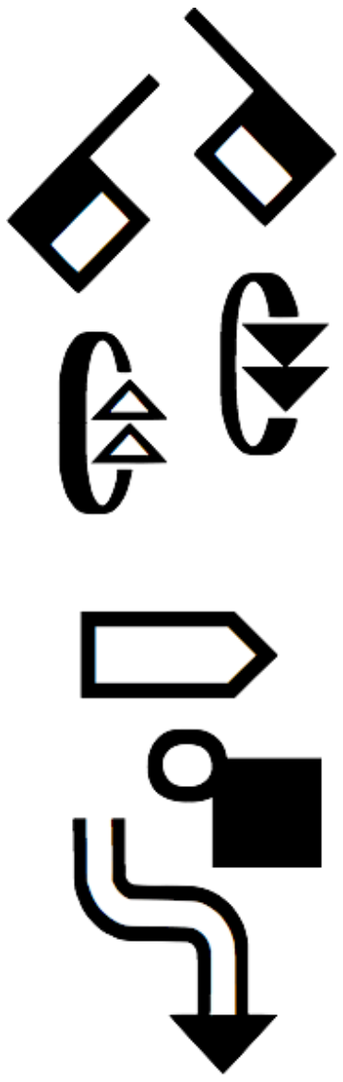
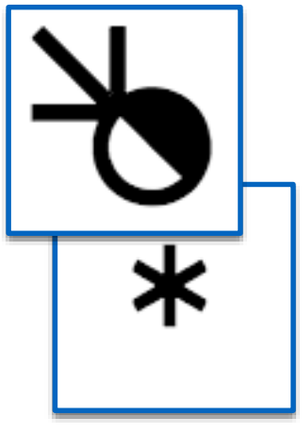


SignWriting in Unicode dot SWU

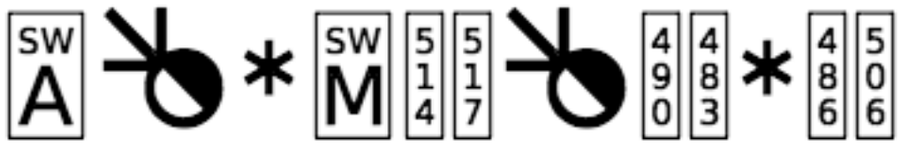
Prepared for the
SignWriting Symposium 2017

by Stephen E Slevinski Jr
in association with the Center for Sutton Movement Writing



