

```
1| #! /usr/bin/env raku
2|
3| # Get the Pod vs. Code structure of a Raku/Pod6 file.
4| # © 2023 Shimon Bollinger. All rights reserved.
5| # Last modified: Sun 10 Sep 2023 01:27:36 PM EDT
6| # Version 0.0.1
7|
```

**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.

## Convenient tokens

Let's create four tokens for convenience.

```
8|      my token hws      {      <!ww>\h*      } # Horizontal White Space
9|      my token leading-ws { ^^ <hws>      } # Whitespace at start of line
10|     my token optional-chars {      \N*?      }
11|     my token rest-of-line {      \N*      [\n | $] }
12|     my token ws-till-EOL {      <hws>      [\n | $] }
13|     my token blank-line {      ^^ <ws-till-EOL> }
```

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.

# The Grammar

```
14| #use Grammar::Tracer;
15| grammar Semi::Literate is export {
```

Our file will exclusively consist of Pod or Code sections, and nothing else. The Code sections are of two types, a) code that is woven into the documentation, and b) code that is not woven into the documentation. The TOP token clearly indicates this.

```
16|     token TOP {
17|         [
18|             || <pod>
19|             || <woven-code>
20|             || <non-woven-code>
21|         ]*
22|     } # end of token TOP
```

## 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-pod token

```
23|     token begin-pod {
24|         <leading-ws>
25|         '=' begin <hws> pod
26|         <ws-till-EOL>
27|     } # end of token begin-pod
```

### The end-pod token

The end-pod token is much simpler.

```
28|     token end-pod { <leading-ws> '=' end <hws> pod <ws-till-EOL> }
```

## Replacing Pod6 sections with blank lines

Most programming applications do not focus on the structure of the executable file, which is not meant to be easily read by humans. Our tangle would replace all the Pod6 blocks with a single `\n`. That can clump code together that is easier read if there were one or more blank lines.

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 Pod6 comment immediately after the `=begin pod` statement. The comment can say anything you like, but must end with a digit specifying the number of blank lines with which to replace the Pod6 section.

```
29|     token num-blank-line-comment {
30|         <leading-ws>
31|         '=' comment
32|         \N*?
33|         $<num-blank-lines> = (\d+)?
34|         <ws-till-EOL>
35|     } # end of token num-blank-line-comment
```

## 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.

```
36|     token pod {
37|         <begin-pod>
38|         <num-blank-line-comment>?
39|         [<pod> || <plain-line>]*
40|         <end-pod>
41|     } # end of token pod
```

## The Code tokens

The Code sections are similarly easily defined. There are two types of Code sections, depending on whether they will appear in the woven code. See [below](#) for why some code would not be included in the woven code.

### Woven sections

These sections are trivially defined. They are just one or more `plain-lines`.

```
42|     token woven-code {
43|         [
44|             || <plain-line>
45|         ]+
46|     } # end of token woven-code
```

### Non-woven sections

Sometimes there will be code you do not want woven into the document, such as boilerplate code like `use v6.d;`. You have two options to mark such code. By individual lines or by delimited blocks of code.

```
47|     token non-woven-code {
48|         [
49|             || <one-line-no-weave>
50|             || <delimited-no-weave>
51|         ]+
52|     } # end of token non-woven
```

### One line of code

Simply append `# begin-no-weave` at the end of the line!

```
53|     token one-line-no-weave {
54|         ^^ \N*?
55|         '#' <hws> 'no-weave-this-line'
56|         <ws-till-EOL>
57|     } # end of token one-line-no-weave
```

### Delimited blocks of code

Simply add comments `# begin-no-weave` and `#end-no-weave` before and after the code you want ignored in the formatted document.

```
58| token begin-no-weave {
59|     <leading-ws>                # optional leading whitespace
60|     '#' <hws> 'begin-no-weave'  # the delimiter itself (# begin-no-weave)
61|     <ws-till-EOL>              # optional trailing whitespace or comment
62| } # end of token <begin-no-weave>
63|
64| token end-no-weave {
65|     <leading-ws>                # optional leading whitespace
66|     '#' <hws> 'end-no-weave'    # the delimiter itself (#end-no-weave)
67|     <ws-till-EOL>              # optional trailing whitespace or comment
68| } # end of token <end--no-weave>
69|
70| token delimited-no-weave {
71|     <begin-no-weave>
72|     <plain-line>*
73|     <end-no-weave>
74| } # end of token delimited-no-weave
75|
76| token code-comments {
77|     <leading-ws>
78|     '#' <rest-of-line>
79|     <![ / <begin-no-weave> | <end-no-weave> / ]>
80| } # end of token code-comments
```

## The plain-line token

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

```
81| token plain-line {
82|     :my $*EXCEPTION = False;
83|     [
84|         || <begin-pod>          { $*EXCEPTION = True }
85|         || <end-pod>           { $*EXCEPTION = True }
86|         || <begin-no-weave>     { $*EXCEPTION = True }
87|         || <end-no-weave>       { $*EXCEPTION = True }
88|         || <one-line-no-weave> { $*EXCEPTION = True }
89|         || $<plain-line> = [^^ <rest-of-line>]
90|     ]
91|     <?{ !$*EXCEPTION }>
92| } # end of token plain-line
```

