

of course[®]

VOLUME 3, NUMBER 1

SPRING 2003

**We're letting
the cat out
of the box**



The loo is done!

Plus the latest news on
Perl • PHP • MySQL • Java • Tcl/Tk



SPRING CLEANING

A nip, a tuck, and a bit of a facelift

- new logo
- new lounge
- new loo
- new letterhead
- new...lots more

by Graham Ellis

Normally people do their Spring cleaning at about the same time the flowers are just starting to emerge and the sunshine is actually warming rather than just merely lighting. We've been making revisions all along; we're just choosing to devote our Spring issue to telling you about it because we've come a long way since last Spring. Those who have not come on courses since November 2002 will see the biggest change, and we can't wait to show you!

With this issue comes our official launch of a new logo and product design. And as ever, the course quality only improves with each instance.

We make these constant changes because we pride ourselves for having "Practical Quality". In this issue, you'll note a number of technical articles covering software quality – code that does its job well and is easy to support, maintain and develop further. These aspects of programming are often mistakenly overlooked; up to 80 percent of the cost of bespoke software can be in the ongoing attention that it requires, rather than in the original writing!

We also pride ourselves in providing practical quality on our training courses too – courses that are up to date, covering the subjects that you want to cover, incorporating an excellent tutor-to-student ratio.

Our training centre facilities have been outfitted with you in mind. Every student who attends a public course has exclusive use of a laptop running a choice

of Linux, Mac OS X or Windows XP with a 1 Mbit broadband link to the Internet for the duration of the course. Wireless, 100 base T, 10 base T and analogue modem connections are all available

We've never varied on our quest for practical quality; this is why we're constantly going through changes and upgrades. It's not that the product is bad, it's that it can always be made better.

for trainees who bring their own laptops too. All trainees' work is emailed back to them upon course completion, and all the tutor's examples are available on the Internet. Even the manuals covering the subject of the course are in much more detail than you might expect.

For those who saw the front cover and smiled, you've no doubt been to our training centre. Yes, we're happy to report that our customer loo is finally finished. And like any proud parent, we wanted to show off our pictures. But the tour isn't stopping at the loo. If you haven't been here for a while, you might also be interested in the other things we've been up to. See pages 4 & 5 for photos and a peek at what we've got planned coming up.

Oh, and another thing about that cover: If you've scratched your head in confusion, then you just might be in for a different sort of experience if you do happen to book on a course here. And the cat? She's part of our centre.

Just in case you do not dare to ask us ...

We thought it looked like a piece of art, so what better place than to show it off but in our loo! It has its function, but not easily guess-able. And those too polite to ask may want to know what it is, just in case.

This must-have toilet brush and holder (right) was begged off the display from exhibitor Roman Dietsche GmbH & Co. KG (www.dietsche.com) at 100% Design in London, showcasing design innovations.



Public Sector Enemy #1: putting your ps data online

Local Government and the Health Service have been set stiff targets to get their information on line in the next few years. As a result, Public Sector staff are finding themselves having to learn how to tell their Flash from their JavaScript, and their Rubies from their Perls. But where to start? It's confusing enough without having certain vendors try to sell you on their own technology, never mentioning that another product might suit your use better.

Our "Getting your Public Sector Data Online" course takes public employees who are aware of the Web and Intranets as users, explains which technology would be right for their own use, and gives them a good understanding of how all the various components fit together. We're unbiased about all the competing technologies because we teach most of them, use even more of them ourselves and have a good understanding of almost all of them.

A WHC COURSE EXTRA FOR PUBLIC SECTOR WEBSITES



It's one thing to be putting flat data, such as your opening hours, online, and quite another to provide facilities to let the public make appointments via the Internet. This one-day course takes you behind the scenes, introduces the technologies available, and helps you understand which will be best for you.

Getting Your Public Sector Data Online

Wednesday, March 12, 2003 - or - Thursday, May 8, 2003

Melksham, Wiltshire, England

a one-day, purpose-run course targeted to your needs

£290.00 plus VAT (£200 each for 2nd and subsequent on same invoice)



404, The Spa, Melksham, Wiltshire SN12 6QL 01225 708 225 www.wellho.net

graham@wellho.net

When size matters in your field

MySQL is an example of a relational database. It typically provides a bewildering choice of field types you can specify in your tables, so you'll have to consider very carefully which one to use.

If you have a bit of text to hold, should you use a CHAR, a VARCHAR or a TEXT? To answer this, you'll need to think further about the characteristics of your bit of text (or "blob").

If you use a CHAR (for example CHAR(30)), MySQL will allocate 30 bytes in every row for your text. Any shorter data that you enter will be space filled. An excellent example of a piece of data for which CHAR is well suited is a part number or part code, especially if it has a fixed length.

A VARCHAR (for example VARCHAR(30)) allocates up to 30 characters for your text, so it's more efficient in memory use, but it does apply a strict limit. Use a VARCHAR for a line that's to be printed on an adhesive label for instance. You must not exceed the maximum number of characters so that you don't run off the label.

