In this second session, we’ll start going over Perl in more detail, following order of material from *Learning Perl*.

We’ll look at **scalar data** (numbers, strings, variables), **expressions**, **control structures** and **input/output**.

A Note on Perl History

- **Perl 5**, 1994-now: complete interpreter rewrite, significant language extensions. Maintained, patched and extended for over 20 years.
- Since 2010: developers do annual releases of Perl 5, eg **Perl 5.16 in Spring 2012, Perl 5.18 in Spring 2013, Perl 5.20 in Spring 2014**. (DoC Linux machines: 5.18; DoC Windows: 5.16.3).
- **Perl 6**, 2001-????: fundamental redesign of Perl, a new programming language! BUT: never finished! Will it ever be?

Scalar Variables and Values

- A scalar variable is declared via `my $variablename`, the `$` is **always** present. A variable name starts with a letter, and then may continue with any number of letters, digits or underscores! Every character in the name is significant, as is case.
- Can write as `@{variablename}`, just like the Unix shell.
- Stores any single scalar value:
  - An integer.
  - A real number (floating point).
  - An arbitrarily long string.
  - A reference to some other piece of data.
- Integer literals may be written in decimal, octal or hexadecimal: eg. `-129`, `0377` and `0xffe0`.
- Real literals may be written in **exponential notation**, with or without a `decimal point`: eg. `6.27`, `-2.4e30`, `7e-5`.
- A string literal can be either delimited with single-quotes or double-quotes. Different rules apply inside the two kinds of string:

  Single-Quoted String Literals

  - A single-quoted string, eg `'hello there'`, is an arbitrarily-long sequence of characters placed inside single quote marks.
  - The quotes are not part of the string - rather, they delimit both ends of the string.
  - The shortest (empty) single-quoted string is `'`.
  - Each character inside the quotes is placed into the string exactly as-is - even a literal newline!
  - There are only two exceptions to this:
    - A single-quote on it’s own would terminate the current string - to embed a single quote in a single-quoted string, write `\'`.
    - A backslash may be embedded in a single-quoted string as `\`.
  - Apart from these exceptions, nothing is modified inside a single quoted string.
  - In particular, `$` symbols are embedded as-is, and C-style escapes like `\n` do not function in a single-quoted string.
A double-quoted string literal, like "hello $name\n", is much more powerful. Again, it is an arbitrary sequence of characters which can be split across lines.

However, the backslash can now specify several important control or escape operations as follows:

- Newline (\n)
- Tab (\t)
- Carriage Return (\r)
- Ring bell (\a)
- Any octal ASCII value (7*8+2 = 58 = ':')
- Any hexadecimal ASCII value (6*16+13 = 109 = 'm')
- Backslash (\)
- Dollar ($) 
- A double-quote (")
- Lower-case the next letter (\l)
- Upper-case the next letter (\u)
- Start lower-casing the rest of the string (\L)
- Start upper-casing the rest of the string (\U)
- Stop lower- or upper-casing the rest of the string (\E)

For example, "hello \Uthere\EU how are you" generates "hello THERE how are you".

Also in a double-quoted string...

 expressions in Perl are very like those of C, but without the pointer operators.

Some operators expect numbers, others expect strings. Perl converts one to the other when needed. Thus the string "17.3e7" converts to the number 17.3e7 when needed, and the number 31.53 converts to the string "31.53" when needed.

Perl provides all the obvious +, -, *, / arithmetic operators. Unlike C, all operators operate in floating point.

Plus, the exponentiation operator (written **). For example: 3.25 ** 4 (ie. 3.25^4).

The modulus operator: 10%3 gives the remainder when 10 is divided by 3. Both values are truncated to integers before this operator is applied.

A set of numeric comparison operators just like C: <, <=, ==, >=, > and != (not equals).
Expressions

- Alphabetical string comparison via a separate set of comparison operators: `<`, `le`, `eq`, `ge`, `gt` and `ne`.
- A common mistake is comparing numbers with `eq` or strings with `==`, allowed because strings and numbers are converted back and forth freely.
- For example, "0" == "000" is true, but 0 eq "000" is false. (The string "0" which is not the same as "000"!)
- String concatenation done with the fullstop operator (.), eg:
  "hello " . $name . "\n"
  (Warning: No space is added automatically between the strings!)
- Concatenation is not used as much as you would expect - interpolation is often used instead. Write the above as:
  "hello $name\n"

Useful string repetition operator:
"fred" x 3 gives "fredfredfred"
(RHS truncated to an integer first).

Assignment

- The most vital operator of all is assignment (`=`). For example:
  ```perl
  $x = 37;
  $y = $x * 7 + 5;
  $z = $z * 3;
  ```
  - The first simply sets $x to 37.
  - The second calculates $x * 7 + 5 (using the current value of $x) and then stores the result in $y.
  - The last example takes the current value of $z, multiplies it by three, and stores the result back in $z.
- An assignment also has a value, which means you can nest or chain assignments:
  ```perl
  $x = $y = 17;
  $y = 5 * ($a = 7 + $x);
  ```
  - The first example sets both $x and $y to 17.
  - The second example means: evaluate $7 + $x and store the result in $a, then multiply $a by 5 and store the final result in $y.

Autoincrement and Autodecrement

- To increment or decrement $a, write:
  ```perl
  ++$a;
  --$a;
  ```
  - Prefix and postfix form of these operators: `++$a`, `$a++`, `--$a` and `$a--`, work exactly like C and Java.
  - Probably a bad idea to embed these in much larger expressions.

Boolean Expressions

- Expressions used in a boolean context are first evaluated as numbers or strings.
- When the result has been calculated: zero, the empty string or the undefined value are false, anything else is true!

There are a few extra boolean operators:

- `!` is a unary operator meaning not.
- `||` and `&&` come straight from C – these can be written as ‘or’ and ‘and’ in modern Perl.

Just like C and Java, they short-circuit:

- Consider:
  ```perl
  $a < 7 || $b > 6
  ```
  - If $a is less than 7, then the whole expression is bound to be true.
  - No point in evaluating the rest of the expression - so Perl doesn’t!
- Similarly:
  ```perl
  $s eq "hello" && $t ne "bonjour"
  ```
  - If $s is not "hello" then the whole expression must be false.
  - Again, no point in evaluating the second half.
Statements and Control Structures

- Any expression can become a statement by appending `;`.
- SO `3+4;` means *evaluate expression* (result: 7) and *discard the result*. A pointless statement!
- Usually, expression-statements have side-effects:
  - Assignments (the basic `=`, binary assignment operators like `+=`, and `++` and `--` operators).
  - Function/procedure calls (eg. `print` - or your own functions).
- Perl provides single statement modifiers:

  ```perl
  <statement> if <EXPR> ;  
  <statement> unless <EXPR> ;  
  <statement> while <EXPR> ;  
  <statement> until <EXPR> ;
  ```

  - `unless` is a synonym for `if !`; `until` a synonym for `while !`.
- A sequence of statements may be enclosed inside `{}` braces forming a *block*. NB: no `;` after the `{}` of a block.

Perl provides a conventional if.. elsif.. elsif... else statement, to choose between two or more alternatives:

```perl
if ( $i < 20 || $j > 7.4 ) {
  print "case one\n";
} elsif ( $i > 40 && $j > 0 ) {
  print "case two\n";
} else {
  print "case three\n";
}
```

Note that the brackets – both round and curly – are compulsory on an *if* - and the loop below.

- Historically, Perl did **NOT** provide a special case/switch statement for multi-way comparisons. Perl 5.10 added ‘given/when’, but you have to ask for 5.10 features to be enabled:

```perl
use v5.10; # at the top of the program, enable 5.10 features
...  
given( $x ) {  
  when( "hello" ) { print "hello\n"; }
  when( /^Dun\[ck]/ ) { print "dunc\n"; }
  default { print "unknown\n"; }
}
```

Perl also provides a conventional test *at the top* while loop:

```perl
my $w = $x; my $h = 1;  
while ( abs($w-$h) > 0.001 ) {  
  $w = ($w+$h)/2;  
  $h = $x/$w;  
}
```

The above algorithm happens to find the square root of $x!

Perl provides a test *at the bottom* loop, two forms:

```perl
for( my $sum = 0, my $i = 1; $i <= 10; $i++ ) {  
  $sum += $i;  
}
```

- A **C-style counting loop**:

```perl
for( my $sum = 0, my $i = 1; $i <= 10; $i++ ) {  
  $sum += $i;  
}
```

- A very useful *for each element in a list/array loop*:

```perl
foreach my $x ( 1, 3, 7, 6, 54 ) {  
  $total += $x;  
}
```

A *very useful* for each element in a *list/array loop*:

```perl
foreach my $x ( 1, 3, 7, 6, 54 ) {  
  $total += $x;  
}
```

(We’ll see more about arrays and lists next session).