SignWriting in Unicode

SWU



Experimental Unicode design

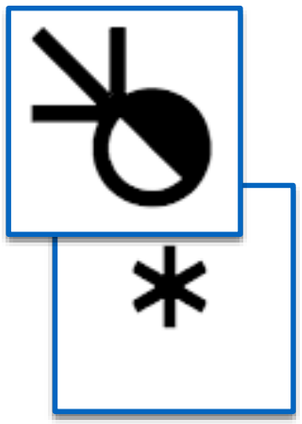
One-Dimensional Fonts available

Two-Dimensional Font being developed

Endorsed by the Center for Sutton Movement Writing

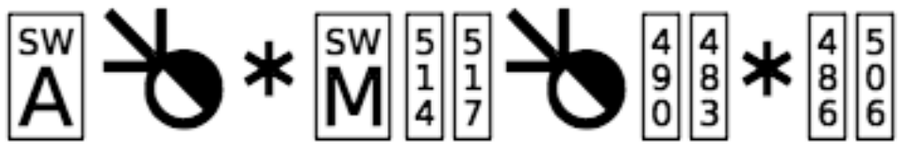
Submitted to the UTC July 2017

http://www.unicode.org/L2/L2017/17220-signwriting-design-opt.pdf



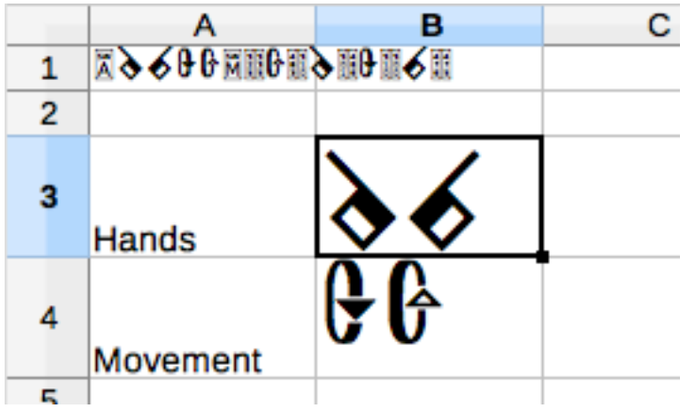
SignWriting in Unicode

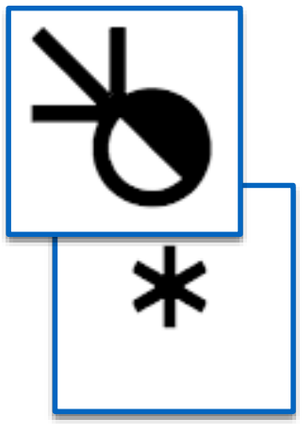
One-Dimensional Font



The Sutton SignWriting One-D font makes it possible to use SWU on a variety of operating systems and across applications with a visual representation rather than data.

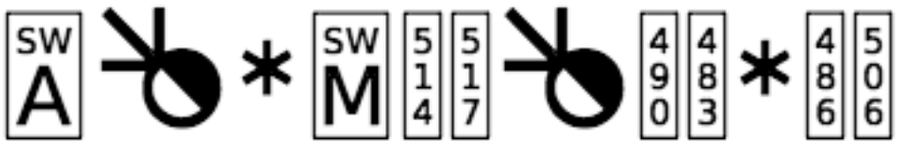
font-family: “SuttonSignWritingOneD”;





SignWriting in Unicode

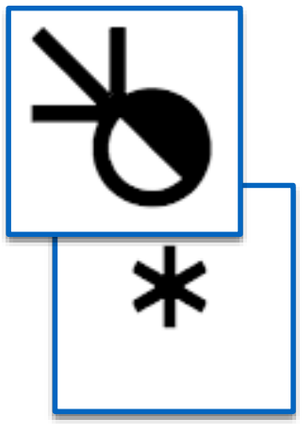
One-Dimensional Font



The Sutton SignWriting One-D font is available for download and installation for a variety of operating systems.

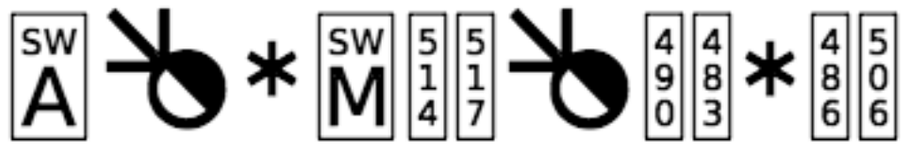
https://slevinski.github.io/SuttonSignWriting/components/fonts.html

