McKeeman Form


crockford.com/mckeeman.html

McKeeman Form is a notation for expressing grammars. It was proposed by Bill McKeeman of Dartmouth College. It is a simplified Backus-Naur Form with significant whitespace and minimal use of metacharacters.

Grammar

We can express the grammar of McKeeman Form in McKeeman Form.

A *grammar* is a list of one or more rules.

grammar rules

The Unicode code point `U+0020` is used as the *space*. The Unicode code point `U+000A` is used as the *newline*.

space `'0020'`

newline `'000A'`

A *name* is a sequence of letters or `_` underbar.

name letter letter name

letter `'a'` . `'z'` `'A'` . `'Z'` `'_'`

An *indentation* is four spaces.

indentation space space space space

Each of the *rules* is separated by a *newline*. A *rule* has a *name* on one line, with *alternatives* indented below it.

rules rule rule newline rules

rule name newline nothing alternatives

If the first line after the *name* of a *rule* is `""`, then the *rule* may match *nothing*.

nothing `""` indentation `'"'` `'"'` newline

Each *alternative* is indented on its own line. Each *alternative* contains *items* followed by a *newline*.

alternatives alternative alternative alternatives

alternative indentation items newline

The *items* are separated by spaces. An *item* is a *literal* or the *name* of a *rule*.

items item item space items

item literal name

literal singleton range exclude `'"'` characters `'"'`

Any single Unicode code point except the 32 control codes may be placed within the single quotes. The *hexcode* of any Unicode code point may also be placed within the single quotes. A *hexcode* can contain 4, 5, or 6 hexadecimal digits.

singleton `'''` codepoint `'''`

codepoint `' '` `.` `'10FFFF'` hexcode

hexcode `"10"` hex hex hex hex hex hex hex hex hex hex hex hex hex

hex `'0'` `.` `'9'` `'A'` `.` `'F'`

A *range* is specified as a *singleton*, a `.` period, and another *singleton*. Literal ranges can optionally be followed by `-` minus sign and characters to be excluded.

range singleton space `'.'` space singleton

exclude `""` space `'-'` space singleton exclude space `'-'` space range exclude

A *character* wrapped in `"` double quote can be any of the Unicode code points except the 32 control codes and `"` double quote. The definition of *character* shows an example of a codepoint range and exclude.

characters character character characters

character `' '` `.` `'10FFFF'` `-` `'"'`

JSON

This is the JSON grammar in McKeeman Form.

json element

value object array string number `"true"` `"false"` `"null"`

object `'{'` ws `'}'` `'{'` members `'}'`

members member member `','` members

member ws string ws `':'` element

array `'['` ws `']'` `'['` elements `']'`

elements element element `','` elements

element ws value ws

string `'"'` characters `'"'`

characters `""` character characters

character `'0020'` `.` `'10FFFF'` `-` `'"'` `-` `'\'` `'\'` escape

escape `'"'` `'\'` `'/'` `'b'` `'f'` `'n'` `'r'` `'t'` `'u'` hex hex hex hex

hex digit `'A'` `.` `'F'` `'a'` `.` `'f'`

number integer fraction exponent

integer digit onenine digits `'-'` digit `'-'` onenine digits

digits digit digit digits

digit `'0'` onenine

onenine `'1'` `.` `'9'`

fraction `""` `'.'` digits

exponent `""` `'E'` sign digits `'e'` sign digits

sign `""` `'+'` `'-'`

ws `""` `'0020'` ws `'000A'` ws `'000D'` ws `'0009'` ws