Duncan White (CSG) Introduction to Perl: Second Lecture January 2015 13 / 22

Statements and Control Structures

- Perl provides a conventional if.. elsif.. elsif... else statement, to choose between two or more alternatives:

```perl
if ( $i < 20 || $j > 7.4 ) {
  print "case one\n";
} elsif ( $i > 40 && $j > 0 ) {
  print "case two\n";
} else {
  print "case three\n";
}
```

Note that the brackets – both round and curly – are compulsory on an *if* - and the loop below.

- Historically, Perl did **NOT** provide a special case/switch statement for multi-way comparisons. Perl 5.10 added ‘given/when’, but you have to ask for 5.10 features to be enabled:

```perl
use v5.10; # at the top of the program, enable 5.10 features
...  
given( $x ) {  
  when( "hello" ) { print "hello\n"; }
  when( /^Dun\[ck]/ ) { print "dunc\n"; }
  default { print "unknown\n"; }
}
```

Perl also provides a conventional test *at the top* while loop:

```perl
my $w = $x; my $h = 1;  
while ( abs($w-$h) > 0.001 ) {  
  $w = ($w+$h)/2;  
  $h = $x/$w;  
}
```

The above algorithm happens to find the square root of $x!

Perl provides a test *at the bottom* loop, two forms:

```perl
do {  
  $y *= $x;  
  $y *= $x;  
  $x += 2;  
  $x += 2;  
} while( $x < 15 );  
```

- A **C-style counting loop**:

```perl
for( my $sum = 0, my $i = 1; $i <= 10; $i++ ) {  
  $sum += $i;  
}
```

- A very useful *for each element in a list/array loop*:

```perl
foreach my $x ( 1, 3, 7, 6, 54 ) {  
  $total += $x;  
}
```

(We’ll see more about arrays and lists next session).

Duncan White (CSG) Introduction to Perl: Second Lecture January 2015 13 / 22

Statements and Control Structures

- Perl provides a conventional if.. elsif.. elsif... else statement, to choose between two or more alternatives:

```perl
if ( $i < 20 || $j > 7.4 ) {
  print "case one\n";
} elsif ( $i > 40 && $j > 0 ) {
  print "case two\n";
} else {
  print "case three\n";
}
```

Note that the brackets – both round and curly – are compulsory on an *if* - and the loop below.

- Historically, Perl did **NOT** provide a special case/switch statement for multi-way comparisons. Perl 5.10 added ‘given/when’, but you have to ask for 5.10 features to be enabled:

```perl
use v5.10; # at the top of the program, enable 5.10 features
...  
given( $x ) {  
  when( "hello" ) { print "hello\n"; }
  when( /^Dun\[ck]/ ) { print "dunc\n"; }
  default { print "unknown\n"; }
}
```

Some Simple I/O

- Reading a line of input

Let's consider getting input from the keyboard, and reporting results to the screen.

- The <> operator (pronounced diamond or readline) fetches a line of input from a filehandle - STDIN is a filehandle, which represents the standard input (normally the keyboard). So:

  ```perl
  $name = <STDIN>;
  ```

  reads an entire line of input, up to and including a newline, and then stores the input in the variable (including the newline).

- After reading a line, you usually want to get rid of the trailing newline. Two operators:

  - Perl 4 provided `chop($name)` which removes the last character from the named variable and returns the character removed.
  - Perl 5 added `chomp($name)` which deletes a trailing newline or does nothing.

  Use `chomp` everywhere; it’s safer and more portable (it will remove the newline from the string even if the newline convention on the OS you’re using is not a single character!).
Some Simple I/O

Printing the results

print(), combined with the variable interpolation in double-quoted strings, gives us a lot of formatting control - although Perl also has a C-style printf() function (see perlfunc -f printf) to give you even better control such as 3 decimal places for $z:

```perl
print "debug: x=$x, y=$y, z=$x\$y\n";
print "debug: x=$x, y=$y, z=\$.3f\n\", $z;
```

Here's an example (eg1), allowing you to investigate different expressions:

```perl
print "Please enter x: "; my $x = <STDIN>; chomp $x;
print "Please enter y: "; my $y = <STDIN>; chomp $y;
my $z = $x + $y;
print "\n$x + $y = $z\n";
```

You may like to spend some time trying some examples of string, numeric and boolean expressions.

