

A grammar to parse a file into Pod and Code sections.

INTRODUCTION

I want to create a semi-literate Raku source file with the extension `.sl`. Then, I will *weave* it to generate a readable file in formats like Markdown, PDF, HTML, and more. Additionally, I will *tangle* it to create source code without any Pod6.

To do this, I need to divide the file into Pod and Code sections by parsing it. For this purpose, I will create a dedicated Grammar.

Convenient tokens

Let's create two tokens for convenience.

```
1|    my token rest-of-line {    \N* [\n | $] }
2|    my token blank-line   { ^^ \h* [\n | $] }
```

The Grammar

Our file will exclusively consist of Pod or Code sections, and nothing else. The TOP token clearly indicates this.

```
3| grammar Semi::Literate is export {  
4|     token TOP { [ <pod> | <code> ]* }
```

The Pod6 delimiters

According to the [documentation](#),

Every Pod6 document has to begin with `=begin pod` and end with `=end pod`.

So let's define those tokens.

The `begin` token

```
5|     my token begin {  
6|         ^^ \h* \= begin <.ws> pod
```

Most programming applications do not focus on the structure of the executable file, which is not meant to be easily read by humans.

However, we can provide the option for users to specify the number of empty lines that should replace a pod block. To do this, simply add a number at the end of the `=begin` directive. For example, `=begin pod 2`.

```
7|         [ \h* $<num-blank-lines>=(\d+) ]?
```

The remainder of the `begin` directive can only be whitespace.

```
8|         <rest-of-line>  
9|     }
```

The `end` token

The `end` token is much simpler.

```
10|     my token end { ^^ \h* \= end <.ws> pod <rest-of-line> }
```

The Pod token

Within the delimiters, all lines are considered documentation. We will refer to these lines as `plain-lines`. Additionally, it is possible to have nested Pod sections. This allows for a hierarchical organization of documentation, allowing for more structured and detailed explanations.

It is also permissible for the block to be empty. Therefore, we will use the 'zero-or-more' quantifier on the lines of documentation, allowing for the possibility of having no lines in the block.

```
11|     token pod {  
12|         <begin>  
13|         [<pod> | <plain-line>]*  
14|         <end>  
15|     }
```

The Code token

The Code sections are trivially defined. They are just one or more plain-lines.

```
16|      token code { <plain-line>+ }
```

The plain-line token

The plain-line token is, really, any line at all...

```
17|      token plain-line {  
18|          $<plain-line> = [^^ <rest-of-line>]
```

Disallowing the delimiters in a plain-line.

... except for one subtlety. They it can't be one of the begin/end delimiters. We can specify that with a Regex Boolean Condition Check.

```
19|          <?{ &not-a-delimiter($<plain-line>.Str) }>  
20|      }
```

This function simply checks whether the plain-line match object matches either the begin or end token.

Incidentally, this function is why we had to declare those tokens with the my keyword. This function wouldn't work otherwise.

```
21|      sub not-a-delimiter (Str $line --> Bool) {  
22|          return not $line ~~ /<begin> | <end>/;  
23|      }
```

And that concludes the grammar for separating Pod from Code!

```
24| }
```

The Tangle subroutine

This subroutine will remove all the Pod6 code from a semi-literate file (.sl) and keep only the Raku code.

```
25| sub tangle (
```

The subroutine has a single parameter, which is the input filename. The filename is required. Typically, this parameter is obtained from the command line through the wrapper subroutine MAIN.

```
26|     Str $input-file!,
```

The subroutine will return a Str, which should be a working Raku program.

```
27|     --> Str ) is export {
```

First we will get the entire .sl file...

```
28|     my Str $source = $input-file.IO.slurp;
```

Remove the

```
29|     $source ~~ s:g{ ^^ \h* '#' <.ws>      'no-weave' <rest-of-line> } = '';
30|     $source ~~ s:g{ ^^ \h* '#' <.ws> 'end-no-weave' <rest-of-line> } = '';
```

Remove blank lines at the beginning and end of the code sections.

```
31|     $source ~~ s:g/\=end (\N*)\n+/\=end$0\n/;
32|     $source ~~ s:g/\n+\=begin      /\n\=begin/;
```

...Next, we parse it using the Semi::Literate grammar and obtain a list of submatches (that's what the caps method does) ...

```
33|     my Pair @submatches = Semi::Literate.parse($source).caps;
```

...And now begins the interesting part. We iterate through the submatches and keep only the code sections...

```
34|     my Str $raku-code = @submatches.map( {
35|         when .key eq 'code' {
36|             .value;
37|         }
```

Most programming applications do not focus on the structure of the executable file, which is not meant to be easily read by humans.

However, we can provide the option for users to specify the number of empty lines that should replace a pod block. To do this, simply add a number at the end of the =begin directive. For example, =begin pod 2 .

```
38|         when .key eq 'pod' {
39|             my $num-blank-lines = .value.hash<begin><num-blank-lines>;
40|             with $num-blank-lines { "\n" x $num-blank-lines }
41|         }
```

... and we will join all the code sections together...

```
42|     }
43|     ).join;
```

remove blank lines at the end

```
44|      $raku-code ~~ s{\n    <blank-line>* $ } = '';
```

And that's the end of the tangle subroutine!

```
45|      return $raku-code;
46|  }
```

The Weave subroutine

The Weave subroutine will *weave* the `.sl` file into a readable Markdown, HTML, or other format. It is a little more complicated than `sub tangle` because it has to include the code sections.

```
47| sub weave (
```

The parameters of Weave

`sub weave` will have several parameters.

`$input-file`

The input filename is required. Typically, this parameter is obtained from the command line through a wrapper subroutine `MAIN`.

```
48|     Str $input-file!;
```

`$format`

The output of the weave can (currently) be Markdown, Text, or HTML. It defaults to Markdown. The variable is case-insensitive, so 'markdown' also works.

```
49|     Str :f(:$format) is copy = 'markdown';
```

`$line-numbers`

It can be useful to print line numbers in the code listing. It currently defaults to `True`.

```
50|     Bool :l(:$line-numbers) = True;
```

`sub weave` returns a `Str`.

```
51|     --> Str ) is export {
```

```
52|         my UInt $line-number = 1;
```

First we will get the entire `.sl` file...

```
53|         my Str $source = $input-file.IO.slurp;
```

```
54|
```

```
55|         my Str $cleaned-source;
```

Clean the source of items we don't want to see in the woven document.

remove code marked as 'no-weave'

```
56|         $source ~~ s:g{^^ \h* '#' <.ws> 'no-weave'      <rest-of-line>
```

```
57|
```

```
58|                 (^^ <rest-of-line> )*?
```

```
59|
```

```
60|                 ^^ \h* '#' <.ws> 'end-no-weave' <rest-of-line>
```

```
61|                 } = '';
```

remove full comment lines followed by blank lines

```
62|     $source ~~ s:g{ ^^ \h* '#' \N* \n+} = '';
```

head3 Remove EOL comments

```
63|     for $source.split("\n") -> $line {
64|         my $m = $line ~~ m{
65|             ^^
66|             $<stuff-before-the-comment> = ( \N*? )
67|
68|             <!after
69|                 ( [
70|                     | \\\
71|                     | \" <- [\">*
72|                     | \' <- [\'>*
73|                     | \[] <- [\[]>*
74|                 ] )
75|             >
76|             "#" \N*
77|             $$ };
78|
79|         $cleaned-source ~= $m ?? $<stuff-before-the-comment> !! $line;
80|         $cleaned-source ~= "\n";
81|     }
```

Remove blank lines at the begining and end of the code

```
82|     $cleaned-source ~~ s:g{\=end (\N*)\n+} = "\=end$0\n";
83|     $cleaned-source ~~ s:g{\n+\=begin (<.ws> pod) [<.ws> \d]?} = "\n\=begin$0";
```

...Next, we parse it using the `Semi::Literate` grammar and obtain a list of submatches (that's what the `caps` method does) ...

```
84|     my Pair @submatches = Semi::Literate.parse($cleaned-source).caps;
```

...And now begins the interesting part. We iterate through the submatches and insert the code sections into the `Pod6`...

```
85|     my Str $weave = @submatches.map( {
86|         when .key eq 'pod' {
87|             .value
88|         }
89|         when .key eq 'code' { qq:to/EOCB/; }
90|             \=begin pod
91|             \=begin code :lang<raku>
92|             { my $fmt = ($line-numbers ?? "%3s| " !! ' ') ~ "%s\n";
93|               .value
94|               .lines
95|               .map($line-numbers
96|                   ?? {"%4s| %s\n".sprintf($line-number++, $_) }
97|                   !! { "%s\n".sprintf(                                     $_) }
98|               )
99|               .chomp;
100|            }
101|             \=end code
102|             \=end pod
103|             EOCB
```



```
104|  
105|    }  
106|    ).join;
```

remove useless Pod directives

```
107|    $weave ~~ s:g{ \h* \=end    <.ws> pod  <rest-of-line>  
108|                \h* \=begin <.ws> pod <rest-of-line> } = '';
```

remove blank lines at the end

```
109|    $weave ~~ s{\n  <blank-line>* $ } = '';
```

And that's the end of the tangle subroutine!

```
110|    return $weave  
111| }
```

NAME

Semi::Literate - Get the Pod vs Code structure from a Raku/Pod6 file.

VERSION

This documentation refers to Semi-Literate version 0.0.1

SYNOPSIS

```
use Semi::Literate;
```

DESCRIPTION

A full description of the module and its features. May include numerous subsections (i.e. =head2, =head2, etc.)

BUGS AND LIMITATIONS

There are no known bugs in this module. Patches are welcome.

AUTHOR

Shimon Bollinger (deoac.bollinger@gmail.com)

LICENSE AND COPYRIGHT

© 2023 Shimon Bollinger. All rights reserved.

This module is free software; you can redistribute it and/or modify it under the same terms as Raku itself. See [The Artistic License 2.0](#).

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.