Digital AV – Files & Record Layouts

Authorized Version (aka King James Bible) – SDK 2011 Edition – Rev H815

AVBible.2011 (4-byte records; 32 bits)

Record #	WordKey(2) see AVEnglish.2011	Punctuation(1)	VerseNum(1)
0	0x0000 (delimiter)	size^{gen} & 0xFF	size^{gen} / 0x100
1	WordKey of “In”	0x0	1
2	WordKey of “the”	0x0	1
3	WordKey of “beginning”	0x0	1
… (until end of chapter, then start again with chapter 2 of Genesis)			
38263	WordKey of “Thus”	0x0	1
38263	WordKey of “the”	0x0	1
38264	WordKey of “heavens”	0x0	1
… (until end of book; Genesis depicted)			
size^{gen} -1	WordKey of “coffin”	0x0	26
size^{gen}	WordKey of “in”	0x0	26
size^{gen} +1	WordKey of “Egypt”	0xE0	26
size^{gen} +2	0x0000 (delimiter)	size^{exo} & 0xFF	size^{exo} / 0x100
… (until last word of last chapter of the 66th book: namely Revelation)			
789716	WordKey of “Amen”	0xE0	21
789717	0x0000 (delimiter)	0x00	0x00

Bit patterns for Punctuation Byte

Description	Bit Pattern (Hex)
PUNCclause	0xE0
PUNCexclamatory	0x80
PUNCinterogative	0xC0
PUNCdeclarative	0xE0
PUNCdash	0xA0
PUNCsemicolon	0x20
PUNCcomma	0x40
PUNCcolon	0x60
PUNCpossessive	0x10
PUNCparagraph	0x08
MODEparenthetical	0x04
MODEitalics	0x02
MODEjesus	0x01

Bit patterns for Top 2 bits of WordKey

Description	Bit Pattern (Hex)
English Word	0x3FFF (This mask provides index into AVEnglish.2011)
Illegal Word	0x0000 (Delimiter that represents end-of-lists)
1st Letter Cap	0x8000 (example: Lord)
All Letters CAPS	0x4000 (example: LORD)
Use Mixed CaPs	0xC000 (Use captitolization from AVEnglish.2011)

AVEnglish.2011 (variable-length records)

Offset	WordSize (2 bytes)	Word Count (2 bytes)
0: Max Len	Two-Byte record = 18 (Word Count field not in 1st record)	
2: Header	1	3
6 6 = 2 + 4	CLOB[1] = {'a'}, key=1 {'i'}, key=2 {'o'}, key=3	
9: Header 9 = 6 + (3 x 1)	2	40
12 12 = 9 + 3	CLOB = {'a','h'}, key = 4 {'a','i'}, key = 5 {'a','m'}, key = 6 {'a','n'}, key = 7 {'a','r'}, key = 8 {'a','s'}, key = 9 {'a','t'}, key = 10 {'b','e'}, key = 11 {'b','y'}, key = 12 {'d','o'}, key = 13 {'e','d'}, key = 14 {'e','r'}, key = 15 {'g','o'}, key = 16 {'h','a'}, key = 17 {'h','e'}, key = 18 {'h','o'}, key = 19 {'i','f'}, key = 20 {'i','n'}, key = 21 {'i','r'}, key = 22 {'i','s'}, key = 23 {'i','t'}, key = 24 {'l','o'}, key = 25 {'m','e'}, key = 26 {'m','y'}, key = 27 {'n','o'}, key = 28 {'o','f'}, key = 29 {'o','g'}, key = 30 {'o','h'}, key = 31 {'o','n'}, key = 32 {'o','r'}, key = 33 {'o','x'}, key = 34 {'p','e'}, key = 35 {'s','o'}, key = 36 {'t','o'}, key = 37 {'u','p'}, key = 38 {'u','r'}, key = 39 {'u','s'}, key = 40 {'u','z'}, key = 41 {'w','e'}, key = 42 {'y','e'}, key = 43	
92: Header 92 = 12+(40x2)	3	Word count of 3-letter words

96
96 = 92 + 4
	CLOB = {'a','b','i'}, key = 44 {'a','c','t'},
	key = 45 …	

[1] CLOB is an acronym for Character-Large-Object (these character arrays are NOT null-terminated).

Digital AV – Files & Record Layouts

Authorized Version (aka King James Bible) – SDK 2011 Edition – Rev H815

AVInfo.2011 (8-byte records × 4 words; 32 bits)

Record #	Strongs #1	Strongs #2	Strongs #3	Strongs #4
0 (record-cnt)	size^{gen}	0	0	0
1 (avnum = 1)	1st Strongs #	2nd Strongs #	3rd Strongs #	0
2 (avnum = 2)	1st Strongs #	2nd Strongs #	3rd Strongs #	0
3 (avnum = 3)	1st Strongs #	2nd Strongs #	3rd Strongs #	0

… (until end of book; Genesis depicted)

. . .

Matthew	size$_{gmatt}$	0	0	0
(record-cnt)	1st Strongs # xor 0x8000	2nd Strongs # xor 0x8000	3rd Strongs # xor 0x8000	4th Strongs # xor 0x8000

… (until end of book; Matthew depicted)

Revelation	Size^{rev}	0	0	0
(record-cnt)	1st Strongs # xor 0x8000	2nd Strongs # xor 0x8000	3rd Strongs # xor 0x8000	4th Strongs # xor 0x8000

… (until last word of last chapter of the 66th book: namely Revelation)

(record-cnt)	0	0	0	0

All revisions of the Digital-AV encode date-of-last-revision in the revision number. When the AV-SDK 2011 became re-branded as the Digital-AV, it also embraced the revision encoding scheme of all other Digital-AV editions. The revision number is encoded using a four-digit character sequence. These four digits can be interpreted as YMDD. Year (Y) is encoded as a single base-36 digit. Year Zero (Y=0) represents the year of our LORD 2000. Year Seventeen (Y = H) represents 2017. The M digit for month is similar, but easier. Digits 1 through 9 are as expected. A is October, B is November, and C is December. The two DD digits range from 01 to 31.

As the 2017 edition of the Digital-AV has been released, revision #H815 is expected to be the final revision of the 2011 Edition of Digital-AV [aka, this SDK]. If you encounter documents identified as the 2011 Edition of the AV-SDK, then these likely predate the revision scheme introduced by the rebranding of Digital-AV. Such documents are identified as they do not mention Digital-AV in the header of the SDK document. If you are using one of these earlier SDK's, you may want to update this this release. Check the date of the files underlying the SDK: if they are earlier than 2017/08/15, then consider updating.

It is interesting to note that the 2017 Edition was initially built using revision #H815 of the 2011 Edition. As the author of Digital-AV utilized the 2011 Edition of the SDK for at least a decade prior to open sourcing the files in 2011, he learned a lot about what would make the SDK easier to use. For that reason, he provided a new edition of the SDK in 2017. The 2017 Edition is recommended for all new development. This document is maintained merely as a service to any developer who has already standardized on the 2011 Edition. Both editions sport the same permissive MIT-style open-source license. There were no known bugs in the 2011 SDK, but the developer had discovered far too many use-cases where the usage of the 2011 Edition was tedious. For example, in the 2011 lexicon, lexical entries are not null terminated: this saved about 16K on disk space. In 1996, this design made some sense; but in 2017, not so much. Similar processing optimizations were made throughout the SDK. But to be clear, the 2017 Edition is far beefier. So perhaps a mobile app developer should weigh the pros and cons of this SDK vis-à-vis the 2017 Edition. The choice is yours; the license is the same. But the author is moving full steam ahead with the 2017 Edition.