER EPENTHETIC YUT |
| 0824 | ࠤ | SAMARITAN MODIFIER LETTER SHORT A |
| 0828 | ࠨ | SAMARITAN MODIFIER LETTER I |
| 10FC | ჼ | MODIFIER LETTER GEORGIAN NAR |
| 1D2F | ᴯ | MODIFIER LETTER CAPITAL BARRED B |
| 1D3B | ᴻ | MODIFIER LETTER CAPITAL REVERSED N |
| 1D5E | ᵞ | MODIFIER LETTER SMALL GREEK GAMMA |
| 1D60 | ᵠ | MODIFIER LETTER SMALL GREEK PHI |
| 1D78 | ᵸ | MODIFIER LETTER CYRILLIC EN |
| 2D6F | ⵯ | TIFINAGH MODIFIER LETTER LABIALIZATION MARK |
| A69C | ꚜ | MODIFIER LETTER CYRILLIC HARD SIGN |
| A69D | ꚝ | MODIFIER LETTER CYRILLIC SOFT SIGN |
| A700 | ꜀ | MODIFIER LETTER CHINESE TONE YIN PING |
| A701 | ꜁ | MODIFIER LETTER CHINESE TONE YANG PING |
| A702 | ꜂ | MODIFIER LETTER CHINESE TONE YIN SHANG |
| A703 | ꜃ | MODIFIER LETTER CHINESE TONE YANG SHANG |
| A704 | ꜄ | MODIFIER LETTER CHINESE TONE YIN QU |
| A705 | ꜅ | MODIFIER LETTER CHINESE TONE YANG QU |
| A706 | ꜆ | MODIFIER LETTER CHINESE TONE YIN RU |
| A707 | ꜇ | MODIFIER LETTER CHINESE TONE YANG RU |
| A708 | ꜈ | MODIFIER LETTER EXTRA-HIGH DOTTED TONE BAR |
| A709 | ꜉ | MODIFIER LETTER HIGH DOTTED TONE BAR |
| A70A | ꜊ | MODIFIER LETTER MID DOTTED TONE BAR |
| A70B | ꜋ | MODIFIER LETTER LOW DOTTED TONE BAR |
| A70C | ꜌ | MODIFIER LETTER EXTRA-LOW DOTTED TONE BAR |
| A70D | ꜍ | MODIFIER LETTER EXTRA-HIGH DOTTED LEFT-STEM TONE BAR |
| A70E | ꜎ | MODIFIER LETTER HIGH DOTTED LEFT-STEM TONE BAR |
| A70F | ꜏ | MODIFIER LETTER MID DOTTED LEFT-STEM TONE BAR |
| A710 | ꜐ | MODIFIER LETTER LOW DOTTED LEFT-STEM TONE BAR |
| A711 | ꜑ | MODIFIER LETTER EXTRA-LOW DOTTED LEFT-STEM TONE BAR |
| A712 | ꜒ | MODIFIER LETTER EXTRA-HIGH LEFT-STEM TONE BAR |
| A713 | ꜓ | MODIFIER LETTER HIGH LEFT-STEM TONE BAR |
| A714 | ꜔ | MODIFIER LETTER MID LEFT-STEM TONE BAR |
| A715 | ꜕ | MODIFIER LETTER LOW LEFT-STEM TONE BAR |
| A716 | ꜖ | MODIFIER LETTER EXTRA-LOW LEFT-STEM TONE BAR |
| A717 | ꜗ | MODIFIER LETTER DOT VERTICAL BAR |
| A718 | ꜘ | MODIFIER LETTER DOT SLASH |
| A719 | ꜙ | MODIFIER LETTER DOT HORIZONTAL BAR |
| A71A | ꜚ | MODIFIER LETTER LOWER RIGHT CORNER ANGLE |
| A71B | ꜛ | MODIFIER LETTER RAISED UP ARROW |
| A71C | ꜜ | MODIFIER LETTER RAISED DOWN ARROW |
| A720 | ꜠ | MODIFIER LETTER STRESS AND HIGH TONE |
| A721 | ꜡ | MODIFIER LETTER STRESS AND LOW TONE |
| A770 | ꝰ | MODIFIER LETTER US |
| A78A | ꞊ | MODIFIER LETTER SHORT EQUALS SIGN |
| A9E6 | ꧦ | MYANMAR MODIFIER LETTER SHAN REDUPLICATION |
| AA70 | ꩰ | MYANMAR MODIFIER LETTER KHAMTI REDUPLICATION |
| AB5B | ꭛ | MODIFIER BREVE WITH INVERTED BREVE |
| 10780 | 𐞀 | MODIFIER LETTER SMALL CAPITAL AA |
| 10781 | 𐞁 | MODIFIER LETTER SUPERSCRIPT TRIANGULAR COLON |