TINYTEXT limits your field width to 255 characters.

If you want more than that, TEXT allows up to 65,535 characters. For example, use TINYTEXT to hold someone's surname, and use TEXT to hold a complete email header.

Techbuzz Quickdefinitions

Data Mining

Gathering information about visitors to a web site in order to tailor the presentation of the site to them later in their visit and on subsequent visits.

Content Management System

A way of entering and managing data (usually with the intent of presenting that data through a web site) that provides a consistent, easy-to-use and easy-to-administer interface for both the data provider and the data accessor.



String types	Max. number of characters
char	255
varchar	255
tinyblob	255
tinytext	255
blob	65535
text	65535
mediumblob	16777215
mediumtext	16777215
longblob	429496295
longtext	429496295

Global technology and trend watching

Whenever staff from our company visit other countries, particularly the USA, we browse in bookshops, such as Borders and Barnes and Noble, at length. Computer stores such as CompUSA don't get bypassed either. Just what are we looking for?

Trends. What product fills the bookshelves? How many books are on a particular topic compared to another? Where is the interest in technology next heading?

A recent excursion to Washington DC revealed:

- A big increase in books on relational databases in general, and on MySQL in particular
- A big increase in the number of books covering PHP
- An increase in the number of books on more advanced web

Data Protection buyer beware

In the UK, the Data Protection act of 1998 requires that Data Controllers register with the Data Protection Registrar, for an annual fee of £35.00, and is not VAT-able. Registration involves completing a number of forms that describe where you get your data and what you do with it.

We received a "Final Notice" the other day, asking for a fee of

£95.00 + VAT to register (ignoring the fact we've been registered all along). The letter looks quite convincing and a number of people have been misled into paying it.

If you have received one of these notices we strongly suggest you look online to see what the Information Commissioner has to say at <http://www.dpr.gov.uk/donotbemised.html>.



topics that are cross-language, such as Content Management Systems, E-Commerce, Data Mining, Network and web site security, etc.

- Java remains heavily featured on the shelves
- There remains significant coverage of Perl (but no big increase)
- Tcl and Ruby remain very much niche topics. You might be able to find the books if you look hard enough, and unfortunately, they are usually mis-filed
- Online sales of books (ie Amazon) seem to have clearly affected some stores; Borders and Barnes and Noble are still carrying extensive stocks, but the book section at CompUSA has shrunk, as has coverage at other

bookstores such as Waldenbooks.

On a previous trip to the US last year, there was noticed a huge increase in wireless networking and broadband products when there were just a few here in the UK. The interest for these technologies in the US is still there, but the UK has now caught up to it, indicated by what's available in the UK shops today.

There seems to be a growth of Apple products available in the shops. Macintosh OS X is working hard to bite a chunk out of the sales of other operating systems.

We weren't taken by surprise by any new types of products on this trip. Is this perhaps indicating the breakneck progress of hardware is slowing down a little?

Ah well. It gives us time to catch up to it all!

The editors of THE FUTURIST present "Five Tips for Trend Watching" excerpted at <http://www.wfs.org/fivetips.htm>.

Introducing:



our loo

Three years on the drawing board and six months in the making.

It's done.



Our training centre is in a Georgian Spa House on the outskirts of Melksham, Wiltshire. When we bought 404 it was in a sorry state, with collapsing ceilings, floors that you were likely to go straight through, one power point, no central heating, one water tap (that came through the adjoining building) and no toilet at all! That's all changed

– it's been refurbished from below the ground up and it now includes a training centre like no other.

We've tried to think of all the details, giving you maximum comfort in an environment you feel at home in. We haven't scrimped on technology either, as our 1Mbit broadband connection is hooked up to all the training machines.

All our courses are run at our Melksham training centre, which is also available for tailored private courses for groups of up to eight students.

Whether you come from near or far, on a beginner's or an advanced course, we're determined to make our courses an enjoyable experience as well as an educational one.

From Humble and Bumble Beginnings



For most of you, what you saw of this customer loo was a closed door. Those caught by curiosity will probably remember a cleared and just-plain-plaster, unpainted, clean room.

The picture (far left) was taken at the very beginning, the day we bought the centre.

Before the centre opened (left), just after the first interim ceiling collapsed, we had to discard the second, honey-drenched and badly damaged ceiling so as to remove several seasons of beehives, both honey and paper, and large amounts of bee glop.

Oh yes, there were live bees swarming as well!

Sneak Preview

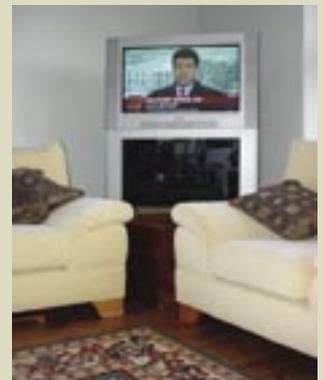
We're in the middle of more exciting projects as well and in future issues we'll give you a closer look as we progress. But for now, here are some previews:



Our library includes more than 300 technical books. Trainees are welcome to browse the books while they're with us, and borrow books overnight if they wish.



Our entire technical library is recorded on a vast database, using PHP and MySQL, which can be found online on our website. We'll even tell you how we did it on the relating course.



Our customer lounge is available from 08:00 on course mornings with coffee, snacks and soft drinks; the day's newspaper is there for the reading, and BBC News 24 or similar is on the TV in the background as students arrive.

Want to take on line payments?

Interfacing to online payment systems such as Barclay's EPDQ isn't as hard as you might imagine. If you have a web site from which you want to accept online payments, it will need the following facilities as a minimum:

- a) Server-side programming support in Perl or PHP (some other languages are acceptable alternatives; check with your EDPQ sponsor for a list of these)
- b) A fixed domain name and URL from which orders will be checked out (you need a single checkout page, and it must not be an IP address-based page)
- c) The capability to act as a client to another web server (if you have access to the LWP::UserAgent module in Perl, it provides this ability; it comes standard in PHP)
- d) A directory on your server that can contain a Perl or PHP script to which

continues on page 8

The management facilities of your bank's online order processing system will allow you, the account operator (with appropriate passwords), to enter orders directly without using your own site at all. If you accept postal, fax and telephone orders as well as orders via the net, you can enter all payments directly and get rid of that old card-swiping machine. If you require "cardholder present" facilities (e.g. if you have a shop), though, you might find that your customers prefer the swiping machine at present.

If going the whole way to an automatic payment-on-order isn't practical for you, there are alternatives:

- Online selection of goods, payment details by phone or fax
- Online selection of goods, then you email your customer requesting their payment via the bank's system
- Online selection of goods and entry of payment details via your own secure site (be careful here, there are many security issues)

You should strongly discourage your customers from sending you their payment details in unencoded emails, and you should not provide a form that invites them to send their details via an open (http rather than https) form.

Although the majority of our business is in providing training, we do undertake a small number of contracts using the most up-to-date technologies to keep our finger on the pulse.

We also learn these technologies so that we can in turn train you (or write about it in our newsletter). Contract work we have undertaken recently includes a web site that interfaces to Barclay's EPDQ system to accept automated online payments.

**direct
entry**

**we
did it
too**



Setting up online
e-commerce is not
as daunting as you
might think.



THE SPA B&B

CONIGRE FARM

THE OLD MANOR

Local accommodation around Melksham

People attend courses at our Melksham training centre from far and wide, and there's a good range of well priced accommodation available nearby. Whilst we try to ensure the accuracy of these details, we might

not be informed of any changes, so please verify this information with the establishment when you book your own accommodation. Pricing shown is for single occupancy; many offer family and double occupancy facilities.

There are also self catering, caravan and camping facilities nearby, as well as other hotels, guest houses and B&Bs. Melksham Tourist Information Centre (01225 707 424) can provide you with up-to-date details.