And that concludes the grammar for separating Pod from Code!

```
93| } # end of grammar Semi::Literate
```

# The Tangle subroutine

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

```
94| #TODO multi sub to accept Str & IO::PatGh
95| 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 or passed from the subroutine MAIN.

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

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

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

First we will get the entire Semi-Literate .sl file...

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

## Clean the source

### Remove unnecessary blank lines

Very often the code section of the Semi-Literate file will have blank lines that you don't want to see in the tangled working code. For example:

```
sub foo () {
    { ... }
} # end of sub foo ()

# <== unwanted blank lines
# <== unwanted blank lines

# <== unwanted blank lines
# <== unwanted blank lines
```

So we'll remove the blank lines immediately outside the beginning and end of the Pod6 sections.

```
99|     my Str $cleaned-source = $source;
100|     $cleaned-source ~~ s:g{\=end (\N*)\n+} = "\=end$0\n";
101|     $cleaned-source ~~ s:g{\n+\=begin (<hws> pod) [<hws> \d]?} = "\n\=begin$0";
```

## The interesting stuff

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

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

...and iterate through the submatches and keep only the code sections...

```
103| #     note "submatches.elems: {@submatches.elems}";
104|     my Str $raku-code = @submatches.map( {
105|         note .key;
106|         when .key eq 'woven-code' | 'non-woven-code' {
107|             .value;
108|         }
```

## Replace Pod6 sections with blank lines

```
109|         when .key eq 'pod' {
110|             my $num-blank-lines =
111|                 .value.hash<num-blank-line-comment><num-blank-lines>;
112|             "\n" x $num-blank-lines with $num-blank-lines;
113|         }
114|
```

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

```
115|     } # end of my Str $raku-code = @submatches.map(
116|     ).join;
```

## Remove the *no-weave* delimiters

```
117|     $raku-code ~~ s:g{ <leading-ws> '#' <hws> 'begin-no-weave'      <rest-of-line> }
118|         = '';
119|     $raku-code ~~ s:g{ <leading-ws> '#' <hws> 'no-weave-this-line' <rest-of-line> }
120|         = "$0\n";
121|     $raku-code ~~ s:g{ <leading-ws> '#' <hws> 'end-no-weave'        <rest-of-line> }
122|         = '';
```

## remove blank lines at the end

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

And that's the end of the tangle subroutine!

```
124|     return $raku-code;
125| } # end of sub tangle (
```



# 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.

```
126| 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`.

```
127|     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.

```
128|     Str :f(:$format) is copy = 'markdown';
129|     #= The output format for the woven file.
```

### `$line-numbers`

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

```
130|     Bool :l(:$line-numbers) = True;
131|     #= Should line numbers be added to the embedded code?
```

`sub weave` returns a `Str`.

```
132|     --> Str ) is export {
133|
134|     my UInt $line-number = 1;
```

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

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

## Remove blank lines at the beginning and end of the code

### EXPLAIN THIS!

```
136|     my Str $cleaned-source = $source;
137|     $cleaned-source ~~ s:g{\=end (\N*)\n+} = "\=end$0\n";
138|     $cleaned-source ~~ s:g{\n+\=begin (<hws> pod) [<hws> \d]?} = "\n\=begin$0";
```

## Interesting stuff

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

```
139|     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...

```
140| #     note "weave submatches.elems: {@submatches.elems}";
141| #     note "submatches keys: {@submatches».keys}";
142|     my Str $weave = @submatches.map( {
143|         when .key eq 'pod' {
144|             .value
145|         } # end of when .key
146|
147|         when .key eq 'woven-code' {qq:to/EOCB/; }
148|             \=begin pod
149|             \=begin code :lang<raku>
150|                 { my $fmt = ($line-numbers ?? "%3s| " !! ' ') ~ "%s\n";
151|                     .value
152|                     .lines
153|                     .map($line-numbers
154|                         ?? {"%4s| %s\n".sprintf($line-number++, $_) }
155|                         !! {      "%s\n".sprintf(                $_) }
156|                     )
157|                     .chomp # get rid of the last \n
158|                 }
159|             \=end code
160|             \=end pod
161|             EOCB
162|
163|         when .key eq 'non-woven-code' {
164|             ''; # do nothing
165|             #TODO don't insert a newline here.
166|         } # end of when .key eq 'non-woven-code'
167|
168|     } # end of my Str $weave = @submatches.map(
169|     ).join;
```

### remove blank lines at the end

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

And that's the end of the `tangle` subroutine!

```
171|     return $weave
172| } # end of sub weave (
```

# NAME

Semi::Literate - A semi-literate way to weave and tangle Raku/Pod6 source code.

# **VERSION**

This documentation refers to Semi-Literate version 0.0.1

# **SYNOPSIS**

```
use Semi::Literate;  
# Brief but working code example(s) here showing the most common usage(s)  
  
# This section will be as far as many users bother reading  
# so make it as educational and exemplary as possible.
```

# DESCRIPTION

`Semi::Literate` is based on Daniel Sockwell's `Pod::Literate` module

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.