Perl 5.10 added a new function 'say' - basically 'print' with a newline added - again, you have to ask for it:

```perl
use v5.10; # at the top of the program, enable 5.10 features
...
say "debug: x=$x, y=$y, z=$z";
```

End of File

We've seen that $line = <STDIN> reads a whole line of input from stdin.

But what if there's no more input? Files have a concept of end of file (eof) - after you have read the last line of data!

So does the keyboard: On Unix systems, end-of-input is signalled by the user typing CTRL-D on a line on its own.

No CTRL-D character (ASCII code 4) is delivered; CTRL-D is a terminal control signal that input has ended.

When eof occurs, <> returns the empty string. This never otherwise occurs - why? Because <> leaves the newline in!

So test for eof by comparing the result against the empty string:

```perl
if( $line eq "" )...
```

More typically, test for not eof:

```perl
if( $line ne "" )
```

Or simply

```perl
if( $line )
```

Can combine line reading, assignment, and test for not eof into one:

```perl
if( my $line = <STDIN> ) # if we've read a line..
```

Can even declare $line at the same time:

```perl
if( my $line = <STDIN> ) # if we've read a line..
```

File Input/Output

Reading a File

Perl can read the contents of named files (or Unix pipelines), create/overwrite a named file with new contents, and append new output on the end of a file.

Suppose we wish to read a file called fred. The first step is to create a filehandle like STDIN but connected to the file "fred":

```perl
open( my $in, '<', "fred" );
```

If the file "fred" doesn't already exist in the current directory, open() fails, setting $in to undefined (boolean false) and returning 0 (another boolean false). So you might handle an error by:

```perl
open( my $in, '<', "fred" );
or 
unless( open( my $in, '<', "fred" ) )
{
    # handle error...
}
```

A common way of handling the error is to print an error message (on STDERR) and exit - use die:

```perl
unless( open( my $in, '<', "fred" ) )
{
    die "can't open fred\n";
}
```

For Each Line Idiom

Replacing 'if' with 'while' gives us the for each line from stdin idiom:

```perl
while( my $line = <STDIN> ) # for each line from stdin
{
    chomp $line;
    # now process $line
}
```

Let's see an simple example of that (eg2):

Write a program that reads a list of numbers (one per line) from STDIN, adds all these numbers up, and prints the total.

```perl
my $sum = 0;
while( my $line = <STDIN> ) # for each line from stdin
{
    chomp $line;
    $sum += $line;
}
print "total: $sum\n";
```

Above the while loop, we initialize: my $sum = 0;

Processing is: $sum += $line;

At the end, add print "total: $sum\n".

Giving us:

```perl
my $sum = 0;
while( my $line = <STDIN> ) # for each line from stdin
{
    chomp $line;
    $sum += $line;
}
print "total: $sum\n";
```
Better still, use the *Die unless Happy* idiom:
```perl
die "can't open fred\n" unless open( my $in, '<', "fred" );
```

Best of all, we can use the common *Do or die* idiom. This wouldn’t work without short-circuit evaluation:
```perl
open( my $in, '<', "fred" ) || die "can't open fred\n";
```

Whichever variant we use, after a successful `open()`, `$in` is a filehandle, just like `STDIN`, from which we can read data.
```perl
my $line = <$in>
```
reads the next line of input from `fred`. When there is no more input, `$line` will contain the empty string.

A typical *read every line* program looks like **eg3**:
```perl
open( my $in, '<', "fred" ) || die "can't open fred\n";
while( my $line = <$in> ) {
    chomp $line;
    print "read '$line'
";
}
close( $in );
```

Aside: As well as `die` there’s a function `warn` which prints a message to `STDOUT` but doesn’t exit.

---

**Writing a File**

To open a file for writing (overwriting the old contents):
```perl
open( my $out, '>', "bob" ) || die "sob.. can't create bob\n";
```

Now write new data into `bob` by any number of:
```perl
print $out "...some data...\n";
```

As with reading, don’t forget `close($out)` when you finish writing data.

We can append data to `bob`, instead of overwriting:
```perl
open( my $out, '>>', "bob" ) || die "sob.. can't append to bob\n";
```

You can also open a pipe to/from a Unix pipeline for reading/writing:
```perl
open( my $in, '>-', "ls | sort -r" );
```

We can write data to the pipeline:
```perl
open( my $out, '|-', "expand" );
```

You can also do more advanced things like open files for random access I/O, open pipelines feeding data in and receiving data back etc. See `perldoc -f open` and `perldoc IPC::Open2` for more details.