1 The Spa B&B 402, The Spa, Melksham SN12 6QL H. Pavey P: 01225 707 984	distance from centres (mi)		phone in room	restaurant on premises	takes credit cards
	training	town			
	0.1	0.7			
£20.00-25.00 single, one room en-suite, includes breakfast Large comfortable listed Georgian home. Quiet position with open views. Ample parking.					
5 The Old Manor 48 Spa Road, Melksham SN12 7NY L. Bailey P&F: 01225 793 803 E: theoldmanor@yahoo.co.uk	distance from centres (mi)		phone in room	restaurant on premises	takes credit cards
	training	town			
	0.5	0.3			
£25.00-35.00 single, one room en-suite, includes breakfast Set in 3/4 acre, delightful gardens; large English breakfasts a speciality; lovely decor.					
7 Conigre Farm Semington Road, Melksham SN12 6BZ D. Edwards P: 01225 702 229 F: 01225 707 392 W: www.cfhotel.co.uk	distance from centres (mi)		phone in room	restaurant on premises	takes credit cards
	training	town			
	0.9	0.1			
£43.00-52.00 single, all rooms en-suite, SkyTV, includes breakfast Family-run hotel and restaurant offers finest foods, quality accommodation and friendly environment.					
14 The Kings Arms Hotel 20 Market Place, Melksham SN12 6EX P: 01225 707 272	distance from centres (mi)		phone in room	restaurant on premises	takes credit cards
	training	town			
	0.8	0.0			
£35.00 single, all rooms en-suite + £6.95 full/£4.95 continental bkfst Recently completely renovated; located in the heart of town.					
40 Longhope Guest House 9 Beanacre Road, Melksham SN12 8AG D. Hyatt P&F: 01225 706 737 E: longhope@aol.com	distance from centres (mi)		phone in room	restaurant on premises	takes credit cards
	training	town			
	2.0	0.9			
£28.00 single, all rooms en-suite, includes breakfast Large Victorian house set in its own grounds; warm and friendly, well decorated.					
48 Beechfield House Beanacre, Melksham SN12 7PU P: 01225 703 700 F: 01225 790 118 E: csm@beechfieldhouse.co.uk	distance from centres (mi)		phone in room	restaurant on premises	takes credit cards
	training	town			
	2.7	1.6			
£80.00-90.00 single, all rooms en-suite, includes breakfast Comfortable country house, 21 en-suite bedrooms, four-poster beds available, heated outdoor swimming pool.					
42 The Shaw Country Hotel Bath Road, Shaw, Melksham SN12 8EF J. Lewis P: 01225 702 836 F: 01225 790 275	distance from centres (mi)		phone in room	restaurant on premises	takes credit cards
	training	town			
	2.8	1.7			
£46.00 single, all rooms en-suite, includes breakfast A 400-year-old farm house, supremely restored. Mulberry Restaurant has a growing reputation for good food.					
54 Woolley Grange Hotel Woolley Green, Bradford-on-Avon N. Chapman P: 01225 864 705 F: 01225 864 059 E: info@woolleygrange.com	distance from centres (mi)		phone in room	restaurant on premises	takes credit cards
	training	town			
	6.1	Bradford-on-A approx. 1 mile			
from £90.00 single, all rooms en-suite, includes breakfast Jacobean manor house in open countryside with outdoor heated pool, croquet, tennis, badminton.					
50 Travel Inn, Cepen Park (#267) W. Chippenham Way (M4, J17) Chippenham SN14 6UZ P: 01249 462 096 F: 01249 461 387 W: www.travelinn.co.uk	distance from centres (mi)		phone in room	restaurant on premises	takes credit cards
	training	town			
	8.6	Chippenham approx. 1 mile			
£41.95 single, all rooms en-suite + £6.00 full/£4.00 continental bkfst 79 rooms with Brewsters and petrol station nearby.					
56 Travelodge (#813) Moto Services, A36 / A350 Bypass Warminster BA12 7RU P: 01985 219 539 F: 01985 214 380 W: www.travelodge.co.uk	distance from centres (mi)		phone in room	restaurant on premises	takes credit cards
	training	town			
	12.1	Warminster approx. 1 mile			
£49.95 single, all rooms en-suite + breakfast from £4.45 31 rooms with Burger King, Little Chef and petrol station nearby.					

Setting up an EPDQ

continued from previous page

access is controlled using basic authentication. If you can set up your own .htaccess files, then you probably have this facility.

Contacting the System

You write your code to contact the bank's system with an order number and an amount of money, and your user goes off to them to make the payment. Once they have completed their payment data entry, the bank's server contacts your server with a response (accepted or declined), which you'll save via a second piece of code which you provide. A third piece of code which you provide is then contacted by your user following a link away from the bank's site and back to yours. This piece of code needs to match the returning user to the authorisation you should have just received, and you can proceed "post payment".

Each of the steps along the way has appropriate security encoding and passwords applied to keep it secure, but you'll be relieved to note that you don't need to support secure (https) protocol or authentication on your own server.

Important Considerations

Before you can accept online payments, you'll need to consider wider business issues too. As a starter, you'll need to have an account set up by the bank and reach appropriate agreements with them. Then you'll need to gain their trust that you know what you're doing in writing your own code, or work with one of their authorised interfacers, perhaps using standard software. If payments are automated, you'll need to ensure that you can filter orders before payment is accepted to make sure that the order is genuine, fulfillable from stock (or from short term delivery), and comes from a location to where you're prepared to ship.

SERVICES

- Banks
- 34 Barclays
- 26 HSBC
- 17 Lloyds TSB
- 27 NatWest
- Emergency**
- 2 Hospital
- General**
- 20 Post Office
- 33 Public Library
- 22 Tourist information

LEISURE

- 21 The Blue Pool
- 3 Christie Miller Sports Centre

FOOD & DRINK

Groceries

- 44 Aldi
- 24 Iceland
- 51 Made in America
- 37 Sainsbury
- 38 Somerfield

Takeaways

- 19 Acropolis
- 16 BaguetteBar.com
- 8 Beijing Chinese Takeaway
- 3 Chop Suey House
- 13 Cod Joe's Caribbean Takeaway
- 11 Cornerstone Coffee Shop
- 28 Country Cottage Pasties
- 39 The Fisheries
- 9 Golden Falls Chinese
- 59 Happy Valley Chinese & Fish Bar
- 31 Kebab House
- 43 Lee's Fish and Chips
- 29 Melksham Chinese
- 18 The Three Cooks Bakery
- 58 Tricia's Plaice

Restaurants (most do takeaways)

- 15 Big J's Farmer's Cafe
- 51 Brewsters
- 46 cafe inside Leeke's
- 43 Lee's Chinese Restaurant
- 47 Little Chef
- 45 McDonald's (with drive through)
- 23 Melksham Tandoori
- 41 Mulberry Restaurant
- 10 Refa Balti House
- 6 West End Inn

Sampling of Pubs

- 35 The Bear
- 32 Bud's
- 53 Foresters Arms
- 30 The Grapes
- 55 The New Inn
- 25 The Parson's Nose
- 12 The Tavern

HOTELS/B&Bs

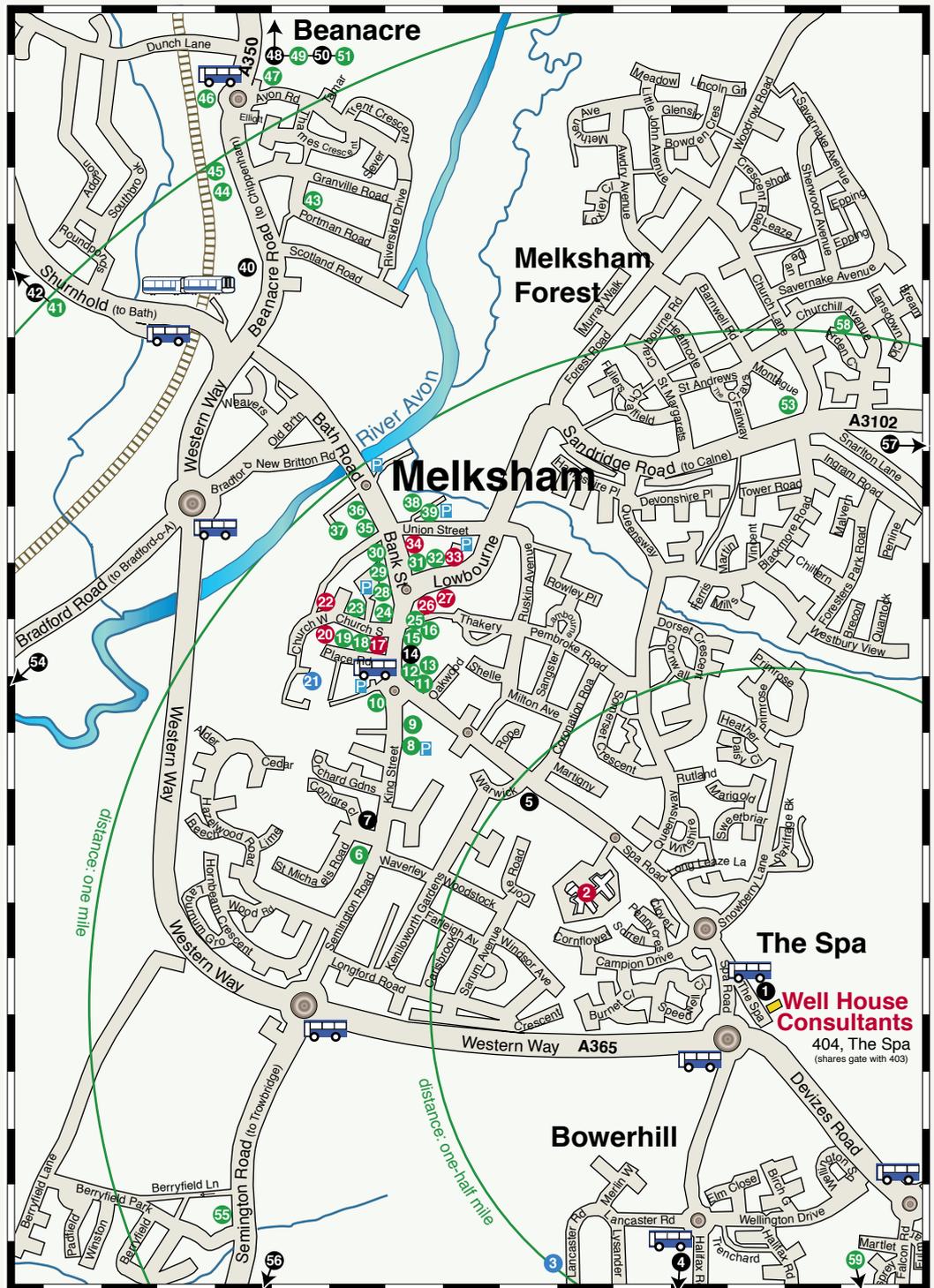
- 48 Beechfield House
- 7 Conigre Farm
- 4 Duxford Close B&B
- 14 The Kings Arms Hotel
- 40 Longhope Guest House
- 5 The Old Manor
- 57 Sandridge Park
- 42 The Shaw Country Hotel
- 52 The Sign of the Angel
- 1 The Spa B&B
- 50 Travel Inn
- 56 Travelodge
- 54 Woolley Grange Hotel



From the city or town to Well House Consultants' door step as the crow flies* ... (in order of distance away)

Bath	10	Leicester	120
Swindon	20	Cambridge	140
Salisbury	20	Sheffield	160
Bristol	25	Manchester	165
Southampton	45	Liverpool	165
Cardiff	50	Dover	175
Oxford	55	Leeds	185
Reading	60	Norwich	200
Portsmouth	70	York	205
Swansea	90	Newcastle	275
Birmingham	90	Glasgow	355
London	105	Edinburgh	355
Plymouth	115		

* These are the mileages we use to determine our expenses for on-site courses.



WEB FORENSICS



ANALYSING LOGS FOR SEARCH TYPES

It's amazing what you can learn about your own business from the way people use your web site. Web servers routinely

log all requests for pages. Since ours also tells us the previous page people visited, we can see what they searched for when

they entered keywords in search engines, such as Google.

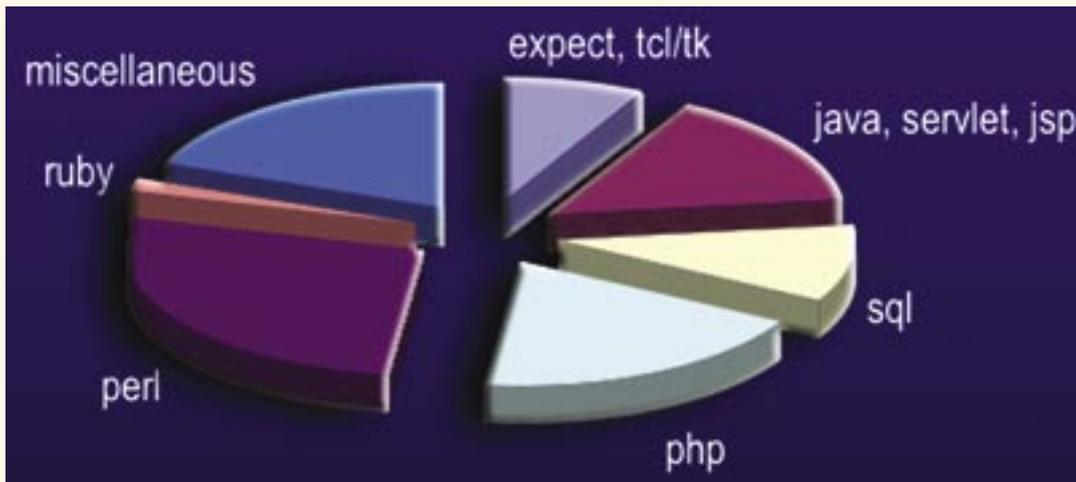
We get thousands of hits every week from the search engines and shown (left) is what people were looking for during a typical week.

Another interesting search is for the phrase "regular expressions". We've not separated it out as it's often a search term used with "Perl", "PHP", or "Java", but if we had done so, it would have had a larger slice than MySQL.

Aside from Open Source languages, what were some of the other topics people searched for (miscellaneous)? Ten people were looking for information about flights from Bristol International Airport and six searches were for "Melksham".

Also comprised in the miscellaneous area were other computer-related topics, such as the 92 searches for Linux, 62 for Windows, seven for Mac OS X, five for Unix and two each for Solaris and AIX.

If you haven't analysed your own web site traffic, you're missing an excellent marketing tool.



During one typical week, our web site logged 3,877 searches by users for various topics.

PROVIDING ANSWERS TO SEARCHING QUESTIONS

The logs of our web site tell us what keywords people have entered when they come to visit us via a search engine. Let's see if we can provide answers to what appears to be questions they might be asking ...

Q: mysql delete table

A: A table in MySQL can be deleted using the command:

```
DROP TABLE [tablename]
```

but you're only allowed to do so if you have drop privileges.

Q: read file PHP

A: You can read a whole file into an array using the file function, or if you want to read a file line by line you can use fopen followed by a loop of fgets or fgetsv functions.

Q: perl hashes

A: A hash in Perl is a collection of scalars (variables), each of which has a unique key which is itself a scalar.

Q: add linux xp boot.ini

A: The answer needs more than the space available here – see

an example of how we set up a machine to dual boot in our solutions centre at:

www.wellho.net/solutions

following the link

"Can Windows XP and Linux co-exist on the same system?"

Q: tutor examples

All the examples from our training courses can be viewed from our website via:

www.wellho.net/resources

Type the name of the example that you want into the form (you'll find the example name in the notes you were given on the course), and the example will be displayed.

Q: tk introduction

A: Tk "Tool Kit" is a GUI (graphic user interface) that sits on top of Tcl or Perl, and allows the programmer to write applications which can be used through buttons, sliders and other widgets from within a window on a Unix, Linux, Windows or OS X computer.

Q: perl test if file exists

A: print "\$file exists\n" if (-e \$file);

Q: wiltshire spiders

A: Several possibilities; of them:



The Bug & Spider

FA Spider

RUNNING A ROBOT APPLICATION

Although the intention of most web sites is to provide information for human visitors using Internet Explorer, Netscape, or one of a number of other browsers, a good proportion of web hits these days are generated by automaton on other computers running "robot" applications.

Search engines use robots to crawl a site, visiting each page and indexing the content. Linkcheckers will systematically visit every page in a domain,

alerting the web site owner to pages that are missing. Current exchange rates, news feeds and dynamic data can be harvested from other web sites to include in yours. If you want to monitor a site so that you know when it changes, a robot is going to be much more efficient than you visiting the site in person.

The LWP::UserAgent module in Perl is great for obtaining data from remote URLs (see our own sample script on the facing page).

Sample Script

Our sample script visits a domain named on the command line, and follows links to the first 100 pages within

the site looking for changes. It's ideal for running as a regular ("cron") job if you need to be alerted when a site is

changed, and is easily modified to keep on running until it exhausts a site to check for the integrity of links.

```
# Check if the top of a site
has changed
```

```
# Domain name to traverse
```

```
# Number of distinct pages
to traverse
```

```
# Locations not to watch
```

```
# Check for the "robots.txt"
file
```

```
# Proceed further if there
is no robots file, or if it is
unchanged
```

```
# If robots.txt is altered,
inform the process owner
```

```
# robots.txt is unchanged
since last time, or does
not exist, so proceed into
the site (script assumes
that process owner has
manually checked robots
exclusion standard)
```

```
$site = $ARGV[0] || "www.wellho.net";
```

```
$limit = 100;
```

```
@omit = ("cgi");
```

```
use LWP::UserAgent;
```

```
mkdir $site if (! -e $site) ;
```

```
$agent = LWP::UserAgent->new;
```

```
$agent->agent("Well House Consultants");
```

```
@alt = ("Not Available", "Changed", "Unchanged", "Freshly Collected");
```

```
getpage("http://$site/robots.txt");
```

```
die "robots.txt was $alt[$altered]\n" if ($altered%2);
```

```
@pagequeue = ("http://$site/index.html");
```

```
while ($file = shift @pagequeue) {
```

```
    next if ($already{$file});
```

```
    last if (++$npages > $limit);
```

```
    sleep 5 unless ($npages%10);
```

```
    $already{$file} = $stat = getpage($file);
```

```
    if ($altered == 1 or $altered == 3) {
```

```
        print "$file: $stat, page was $alt[$altered]\n";
```

```
        $alert++; } }
```

```
    foreach (@links) {
```

```
        $fully = URI ->new_abs($_, $file) ->canonical;
```

```
        $fully =~ s/#.*//;
```

```
        $shopper = 1;
```

```
        foreach $leave (@omit) {
```

```
            $shopper = 0 if ($fully =~ /$leave/)
```

```
            if ($fully =~ m!^http://$site/!i and $shopper) {
```

```
                push @pagequeue, $fully; } } }
```

```
$alert and print "alter count is $alert\n";
```

```
sub getpage {
```

```
    $req = HTTP::Request->new(GET => ($pagewant=$_[0]));
```

```
    $res = $agent->request($req);
```

```
    $altered = 0; @links = ();
```

```
    if (($status = $res->code()) == 200) {
```

```
        ($mirrorname = $pagewant) =~ tr!/:!%=:;
```

```
        $page = $res -> content();
```

```
        $altered = 3;
```

```
        if (-e "$site/$mirrorname") {
```

```
            open (FH, "$site/$mirrorname");
```

```
            read (FH, $oldpage, -s "$site/$mirrorname");
```

```
            $altered = ($page eq $oldpage) + 1; }
```

```
        open (FH, ">$site/$mirrorname");
```

```
        print FH $page;
```

```
        @links = ($page =~ m/href\s*=\s*"?([\^">\s]+)/ig); }
```

```
    return ($status); }
```

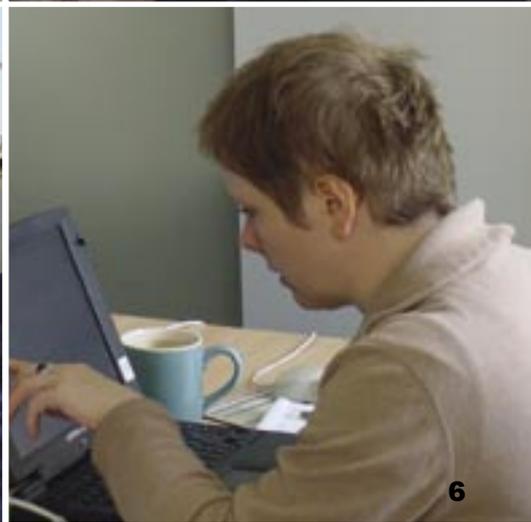
Only visit each page once

Limit site traversal;

Don't swamp the server!