| 10782 | 𐞂 | MODIFIER LETTER SUPERSCRIPT HALF TRIANGULAR COLON |
| 1079C | 𐞜 | MODIFIER LETTER SMALL CAPITAL L WITH BELT |
| 1079F | 𐞟 | MODIFIER LETTER SMALL LEZH WITH RETROFLEX HOOK |
| 107A1 | 𐞡 | MODIFIER LETTER SMALL TURNED Y WITH BELT |
| 107A7 | 𐞧 | MODIFIER LETTER SMALL TURNED R WITH LONG LEG AND RETROFLEX HOOK |
| 107B9 | 𐞹 | MODIFIER LETTER RETROFLEX CLICK WITH RETROFLEX HOOK |
| 107BA | 𐞺 | MODIFIER LETTER SMALL S WITH CURL |
| 111CB | ◌ | SHARADA VOWEL MODIFIER MARK |
| 13447 | ◌ | EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP START |
| 13448 | ◌ | EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT BOTTOM START |
| 13449 | ◌ | EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT START |
| 1344A | ◌ | EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP END |
| 1344B | ◌ | EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP |
| 1344C | ◌ | EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT BOTTOM START AND TOP END |
| 1344D | ◌ | EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT START AND TOP |
| 1344E | ◌ | EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT BOTTOM END |
| 1344F | ◌ | EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP START AND BOTTOM END |
| 13450 | ◌ | EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT BOTTOM |
| 13451 | ◌ | EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT START AND BOTTOM |
| 13452 | ◌ | EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT END |
| 13453 | ◌ | EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP AND END |
| 13454 | ◌ | EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT BOTTOM AND END |
| 13455 | ◌ | EGYPTIAN HIEROGLYPH MODIFIER DAMAGED |
| 16F4F | ◌ | MIAO SIGN CONSONANT MODIFIER BAR |
| 1CF42 | ◌ | ZNAMENNY PRIZNAK MODIFIER LEVEL-2 |
| 1CF43 | ◌ | ZNAMENNY PRIZNAK MODIFIER LEVEL-3 |
| 1CF44 | ◌ | ZNAMENNY PRIZNAK MODIFIER DIRECTION FLIP |
| 1CF45 | ◌ | ZNAMENNY PRIZNAK MODIFIER KRYZH |
| 1CF46 | ◌ | ZNAMENNY PRIZNAK MODIFIER ROG |
| 1DA9B | ◌ | SIGNWRITING FILL MODIFIER-2 |
| 1DA9C | ◌ | SIGNWRITING FILL MODIFIER-3 |
| 1DA9D | ◌ | SIGNWRITING FILL MODIFIER-4 |
| 1DA9E | ◌ | SIGNWRITING FILL MODIFIER-5 |
| 1DA9F | ◌ | SIGNWRITING FILL MODIFIER-6 |
| 1DAA1 | ◌ | SIGNWRITING ROTATION MODIFIER-2 |
| 1DAA2 | ◌ | SIGNWRITING ROTATION MODIFIER-3 |
| 1DAA3 | ◌ | SIGNWRITING ROTATION MODIFIER-4 |
| 1DAA4 | ◌ | SIGNWRITING ROTATION MODIFIER-5 |
| 1DAA5 | ◌ | SIGNWRITING ROTATION MODIFIER-6 |
| 1DAA6 | ◌ | SIGNWRITING ROTATION MODIFIER-7 |
| 1DAA7 | ◌ | SIGNWRITING ROTATION MODIFIER-8 |
| 1DAA8 | ◌ | SIGNWRITING ROTATION MODIFIER-9 |