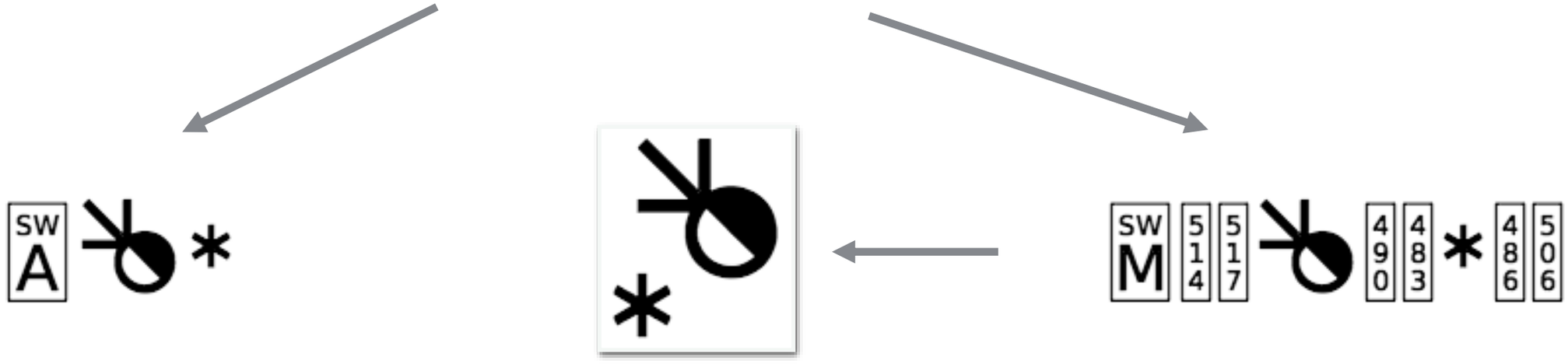
https://github.com/slevinski/signwriting_2010_fonts



SignWriting as Text

Two-Dimensional Font





The Sutton SignWriting Two-D font development is planned for 2018. This font uses the same SWU characters to correctly display signs as two-dimensional clusters across operating systems and applications.

https://meta.wikimedia.org/wiki/Grants:Project/slevinski/ASL_Wikipedia_2-D_Font_Development_for_SignWriting

Supplemental

Background

Sutton SignWriting
Formal SignWriting
Unicode Standard

Details

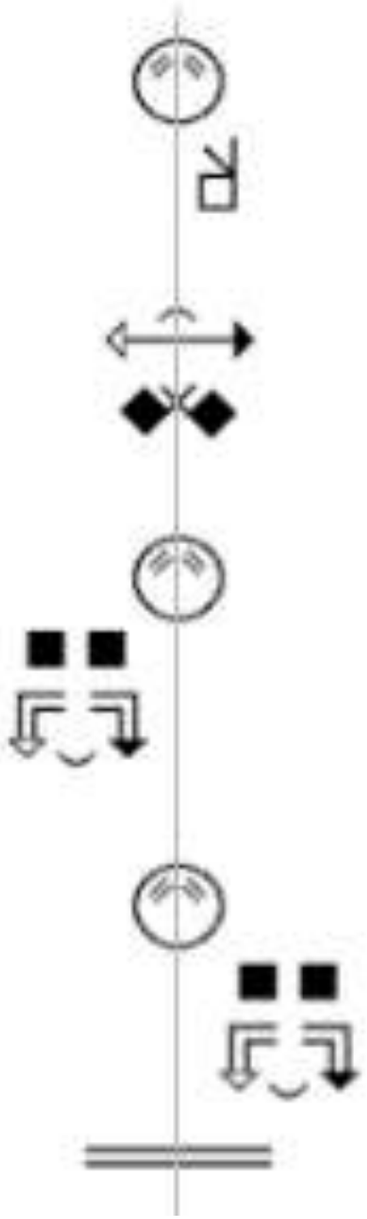
SWU characters
SWU to FSW
SWU codepoits

Technical

Regular Expressions
UTF-8, UTF-16, and UTF-32
Character Ranges and Text

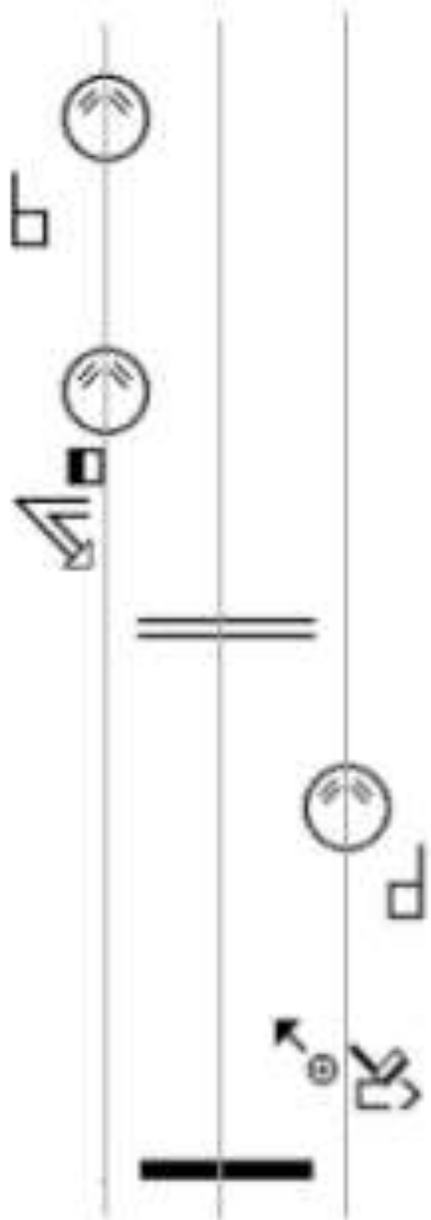
Sutton SignWriting

A script for sign languages.
ISO 15924 Script Code sgnw

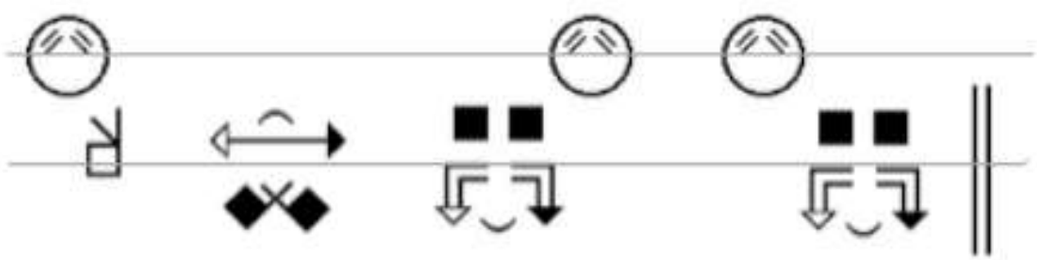


Each sign is written as a word.

The words are 2-dimensional clusters of symbols.



The words are combined with punctuation to form text.



Formal SignWriting

Formal SignWriting is one particular computerized design for Sutton SignWriting that envisions a sign as a two part word.

Each word is written as a string of characters that can be recognized and processed by regular expressions.

The design has been optimized for display, searching, sorting, text flow, and other character processing.

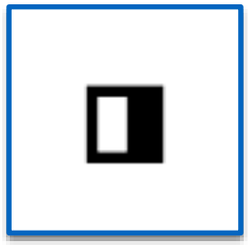
Where as American Sign Language is a natural language, Formal SignWriting is a formal language.

A formal language is useful in mathematics, computer science, and linguistics.

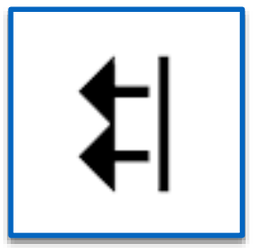
Formal SignWriting

A two part word of time and space.

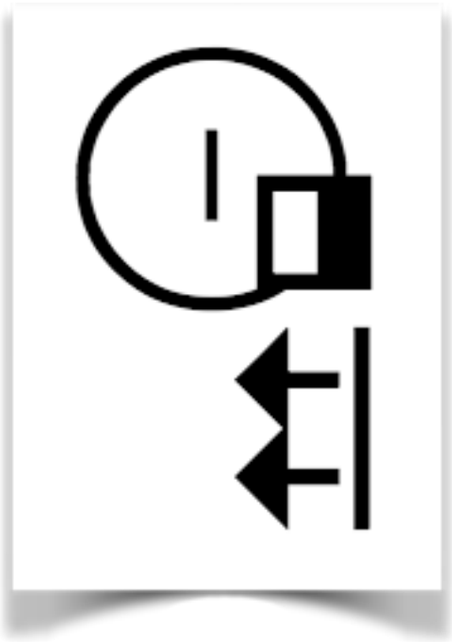
First syllable
Starting hand shapes



Second syllable
Movements and Dynamics



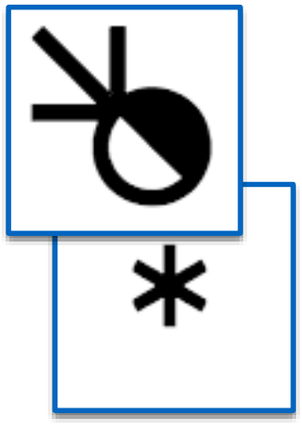
Third syllable
Faces and Locations



Layered writing in
2-Dimensions

Temporal Prefix

Spatial Signbox



Formal SignWriting in ASCII

FSW

AS18711S20500M514x517S18711490x483S20500486x506

Mathematical names

ASCII characters only
ABLMRS0123456789xabcdef

Signs are written as unified words

Stable since January 2012

https://datatracker.ietf.org/doc/draft-slevinski-formal-signwriting/

The Unicode Standard

“a worldwide character standard”

“addresses only the encoding and semantics of text.”

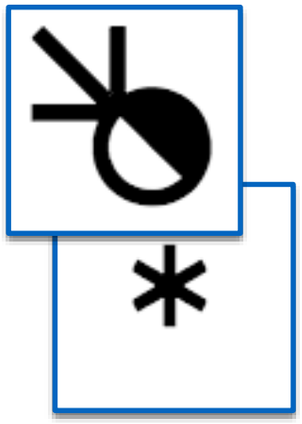
“used for representation of text for computer processing.”

Principles of the Unicode Standard

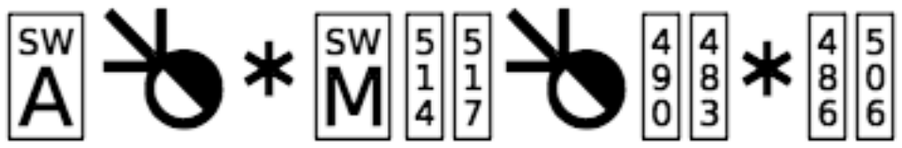
- Universal repertoire
- Logical order
- Efficiency
- Unification
- Characters, not glyphs
- Dynamic composition
- Semantics
- Stability
- Plain Text
- Convertibility

“The Unicode Standard groups characters together by scripts in blocks. A script is any system of related characters.”

http://unicode.org/standard/principles.html



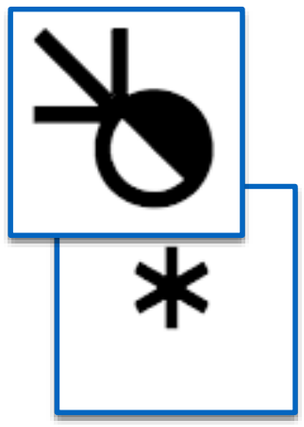
SWU Characters

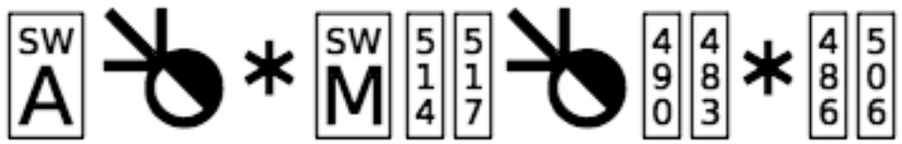


Description	Formal SignWriting	Unicode Characters
Sequence Marker	A	U+1D800
SignBox Markers	B, L, M, R	U+1D801..U+1D804
Numbers	250 to 749	U+1D80C..U+1D9FF
Sutton SignWriting Symbols	S10000 to S38b07	Plane 4

http://www.unicode.org/L2/L2017/17220-signwriting-design-opt.pdf

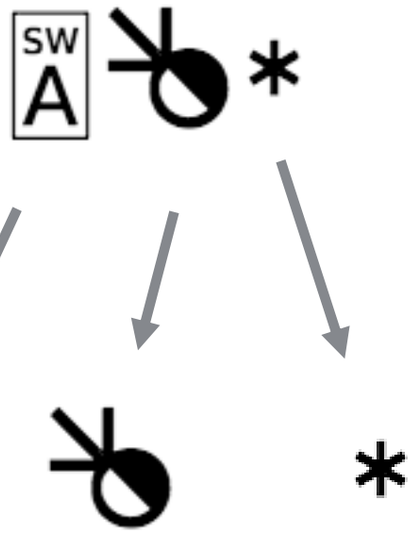
SWU is equivalent to FSW

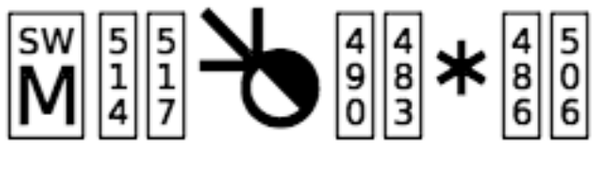


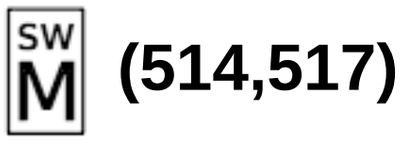


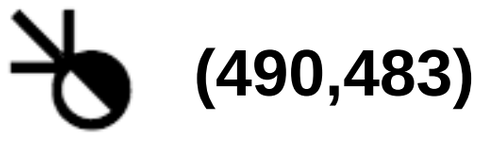
Time

Space









Sequence Marker

Sorting List

Middle Lane Signbox

Max Coord

Spatial Symbol

Spatial Symbol

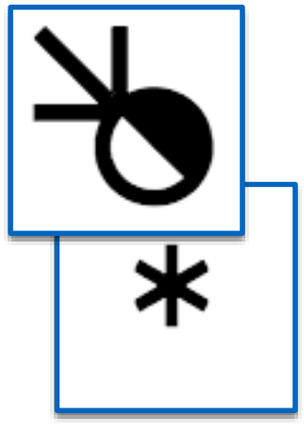
A S18711 S20500

M 514x517

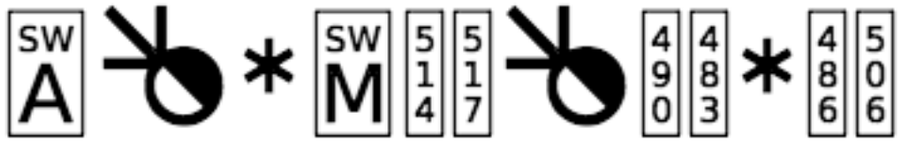
S18711 490x483

S20500 486x506

AS18711S20500M514x517S18711490x483S20500486x506



SWU Codepoints



U+1D800 U+432B2 U+461E1 U+1D803 U+1D914 U+1D917
U+432B2 U+1D8FC U+1D8F5 U+461E1 U+1D8F8 U+1D90C

U+1D800

U+1D803

U+432B2

U+461E1

483 = U+1D8F5

514 = U+1D914

486 = U+1D8F8

506 = U+1D90C

490 = U+1D8FC

517 = U+1D917

Regular Expressions

Basics

*	Match a literal 0 or more times
+	Match a literal 1 or more times
?	Match a literal 0 or 1 times
{#}	Match a literal "#" times
[ ]	Match any single literal from a list
[-]	Match any single literal in a range
( )	Creates a group for matching
(|)	Matches one of several alternatives

