

# How to Perl

By Phil Menger.

This item is not about dropped stitches or how to do the Continental knit stitch. It's about the BCLA Continuing Education Workshop *Introduction to Data Manipulation Using Perl* held April 12<sup>th</sup> and 13<sup>th</sup> at the Brighthouse Branch of the Richmond Public Library. Let me start by assuring you I am very much a newbie when it comes to programming code. The beauty of this Perl basics workshop was you did not need to know programming language or code of any kind.

## The Instructors

John Durno (University of Victoria)  
Mark Jordan (Simon Fraser University)

## What is Perl/perl?

Perl is a general purpose programming language with roots in text processing. It was originally called "Pearl" but the name conflicted with another programming language of the same name. Perl is the name of the language and small case perl is the software (Durno & Jordan, 2012).

## Big Data

Today the talk is all about Big Data and Perl is one tool which helps you manage big data sets. You can use Excel spreadsheets for managing large chunks of data but Excel can only do so much with the data.

A specialized tool to manipulate data is Perl. Another one is MARCEdit. As MARCEdit inventor Terry Reese demonstrated at the May 4<sup>th</sup> BCCATS (British Columbia Cataloguing and Technical Services Interest Group) *Advanced MarcEdit* workshop MARCEdit is a fine tool, and Perl instructors Durno and Jordan emphasized this was not an either or situation.

With Perl you write you own program language that allows you to choose the exact way you wish to manage that date. While I attended only a half day of the MARCEdit workshop I was impressed, as a newbie, with what that program could do. I had only passing familiarity with MARCEdit: knowing it existed, how to find it, download it and use it. Reese's one day class was a more advanced MARCEdit workshop and we were lucky to get him. He is clearly a rock star in the library community. But back to Perl.

## Using Perl

Anyone who has ever performed rudimentary HTML or any other programming language will recognize the procedures. Certain symbols, letters and commands can make the data do various things at your command. Even I could take a set of MARC data and transfer it into Perl for further manipulations.

As the instructors explained, one reason you may want to use Perl to write your own script instead is that Perl can extract data from a large variety of data sources, such as relational databases, web pages (depending on the regularity of the HTML), and structured data such as text files, and then combine all of this data in useful ways. Because of the amount of control that Perl and other programming languages provide, with a little effort it is possible to solve a greater variety of problems than most ready-made applications can (Durno & Jordan, 2012). In other words, Perl (or any other programming language) allows you to create your own tools. Learning how to perform these tasks however requires a more intermediate or advanced course beyond the scope of this workshop.

This was a great introduction to Perl and I commend the instructors for their infinite patience and ability to be right at your elbow to assist with any wrinkles you may encounter. Their methods were comprehensive and thorough. If you have a project that involves handling big data you may wish to catch this workshop again when it is offered. In the meantime John Durno and Marc Jordan have provided a wonderful overview at

<http://interoperating.info/courses/perl4data/>

## References

Durno, J. & Jordan, M. (2012). *Introduction to data manipulation Using Perl*. Retrieved from <http://interoperating.info/courses/perl4data/>

*Phil Menger is a former journalist, corporate suit and new Library Technician. He curates information for the Library Technicians' and Assistants' blog (<http://ltaig.wordpress.com>) for his weekly column Phil the Web Crawler.*