| 1DAA9 | ◌ | SIGNWRITING ROTATION MODIFIER-10 |
| 1DAAA | ◌ | SIGNWRITING ROTATION MODIFIER-11 |
| 1DAAB | ◌ | SIGNWRITING ROTATION MODIFIER-12 |
| 1DAAC | ◌ | SIGNWRITING ROTATION MODIFIER-13 |
| 1DAAD | ◌ | SIGNWRITING ROTATION MODIFIER-14 |
| 1DAAE | ◌ | SIGNWRITING ROTATION MODIFIER-15 |
| 1DAAF | ◌ | SIGNWRITING ROTATION MODIFIER-16 |
| 1E030 | 𞀰 | MODIFIER LETTER CYRILLIC SMALL A |
| 1E031 | 𞀱 | MODIFIER LETTER CYRILLIC SMALL BE |
| 1E032 | 𞀲 | MODIFIER LETTER CYRILLIC SMALL VE |
| 1E033 | 𞀳 | MODIFIER LETTER CYRILLIC SMALL GHE |
| 1E034 | 𞀴 | MODIFIER LETTER CYRILLIC SMALL DE |
| 1E035 | 𞀵 | MODIFIER LETTER CYRILLIC SMALL IE |
| 1E036 | 𞀶 | MODIFIER LETTER CYRILLIC SMALL ZHE |
| 1E037 | 𞀷 | MODIFIER LETTER CYRILLIC SMALL ZE |
| 1E038 | 𞀸 | MODIFIER LETTER CYRILLIC SMALL I |
| 1E039 | 𞀹 | MODIFIER LETTER CYRILLIC SMALL KA |
| 1E03A | 𞀺 | MODIFIER LETTER CYRILLIC SMALL EL |
| 1E03B | 𞀻 | MODIFIER LETTER CYRILLIC SMALL EM |
| 1E03C | 𞀼 | MODIFIER LETTER CYRILLIC SMALL O |
| 1E03D | 𞀽 | MODIFIER LETTER CYRILLIC SMALL PE |
| 1E03E | 𞀾 | MODIFIER LETTER CYRILLIC SMALL ER |
| 1E03F | 𞀿 | MODIFIER LETTER CYRILLIC SMALL ES |
| 1E040 | 𞁀 | MODIFIER LETTER CYRILLIC SMALL TE |
| 1E041 | 𞁁 | MODIFIER LETTER CYRILLIC SMALL U |
| 1E042 | 𞁂 | MODIFIER LETTER CYRILLIC SMALL EF |
| 1E043 | 𞁃 | MODIFIER LETTER CYRILLIC SMALL HA |
| 1E044 | 𞁄 | MODIFIER LETTER CYRILLIC SMALL TSE |
| 1E045 | 𞁅 | MODIFIER LETTER CYRILLIC SMALL CHE |
| 1E046 | 𞁆 | MODIFIER LETTER CYRILLIC SMALL SHA |
| 1E047 | 𞁇 | MODIFIER LETTER CYRILLIC SMALL YERU |
| 1E048 | 𞁈 | MODIFIER LETTER CYRILLIC SMALL E |
| 1E049 | 𞁉 | MODIFIER LETTER CYRILLIC SMALL YU |
| 1E04A | 𞁊 | MODIFIER LETTER CYRILLIC SMALL DZZE |
| 1E04B | 𞁋 | MODIFIER LETTER CYRILLIC SMALL SCHWA |
| 1E04C | 𞁌 | MODIFIER LETTER CYRILLIC SMALL BYELORUSSIAN-UKRAINIAN I |
| 1E04D | 𞁍 | MODIFIER LETTER CYRILLIC SMALL JE |
| 1E04E | 𞁎 | MODIFIER LETTER CYRILLIC SMALL BARRED O |
| 1E04F | 𞁏 | MODIFIER LETTER CYRILLIC SMALL STRAIGHT U |
| 1E050 | 𞁐 | MODIFIER LETTER CYRILLIC SMALL PALOCHKA |
| 1E06B | 𞁫 | MODIFIER LETTER CYRILLIC SMALL ES WITH DESCENDER |
| 1E06C | 𞁬 | MODIFIER LETTER CYRILLIC SMALL YERU WITH BACK YER |
| 1E06D | 𞁭 | MODIFIER LETTER CYRILLIC SMALL STRAIGHT U WITH STROKE |
| 1E948 | ◌ | ADLAM CONSONANT MODIFIER |
| 1E949 | ◌ | ADLAM GEMINATE CONSONANT MODIFIER |
| 1F3FB | 🏻 | EMOJI MODIFIER FITZPATRICK TYPE-1-2 |
| 1F3FC | 🏼 | EMOJI MODIFIER FITZPATRICK TYPE-3 |
| 1F3FD | 🏽 | EMOJI MODIFIER FITZPATRICK TYPE-4 |
| 1F3FE | 🏾 | EMOJI MODIFIER FITZPATRICK TYPE-5 |
| 1F3FF | 🏿 | EMOJI MODIFIER FITZPATRICK TYPE-6 |

