# 1 TERMIUMplus trilingual database

# 1.1 Jay Lawrence <jay (at) lawrence.net> exclaimed:

- Date: Tue, 08 Jan 2002 13:55:07 -0800 (PST)
- URL: http://www.termium.com/

```
TERMIUMplus (www.termium.com) is a trilingual application that
allows translators and terminologists to search a collection of
1.5 million entries in English, French and Spanish. The system is
freely available to any employee of the Canadian Federal government as
well as by subscription to individuals and organizations outside. The
terms and the user interface are both trilingual.

mod_perl plays an integral role in the success of this system.
Because the server experiences significant amounts of traffic during
the middle of the day effecient request handling is of paramount
concern. It is not uncommon to be servicing over 100 concurrent
requests at 2pm. Not only does the system perform very well but it is
also very stable. I don't think our httpd's have ever crashed - and
almost all requests are in the sub-second response range.

If great performance and stability were not enough - mod_perl (Perl) -
has allowed us to provide a very easy to use and enjoyable interface
to our database servers. The servers are actually on NT running a
proprietary database software package.  The database software is very
good at performing both full text and exact term searches of the term
data. However, the software interface to the databse engines is weak
and unusable at best. By using Perl to talk to the database server's
HTTP interface we were able to extract the desired results data and
then use Perl's power to reformat the results into something pleasing
and tailored to the user's preferences. Because each record has over
100 fields and each field can have a number of sub components - I
don't think the job would be doable in any other language than Perl!
In addition to reformatting the output of the database we also employ
some processing of search terms. This processing is unique to our data
collection but helps increase recall by eliminating stopwords such as
"a", "an", "le", "les", etc.

In addition to the fancy user interface TERMIUMplus also offers a
server-to-server term translation service. This allows other search
engines to offer on-the-fly term translation as part of their
service. An excellent feature when dealing with a bi or tri-lingual
document corpus. You are welcome to see this yourself by visiting:

  http://strategis.ic.gc.ca/engdoc/search.html

Check on Bilingual search and try a word such as "turbofan". As a
note, I am not aware of what software the Strategis search system was
built with.

The entire system runs on a dual processor Sun 250 with 2G of RAM (We
discovered how important lots of RAM is for this level of concurrent
user activity) for the front end of the request processing. For the
database queries we have 2 quad Xeon NT boxes which we divide between
Extranet and Internet traffic. We will be replacing the Sun 250 with a
```

quad processor Sun 450 with 8G of RAM.

In addition to mod_perl we use MySQL as our user sessions database and intend to start replacing many functions of our proprietary back end database with functions developed using mod_perl and MySQL. Linux is our front-line development system and CVS is our versioning management system. We use CVS to then move our work on to a Sun staging system for pre-release testing and then finally rsync to push final code on to production servers. All of our code runs as well on Linux as it does on Solaris - with no modifications other than compile time options for the major packages of the application.

I feel that using mod_perl to build TERMIUMplus has allowed for the construction of a high quality service which is capable of handling a significant user load. It is very rare (never?) that we experienced any major problems with the Apache, mod_perl, and Perl portion of our system. Most of our operational difficulties are coming from our vendor supplied software at the database backend where daily server problems are experienced.

Software costs aside I wouldn’t build this appliation using anything but mod_perl, Apache and MySQL!

# Table of Contents: