Hexlish Alphabet for English, Constructed Languages and Cryptography: Automatic, Structural Compression with a Phonetic Hexadecimal Alphabet

Byrl Raze Buckbriar | buck0@octade.net | www.octade.net

DOI: https://doi.org/10.5281/zenodo.13139469

SUMMARY: *Hexlish* is a legible, sixteen-letter alphabet for writing the English language and for encoding text as legible base 16 or compressed binary. Texts composed using the alphabet are automatically compressed by exactly fifty percent when converted from Hexlish characters into binary characters. Although technically lossy, this syntactic compression enables recovery of the correct English letters via syntactic reconstruction. The implementer can predict the size of the compressed binary file and the size of the text that will result from decompression. Generally it is intuitive to recognize English alphabet analogues to Hexlish words. This makes Hexlish a legible alternative to the standard hexadecimal alphabet.

The Hexlish alphabet consists of sixteen characters: (A P C T E V H I J L M N O R S U).

Each character represents multiple English phonemes or letters shown by figure #1.

FIGURE #1
CANONICAL ORDERING

Hexlish	A	P	C	T	E	V	H	I	J	L	M	N	O	R	S	U
English	A	B P	C G K Q	D T	E	F V	H X	I Y	J C H	L	M	N	O W	R	S X Z	U W
Hexadecimal	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F

Transposition of letters or phonemes may sometimes be achieved suitably with more than one Hexlish equivalent symbol. For example the English phoneme or letter *W* may be represented by the Hexlish letters, *O* or *U*. The English letters, *CH* may be represented by the Hexlish pair, *CH*, or the single Hexlish letter, *J*. The translator may exercise discretion of which characters to choose and whether to transpose phonetically or literally, or alternately.

The intuitive transposition of characters will produce intelligible reverse transcription regardless of whether transcribing is literal or phonetic. The reduction from 26 to 16 letters enables automatic 50% compression of text simply by converting the base 16 text to binary. Spaces may be omitted for slightly more compression. Spaces may be marked by a double character such as 'HH' or 'JJ'. Or spaces may be marked with any single character that does not produce a possible new word when inserted between words.

For example the phrase, 'CAT DOC OWL' is marked with insertions to become, 'CATJDOCJJOWL' or, 'CATJDOCIOWL' using both space marking rules. Either result may then be converted to standard hexadecimal [1] for decoding to binary output.

The table in figure #2 shows some arbitrary English to Hexlish transpositions with Hexlish words in red beneath each English equivalent.

FIGURE # 2 - Sample English to Hexlish Literal and Phonetic Transpositions

ALPHABET	LETTERS	COMPUTER	HASHTAG	INTERNET	LANGUAGE
ALVAPET	LETERS	COMPUTER	HASHTAC	INTERNET	LANCUACE
ENGLISH	HEXLISH	TRANSLATE	BINARY	ENCRYPTION	DECIMALS
ENCLISH	HEHLISH	TRANSLAT	PINARI	ENCRIPTION	TESIMALS
COMPRESS	ENCODING	CONLANG	PREPRINT	AMERICA	BUBBLEGUM
COMPRES	ENCOTINC	CONLANC	PREPRINT	AMERICA	PUPLECUM

Note that the plain English words are ASCII [2] binary characters. The Hexlish words can be converted to standard hex characters and decoded to binary characters for 50% compression. The values of each Hexlish character couplet may be directly converted to a binary value using the same mathematical logic as in hexadecimal.

Figure #3 demonstrates analogue characters between Hexlish and hexadecimal with the Hexlish letters sequenced in Roman alphabet order rather than canonical order.

FIGURE #3 - Hexlish to Hexadecimal Mapping using Roman Alphabet Ordering

Hexlish	A	C	E	H	I	J	L	M	N	O	P	R	S	T	U	V
Hexadecimal	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F

Figure #4 shows a transcription from English (blue) to Hexlish (red) to hexadecimal.