W nazwach opisowych znaków kombinatorycznych i modyfikatorów mogą występować frazy identyfikujące pismo, z którego jest brany znak bazowy. Przedstawia to tab. 18.

Tab. . Identyfikowanie pisma w modyfikatorach i znakach kombinatorycznych Unicode

|  |  |  |  |
| --- | --- | --- | --- |
| Pismo | Fraza kluczowa | Kat. | Uwagi |
| Common | MODIFIER LETTER\* | Sk, Lm |  |
| Common | MODIFIER\* | Sk |  |
| Common | COMBINING\* | Mn, Me |  |
| Latin | COMBINING LATIN\* | Mn |  |
| Cyrillic | COMBINING CYRILLIC\* | Mn, Me |  |
| Cyrillic | MODIFIER LETTER CYRILLIC\* | Lm |  |
| Armenian | ARMENIAN MODIFIER LETTER\* | Sk |  |
| Nko | NKO COMBINING\* | Mn |  |
| Samaritan | SAMARITAN MODIFIER LETTER\* | Lm |  |
| Telugu | TELUGU SIGN COMBINING\* | Mn |  |
| Kannada | KANNADA SIGN COMBINING\* | Mn |  |
| Malayam | MALAYALAM SIGN COMBINING\* | Mn |  |
| Georgian | MODIFIER LETTER GEORGIAN\* | Sk |  |
| Ethiopic | ETHIOPIC COMBINING\* | Mn |  |
| Tai Tham | TAI THAM COMBINING\* | Mn |  |
| Balinese | BALINESE MUSICAL SYMBOL COMBINING\* | Mn |  |
| Coptic | COPTIC COMBINING\* | Mn |  |
| Tiffinagh | TIFFINAGH MODIFIED\* | Lm |  |
| Katakana, Hiragana | COMBINING KATAKANA-HIRAGANA\* | Mn |  |
| Bamum | BAMUM COMBINING\* | Mn |  |
| Han | MODIFIER LETTER CHINESE\* | Sk |  |
| Devanagari | COMBINING DEVANAGARI\* | Mn |  |
| Myanmar | MYANMAR MODIFIER LETTER\* | Lm |  |
| Phaistos Disc | PHAISTOS DISC SIGN COMBINING\* | Mn |  |
| Old Permic | COMBINING OLD PERMIC\* | Mn |  |
| Garay | GARAY COMBINING\* | Mn |  |
| Yezidi | YEZIDI COMBINING\* | Mn |  |
| Arabic | ARABIC COMBINING\* | Mn |  |
| Sogdian | SOGDIAN COMBINING\* | Mn |  |
| Old Uyghur | OLD UYGHUR COMBINING\* | Mn |  |
| Sharada | SHARADA VOWEL MODIFIER MARK\* | Mn |  |
| Grantha | GRANTHA SIGN COMBINING\* | Mn |  |
| Grantha | COMBINING GRANTHA\* | Mn |  |
| Bindu | COMBINING BINDU\* | Mn |  |
| Egyptian Hieroglyphs | EGYPTIAN HIEROGLYPH MODIFIER\* | Mn |  |
| Bassa Vah | BASSA VAH COMBINING\* | Mn |  |
| Miao | MIAO SIGN CONSONANT MODIFIER BAR\* | Mn |  |
| Znamenny Chant | ZNAMENNY COMBINING\* | Mn |  |
| Znamenny Chant | ZNAMENNY PRIZNAK MODIFIER\* | Mn |  |
| Musical Notation | MUSICAL SYMBOL COMBINING\* | Mc |  |
| Greek Ancient Instrumental Notation | COMBINING GREEK MUSICAL\* | Mn |  |
| SignWriting | SIGNWRITING \* MODIFIER\* | Mn | a) |
| Glagolitic | COMBINING GLAGOLITIC\* | Mn |  |
| Mende Kikakui | MENDE KIKAKUI COMBINING NUMBER\* | Mn |  |
| Adlam | ADLAM \* MODIFIER\* | Mn | b) |
| Emoji | EMOJI MODIFIER FITZPATRICK\* | Sk |  |

Uwagi:

1. Modyfikatory pisma migowego mają określenie typu pomiędzy słowami kluczowymi SIGNWRITING i MODIFIER. Może to być FILL albo ROTATION. Pierwsze oznacza wypełnienie, drugie – obrót. „Wypełnienie” w SignWriting oznacza widok grzbietu lub wnętrza dłoni dla osoby piszącej w SignWriting. Czarne wypełnienie oznacza, że osoba pisząca widzi grzbiet dłoni; białe wypełnienie – wnętrze dłoni; a półwypełnienie – bok dłoni. Istnieją też wypełnienia, które wskazują, kiedy dłoń znajduje się w płaszczyźnie pionowej i trzy, które wskazują, kiedy dłoń znajduje się w płaszczyźnie poziomej. „Obrót” oznacza jedną z szesnastu pozycji, w których może pojawić się kształt dłoni. Służy również do odróżnienia znaku lewostronnego od prawostronnego. Modyfikatory pisma migowego są w Unicode rozróżniane przez końcową cyfrę (np. SIGNWRITING FILL MODIFIER-2). Program SignWriting korzysta z trypletów znak bazowy, modyfikator wypełnienia, modyfikator obrotu. Pełną listę prawidłowych trypletów można znaleźć w pliku opublikowanym pod adresem http://www.evertype.com/standards/iso10646/pdf/signwritingsequences.txt.
2. Modyfikatory pisma ADLAM mają określenie typu pomiędzy ADLAM i MODIFIER. Może to być CONSONANT albo GEMINATE CONSONANT. Pierwsze oznacza modyfikator spółgłoski, a drugie – modyfikator spółgłoski podwojonej (jak w wyrazie „manna”) lub wydłużonej.

## Litery

Znaki spełniające role liter należą do kategorii Lu (litery wielkie), Ll (litery małe) lub Lo (inne litery). Jest jeszcze kategoria Lt, ale ona oznacza nie pojedyncze litery, lecz ligatury (złożenia) litery wielkiej i małej (patrz podrozdział ‎3.12). Wiele liter ma nazwie opisowej słowo kluczowe „LETTER”, ale nie jest to konieczne. Przykłady:

GREEK LUNATE SIGMA SYMBOL

ARABIC VERTICAL TAIL

DEVANAGARI OM

Litery wielkie mają często w nazwie opisowej słowo kluczowe „CAPITAL”, a litery małe „SMALL”, ale to też nie jest konieczne. Przykłady:

GREEK CAPITAL KAI SYMBOL

MICRO SIGN

Wiele liter ma nazwy własne jednoliterowe, np.:

LATIN CAPITAL LETTER A

CYRILLIC SMALL LETTER A,

ale część ma nazwy dłuższe, np.:

LATIN CAPITAL LETTER SCHWA

## Litery z symbolami diakrytycznymi

Wiele liter (i innych symboli) występuje w postaci zmodyfikowanej z już dodanymi symbolami diakrytycznymi. Rozpoznaje się je po słowie kluczowym WITH w nazwie opisowej. Np. litera o nazwie opisowej LATIN CAPITAL LETTER A **WITH** BREVE ma glif wyglądający jak LATIN CAPITAL LETTER A z dodanym znakiem diakrytycznym COMBINING BREVE.

Ă = A + ◌̆

Litera może być wzbogacona o drugi, a nawet trzeci symbol diakrytyczny. Wówczas klauzula WITH ma dodane klauzule AND. Przykłady:

LATIN CAPITAL LETTER U **WITH** DIAERESIS **AND** MACRON

GREEK SMALL LETTER ALPHA **WITH** PSILI **AND** VARIA **AND** YPOGEGRAMMENI

Nie zawsze bazową literę da się rozpoznać po lewej części nazwy (przed WITH). Dla przykładu w znaku LATIN SMALL LETTER LAMBDA WITH STROKE lewa część nazwy to LATIN SMALL LETTER LAMBDA, ale w piśmie łaciński nie ma małej litery LAMBA, jest za to w piśmie greckim. Podobnie w znaku LATIN **LETTER** TWO **WITH** STROKE znakiem bazowym jest LATIN **DIGIT** TWO.

## Ligatury

Ligatura to złożenie dwóch lub więcej liter (lub innych symboli), które mogą występować jako osobne znaki w taki sposób, że tworzą jeden wspólny znak (najczęściej o mniejszej szerokości. Ligatury. Ligatury rozpoznaje się po słowie kluczowym LIGATURE. Przykład:

LATIN CAPITAL LIGATURE IJ

Jest ponad 500 takich znaków w Unicode, przy czym najwięcej ligatur jest w piśmie arabskim. Ligatury mogą, lecz nie muszą mieć klauzuli WITH w nazwie opisowej. W drugim przypadku WITH nie oznacza złożenia ze znakiem diakrytycznym. Przykłady:

ARABIC LIGATURE SALLALLAHOU ALAYHE WASALLAM

ARABIC SMALL HIGH LIGATURE ALEF WITH LAM WITH YEH

Niektóre ligatury nie mają w nazwie opisowej słowa LIGATURE, lecz LETTER. Przykład:

LATIN CAPITAL LETTER DZ

Ligatury, które są złożeniem litery wielkiej i małej, mają kategorię Lt (zamiast Lu lub Ll). Mają one odpowiednio złożoną nazwę opisową z klauzulą WITH. Przykład:

LATIN CAPITAL LETTER D WITH SMALL LETTER Z

Ligatura może być też połączona ze znakiem diakrytycznym, np.:

LATIN CAPITAL LETTER DZ WITH CARON ma glif ‘Ǆ’, a

LATIN CAPITAL LETTER D WITH SMALL LETTER Z WITH CARON ma glif ‘ǅ’.

## Sylaby

W sylabariuszach są znaki oznaczające całe sylaby. Mają one kategorię So oraz słowo kluczowe SYLLABLE w nazwie opisowej. Przykłady:

TIBETAN SYLLABLE OM

ETHIOPIC SYLLABLE HA

YI SYLLABLE AT

Nie ma sylab z klauzulą WITH.

## Samogłoski i spółgłoski

Samogłoski w sylabariuszach należą do kategorii Lo i mają w nazwie opisowej słowo kluczowe VOWEL lub frazę VOWEL SIGN. Przykłady:

DEVANAGARI VOWEL SIGN AA

TAI VIET VOWEL AA

Samogłoska może mieć dodatkowy przymiotnik INHERENT, który oznacza, że samogłoska musi być wymawiana, nawet jeśli nie jest zapisywana. Przykłady:

KHMER VOWEL INHERENT AQ

KHMER VOWEL INHERENT AA

Spółgłoski mają w nazwie opisowej słowo kluczowe CONSONANT lub frazę CONSONANT SIGN, chociaż należą do różnych kategorii (Mc, Mn, Lo). Przykłady:

TAI THAM CONSONANT SIGN MA

PAHAWH HMONG CONSONANT XAU

## Różne formy znakowe

Są systemy pisma, które mają różne formy znakowe, czyli różne kształty glifów. Pismo arabskie, tybetański i niektóre inne mają różne formy zależne od tego, czy znak występuje na początku wyrazu, w środku, czy na końcu. Takie formy są określane przez słowa kluczowe:

* INITIAL – dla formy początkowej,
* MEDIAL – dla formy środkowej,
* FINAL – dla formy końcowej,
* INDEPENDENT – dla formy samodzielnej,

Inne formy znakowe to:

* WIDE – dla szerokich form znaków,
* NARROW – dla wąskich form znaków,
* FULLWIDTH – dla znaków pionowych o pełnej szerokości
* HALFWIDTH – dla znaków pionowych o szerokości połówkowej,
* SMALL – dla znaków pomniejszonych,
* SMALL CAPITAL – dla pomniejszonych wielkich liter,
* BOLD – forma wytłuszczona,
* ITALIC – forma pochylona (ale OLD ITALIC oznacza stare pismo italskie),
* SANS-SERIF – forma bezszeryfowa (bez krótkich kresek na końcach segmentów),
* MONOSPACE – znaki o równej szerokości (w odróżnieniu od innych, które z reguły mają różne szerokości,
* SUPERSCRIPT – forma indeksu górnego (jak ‘²’),
* SUBSCRIPT – forma indeksu dolnego (jak ‘₂’),
* SCRIPT – forma przypominająca pismo ręczne (jak ‘ℛ’),
* BLACK-CAPITAL – dla form wielkich liter specjalnej czcionce (jak ‘ℜ’), tożsame z FRAKTUR,
* FRAKTUR – dla formy gotyckiej (jak ‘𝔯’),
* DOUBLE-STRUCK – forma z segmentami rysowanymi podwójną linią (jak ‘ℝ’),
* OUTLINED – forma konturowa (jak ‘’),
* CIRCLED – znak w kółku (jak ‘🄬’), ale ‘©’ to COPYRIGHT SIGN,
* NEGATIVE CIRCLED – znak w czarnym kółku (jak ‘🅡’),
* DOUBLE CIRCLED – znak w podwójnym kółku (jak ‘⓵’),
* SQUARED – znak w kwadracie (jak ‘🅁’), ale SQUARE może oznaczać kwadratową formę kilkuliterowego skrótu jednostki (np. ‘㎒’)
* NEGATIVE SQUARED – znak w czarnym kwadracie (jak ‘🆁’),
* PARENTHESIZED – znak w nawiasach (jak ‘🄡’),
* FULL STOP (dotyczy tylko cyfr i liczb od 10 do 20, występuje tylko na końcu nazwy opisowej) – znak z kropką na końcu (np. ‘⒑’),
* SEGMENTED (dotyczy tylko cyfr) – forma jak na wyświetlaczach LED (jak ‘’).
* WHITE – biała forma znaku (który może też być czarny),
* BLACK – czarna forma znaku (który może też być biały),
* NEUTRAL (tylko dla symboli figur szachowych) – biało-czarna forma znaku,
* HEAVY – forma pogrubiona (dotyczy symboli graficznych, w tym strzałek).

Niektóre znaki mają wygląd przypominający inne znaki z graficznym przekształceniem:

* TURNED – obrót o 180 stopni (jak ‘ᴚ’),
* REVERSED (czasami REVERSE) – odbicie lustrzane w poziomie (jak ‘ᴙ’),
* INVERTED – lustrzane odbicie w pionie (jak ‘ʁ’).

Pozostałe przymiotniki nie oznaczają innej formy, a po prostu nazwę znaku. Przykłady:

* SMALL OPEN O – to ‘ɔ’,
* SMALL OPEN E – to ‘ɛ’,
* SMALL CLOSED OPEN E – ‘ʚ’,
* OPEN BOX – ‘␣’,
* CLOCKWISE OPEN CIRCLE ARROW – ‘↻’,
* OPEN SUBSET – ‘⟃’,
* OBLIQUE ANGLE OPENING UP – ‘⦦’.