Regular Expressions

Examples

*	ABC* matches AB, ABC, ABCC, ...
+	ABC+ matches ABC, ABCC, ABCCC, ...
?	ABC? matches AB or ABC
{#}	AB{2} matches ABB
[ ]	[ABC] matches A, B, or C
[-]	[A-C] matches A, B, or C
( )	A(BC)+ matches ABC, ABCBC, ABCBCBC, ...
(|)	(AB|BC|CD) will match AB, BC, or CD

Unicode Transformation Formats

UTF-32	U+1D800	U+432B2	One 32-bit character
UTF-16	uD836 uDC00	uD8CC uDEB2	Two 16-bit characters
UTF-8	%F0%9D%A0%80	%F1%83%8A%B2	Four 8-bit characters

Symbol Ranges

UTF-32

all symbols	U+40001 - U+4F480
writing	U+40001 - U+4EFA0
hand	U+40001 - U+461E0
movement	U+461E1 - U+4BCA0
dynamic	U+4BCA1 - U+4BFA0
head	U+4BFA1 - U+4E8E0
horizontal center	U+4BFA1 - U+4E8E0
vertical center	U+4BFA1 - U+4EC40
trunk	U+4E8E1 - U+4EC40
limb	U+4EC41 - U+4EFA0
location	U+4EFA1 - U+4F2A0
punctuation	U+4F2A1 - U+4F480

Symbol Ranges

UTF-16 regular expressions

all symbols	((\uD8C0[\uDC01-\uDFFF])|([\uD8C1-\uD8DC][\uDC00-\uDFFF])|(\uD8DD[\uDC00-\uDC80]))
writing	((\uD8C0[\uDC01-\uDFFF])|([\uD8C1-\uD8FA][\uDC00-\uDFFF])|(\uD8FB[\uDC00-\uDFA0]))
hand	((\uD8C0[\uDC01-\uDFFF])|([\uD8C1-\uD8D7][\uDC00-\uDFFF])|(\uD8D8[\uDC00-\uDDE0]))
movement	((\uD8D8[\uDDE1-\uDFFF])|([\uD8D9-\uD8EE][\uDC00-\uDFFF])|(\uD8EF[\uDC00-\uDCA0]))
dynamic	\uD8EF[\uDCA1-\uDFA0]
head	((\uD8EF[\uDFA1-\uDFFF])|([\uD8F0-\uD8F9][\uDC00-\uDFFF])|(\uD8FA[\uDC00-\uDCE0]))
horizontal center	((\uD8EF[\uDFA1-\uDFFF])|([\uD8F0-\uD8F9][\uDC00-\uDFFF])|(\uD8FA[\uDC00-\uDCE0]))
vertical center	((\uD8EF[\uDFA1-\uDFFF])|([\uD8F0-\uD8FA][\uDC00-\uDFFF])|(\uD8FB[\uDC00-\uDC40]))
trunk	((\uD8FA[\uDCE1-\uDFFF])|(\uD8FB[\uDC00-\uDC40]))
limb	\uD8FB[\uDC41-\uDFA0]
location	((\uD8FB[\uDFA1-\uDFFF])|(\uD8FC[\uDC00-\uDEA0]))
punctuation	((\uD8FC[\uDEA1-\uDFFF])|(\uD8FD[\uDC00-\uDC80]))

Matching Signs written with SWU

Signs can start with an optional temporal prefix.

Both writing symbols and location symbols
are allowed in the temporal prefix.