FIGURE #4 - English to Hexlish and Hexadecimal Mapping using Roman Alphabet Ordering

HEXLISH	ALPHABET	ENCODING	LANGUAGE	CIPHERS	BUBBLEGUM
HEHLISH	ALVAPET	ENCOTINC	LANCUACE	SIVERS	PUPLECUM
32364C3	06F0A2D	2819D481	6081E012	C4F2BC	AEA621E7

We may compress the hexadecimal characters by decoding to binary which occupies only one byte for each pair of ASCII characters. Thus compression is native to the alphabet without need of complex compression algorithms. Moreover, compression algorithms may be applied to the compressed binary for greater than 50% compression.

Encoding of integers as Hexlish works the same as standard hexadecimal while using the Hexlish mapping. Just as hexadecimal uses the prefixes, '0X' and '$' when represented in source code the Hexlish may be indicated with integer prefixes such as, calculator natation '0h' and the improvised prefix, 'H$'. The table in figure #5 demonstrates prefixes showing Hexlish words in red and hexadecimal words in green.

FIGURE #5 - Hexlish and Hexadecimal Prefix Examples

0hHEHLISH	H$ALVAPET	0hENCOTINC	H$LANCUACE	0hSIVERS	H$PUPLECUM
0x32364C3	0x06F0A2D	0x2819D481	0x6081E012	0xC4F2BC	0xAEA621E7

For ciphers and cryptography multiple Hexlish alphabets may be used by selecting unique sets of the English alphabet letters for each Hexlish set. Up to 26 unique sets may be chosen by assigning one English letter to each set that does not appear in the other sets. Each set may have a random or pseudo random ordering for cryptographic purposes. Letters may be used to encode digits with a special marker to indicate the start of an integer in the sequence.

A Hexlish keyboard could consist of three rows of eight characters per row. Two rows would be the Hexlish alphabet and one row could contain various punctuation and control characters. Typing English text in Hexlish would provide phonetic compression in the English text and binary compression when the text is converted and stored as raw bytes.

This next table in figure #6 shows a arbitrary layout for a simple Hexlish keyboard map.

FIGURE #6
Hexlish keys

!	-	#	.	,	?	(	)
A	P	C	T	E	V	X	I
J	L	M	N	O	R	S	U

This example layout would be of limited use and is provided merely for illustration. A Hexlish keyboard could have four or five rows for extra punctuation, numerical, and control characters commonly used on electronic computers. One advantage of a keyboard with many less keys than a QWERTY keyboard is the ability to make all the keys larger than usual while still maintaining a smaller overall form factor. The example keyboard would be nearly sufficient for mere textual transcription and not for programming.

Hexlish has more potential utility which I leave to the imagination of the reader.

NOTES & ERRATA

The word, *Hexlish* is macaronic like the word, *hexadecimal*.

The word, Hexlish, sounds similar to the german word, *hässlich*, meaning, *ugly*. This pun was not intended in the naming of Hexlish yet fate yields it.

Hexlish should not be confused with *Hexspeak* [3] as the twain are quite different in structure and purpose. This is not that.

The author has unsuccessfully searched existing literature to see if there are any other compressive alphabets similar to Hexlish. If the reader knows of any such alphabets please do inform the author for expanding this research.

REFERENCES

[1] Wikipedia contributors. (2024, June 25). Hexadecimal. In *Wikipedia, The Free Encyclopedia*. < https://en.wikipedia.org/wiki/Hexadecimal >

[2] Wikipedia contributors. (2024, July 12). ASCII. In *Wikipedia, The Free Encyclopedia*. < https://en.wikipedia.org/wiki/ASCII >

[3] Wikipedia contributors. (2024, July 12). Hexspeak. In *Wikipedia, The Free Encyclopedia*. < https://en.wikipedia.org/wiki/Hexspeak >

GREENPRINT DATA

```
== greenprint = # 3HDM.OFCT.656H.VK4J.3YIB.FSSI.F5B3.AR53
== title = Hexlish Alphabet for English, Constructed Languages and Cryptography:
Automatic, Structural Compression with a Phonetic Hexadecimal Alphabet
== author = Byrl Raze Buckbriar
== version = 1
== date == Friday 26 July 2024
```