AV Text Ministries

Digital-AV

Part-of-Speech (Digital-AV SDK Specification)

**Revision:** #i728

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Number | POS bits | Tag | Description | | additional bits | |
| 1. | 0x\_C00 | CC | Coordinating conjunction | | |  |
| 2. | 0x\_F00 | CD | Cardinal number | | |  |
| 3. | 0x\_D00 | DT | Determiner | | |  |
| 4. | 0x0E00 | EX | Existential *there* | | |  |
| 5. | 0x0000 | FW | Foreign word | | |  |
| 6. | 0x0400 | IN | Preposition or subordinating conjunction | | |  |
| 7. | 0x\_A00 | JJ | Adjective | | | Unmarked: 0x0A00 |
| 8. | 0x1A00 | JJR | Adjective, comparative | | |  |
| 9. | 0x2A00 | JJS | Adjective, superlative | | |  |
| 10. | 0x0700 | LS | List item marker | | |  |
| 11. | 0x\_301 | MD | Modal | | |  |
| 12. | 0x401\_ | NN | Noun, singular or mass | | |  |
| 13. | 0x801\_ | NNS | Noun, plural | | |  |
| 14. | 0x501\_ | NNP | Proper noun, singular | | |  |
| 15. | 0x901\_ | NNPS | Proper noun, plural | | |  |
| 16. | 0x2D00 | PDT | Predeterminer | | |  |
| 17. | 0x0008 | POS | Possessive ending | | |  |
| 18. | 0x\_03\_ | PRP | Personal pronoun | Nominative: 0x\_07\_  Oblique: 0x\_0B\_  Reflexive: 0x\_0F\_  Unmarked: 0x\_03\_ | | Neuter: 0x\_\_\_1  Masculine: 0x\_\_\_2  non-feminine: 0x\_\_\_3  Feminine: 0x\_\_\_4 |
| 19. | 0x\_038 | PRP$ | Possessive pronoun |  | | Neuter: 0x\_\_\_1  Masculine: 0x\_\_\_2  non-feminine: 0x\_\_\_3  Feminine: 0x\_\_\_4  Genitive: 0x\_\_\_8 |
| 20. | 0x\_B00 | RB | Adverb | | | Unmarked: 0x0B00 |
| 21. | 0x1B00 | RBR | Adverb, comparative | | |  |
| 22. | 0x2B00 | RBS | Adverb, superlative | | |  |
| 23. | 0x1D00 | RP | Particle | | |  |
| 24. | 0x\_300 | SYM | Symbol | | |  |
| 25. | 0x\_200 | TO | *to* | | |  |
| 26. | 0x\_800 | UH | Interjection | | |  |
| 27. | 0x\_101 | VB | Verb, base form | | |  |
| 28. | 0x\_102 | VBD | Verb, past tense | | |  |
| 29. | 0x\_104 | VBG | Verb, gerund or present participle | | |  |
| 30. | 0x\_106 | VBN | Verb, past participle | | |  |
| 31. | 0x\_103 | VBP | Verb, non-3rd person singular present | | |  |
|  | 0x\_108 |  | Verb, marked for singular case agreement | | | (EModE conjugation) |
| 32. | 0x\_107 | VBZ | Verb, 3rd person singular present | | |  |
| 33. | 0xCD00 | WDT | Wh-determiner | | |  |
| 34. | 0xC030 | WP | Wh-pronoun | | |  |
| 35. | 0xC038 | WP$ | Possessive wh-pronoun | | |  |
| 36. | 0xCB00 | WRB | Wh-adverb | | |  |

The Digitial-AV utilizes the part-of-speech (POS) tags as defined by the Penn Treebank. POS tagging of the bible verses themselves is performed during SDK compilation. In order to minimize the number of unknown words relative to the Penn Treebank, somewhat archaic words are modernized prior to submission to the TextBlob tagger. TextBlob itself uses the default NLTK tagger under the hood. There is not only preprocessing of text submitted for tagging, but also post-processing to handle 2nd-Person singular and associated verb conjugations. Moreover, Hitchcock’s Bible Name Dictionary sets words to Proper-Nouns when they are found in that dictionary (potentially overriding the tag provided by TextBlob/NLTK).

TextBlob is housed in a Python Django web server, whereas the SDK compiler is written in C#/dot-net. The SDK compiler gets tokenized sentences from the TextBlob/NLTK library via Django as a JSON response. Source code to the entire pipeline is available on github at:

<https://github.com/kwonus/avtext>

and

<https://github.com/kwonus/Digital-AV>

The table of the previous page was cloned from the Penn Treebank website identified in the header of this document. After cloning the table, the second column was added to show how the standard Penn Treebank tags map into the bit fields of the Digital-AV.

It should be noted that while the table depicts an accurate mapping of Digital-AV bits from/to standard POS tags, some bits get set by the SDK compiler, most notably: additional bits are set for Person-Number (PN), as defined in the primary Digital-AV documentation.

POS Summary (not entire range of values):

Verb: 0x\_10\_

Modal: 0x\_30\_

Noun: 0x\_01\_

Proper noun:0x1\_1\_

Pronoun: 0x\_03\_

WH: 0xC\_\_\_

Possessive: 0x\_\_\_8

ADJ: 0x\_A00

ADV: 0x\_B00

DET: 0x0D\_0 // unmarked for number (e.g. the)

DET-SING: 0x4D\_0

DET-PLURAL: 0x8D\_0

Particle: 0x1D00

Prep: 0x0400

To: 0x0200

Interject. 0x0800

PN nibble (applies to verbs and nouns only):

1st Person: 0x1\_\_\_

2nd Person: 0x2\_\_\_

3rd Person: 0x3\_\_\_

Unmarked: 0x0\_\_\_

Singular: 0x4\_\_\_

Plural: 0x8\_\_\_