SW A	Writing symbols	Location symbols
U+1D800	U+40001 - U+4EFA0	U+4EFA1 - U+4F2A0

Matching Signs written with SWU

Temporal Prefix

Regular Expression for UTF-32

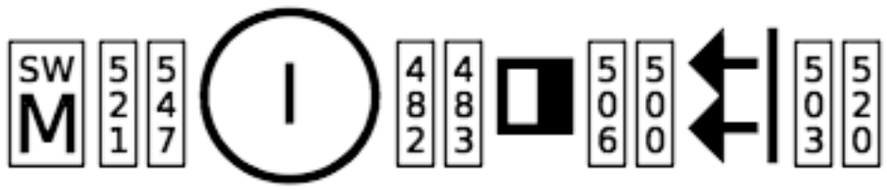
\U0001D800[\U00040001-\U0004F2A0]+

Regular Expression for UTF-16

\uD836\uDC00(((\uD8C0[\uDC01-\uDFFF])|([\uD8C1-\uD8FB][\uDC00-\uDFFF])|(\uD8FC[\uDC00-\uDEA0])))+

Matching Signs written with SWU

All sign have a spatial signbox.

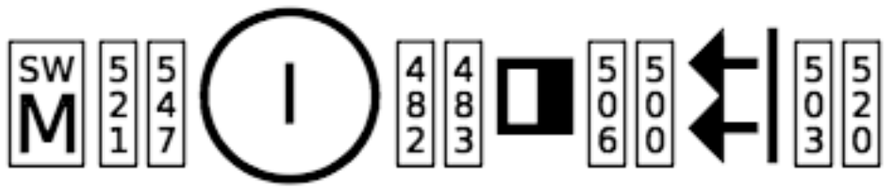


Only writing symbols
are allowed in the spatial signbox.

	Writing symbols	Numbers
U+1D802	U+40001 - U+4EFA0	U+1D80C - U+1D9FF

Matching Signs written with SWU

Spatial Signbox



Regular Expression for UTF-32

[\U0001D801-\U0001D804][\U0001D80C-\U0001D9FF]{2}([\U00040001-\U0004EFA0][\U0001D80C-\U0001D9FF]{2})*

Regular Expression for UTF-16

\uD836[\uDC01-\uDC04](\uD836[\uDC0C-\uDDFF]){2}(((\uD8C0[\uDC01-\uDFFF])|([\uD8C1-\uD8FA][\uDC00-\uDFFF])|(\uD8FB[\uDC00-\uDFA0]))(\uD836[\uDC0C-\uDDFF]){2})*

Matching the Entire Sign

Regular Expression for UTF-32

(\U0001D800[\U00040001-\U0004F2A0]+)? [\U0001D801-\U0001D804][\U0001D80C-\U0001D9FF]{2}([\U00040001-\U0004EFA0][\U0001D80C-\U0001D9FF]{2})*

Regular Expression for UTF-16

(\uD836\uDC00(((\uD8C0[\uDC01-\uDFFF])|([\uD8C1-\uD8FB][\uDC00-\uDFFF])|(\uD8FC[\uDC00-\uDEA0])))+)?\uD836[\uDC01-\uDC04](\uD836[\uDC0C-\uDDFF]){2}(((\uD8C0[\uDC01-\uDFFF])|([\uD8C1-\uD8FA][\uDC00-\uDFFF])|(\uD8FB[\uDC00-\uDFA0]))(\uD836[\uDC0C-\uDDFF]){2})*

Matching Punctuation written with SWU

Punctuation is used between signs

Numbers are required for backwards compatibility with SVG.

Punctuation symbols	Numbers
U+4F2A1 - U+4F480	U+1D80C - U+1D9FF

Matching Punctuation written with SWU

Punctuation

Regular Expression for UTF-32

[\U0004F2A1-\U0004F480][\U0001D80C-\U0001D9FF]{2}

Regular Expression for UTF-16

((\uD8FC[\uDEA1-\uDFFF])|(\uD8FD[\uDC00-\uDC80]))(\uD836[\uDC0C-\uDDFF]){2})*

SignWriting in Unicode dot SWU

by Stephen E Slevinski Jr

slevinski@signwriting.org

Thanks for viewing.

Feedback, comments, and questions are welcomed.

http://signpuddle.com

http://www.slideshare.net/StephenSlevinski/presentations