OBTAINING DATA FROM REMOTE URLS

Before you run our code (or write your own spider), beware! It's very easy to use LWP::UserAgent to write code that's unkind to the web servers, and code that isn't welcome. It's also very easy to harvest information and redistribute it contrary to copyright. Always check the robots.txt file to see if you're welcome at a site (our code above stops at robots.txt if that file has been changed), and always put in some sort of delay between every few pages you harvest so that you don't deny use of the server to others or consume the entire bandwidth of the server.



12

Faces and places

A peek around the training centre at Well House Consultants

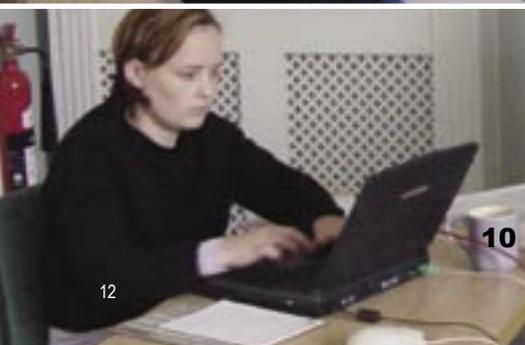
“My mother has a wall of friends and family photos she calls ‘The Brag Board’. I like to think our students are more than just clients; they’re our friends. And we like to brag about them, too.”

– Lisa Ellis [ADMINISTRATOR]

- 1. Adam Sansom from GlaxoSmithKline concentrates on coding.
- 2. Christopher Sheraton from Ford considers whether to use a list or a hash during a practical.
- 3. Paul Clementson of Joy Mining Machinery during a MySQL course.
- 4. Mark King experiments with a range of conditional statements during a Perl

- practical.
- 5. If you arrive by train at Melksham station, let us know and we’ll collect you in our Taxi.
- 6. Caroline Ellam from Three checks the output from her program.
- 7. Dave Williams pours over our first issue of “Of Course”.
- 8. Clifford Weatherall checks his

- MySQL coding.
- 9. Dan Butler concentrates on a MySQL exercise.
- 10. Joanna Norrington-Davies from Eduserv concentrates on an exercise.
- 11. A wide array of food creates lunch decisions for James Walker.
- 12. Suzanne Ruggiero catches up on today’s headlines.



Open Source Software: latest versions available for download

Table and comments were compiled in January, 2003

The numbering systems used by different groups for their software releases are many, varied, and often baffling. We've updated our table format since our last issue to make it easier to follow and more consistent across languages. Where a single version is shown, it's the current production release. Where a group is offering a testing version for general download that's in advance of the production release, we list the production release and also the test release. We are no longer listing future versions that you cannot yet generally download.

Subject	Stable Version	Other Version	Web Site
Perl	5.8.0		www.perl.com
Tcl/Tk	8.4.1.0		www.tcl.tk
Expect	5.38.0		expect.nist.gov
PHP	4.3.0		www.php.net
Ruby	1.6.8		www.ruby-lang.org
MySQL	3.23.54a (production)	4.0.8 (gamma)	www.mysql.com
Java, standard	1.4.1_01 (standard)		java.sun.com
Java, Enterprise	1.3.1 FCS (production)	Enterprise 1.4 (beta)	java.sun.com
C (Gnu)	3.2.1		gcc.gnu.org
Python	2.2.2 (production)	2.3a1 (alpha)	www.python.org

in the table above/right from sourceforge, or from ActiveState who have a ready compiled and bundled version, including a dozen extras from expect through [incr tcl] to TclSOAP and Snack.

The developer's forum for Tcl is beginning to make references to release 9, which will be the next major release of Tcl. It's said that there will be relatively few new features; rather, the structure and C API will be cleaned up to make it a language for the future. No release date has been set, although the web site has an enigmatic "2003?" on it.

PHP 4.2.0, 4.2.1 or 4.2.2. If you're currently running version 4.1.2 or prior, note that upgrading to 4.3.0 may cause your scripts to break; from release 4.2.0 variables are no longer automatically populated from form, cookie and environment contents. You can maintain compatibility through changing your php.ini file if you wish.

4.3.0 incorporates many internal changes. For the regular user, the "keynote" features include speed and command line interface improvements, a bundled GD library for graphics, and a unified approach to stream handling.

testing – 4.0.8 is a gamma release only.

Some new MySQL books (such as the MySQL Cookbook from O'Reilly, ISBN 0-596-00145-2) are already covering the new release and it's anticipated we would expect a production release 4.0.x in early 2003.

COMMENTS ON

Perl

If you're downloading Perl to install afresh, go for 5.8.0; release 5.005_03 is still available for timid souls. If you're running Perl 5.005, 5.6.0 or 5.6.1 and don't need the extended character handling or threading of 5.8.0, you'll probably gain little from upgrading at present.

Perl 6 remains some way in the future, but changes that it will incorporate have become much clearer since our last newsletter - see the separate article.

COMMENTS ON

Tcl/Tk

Release 8.4.1 is now stable; you can download it via the URL

COMMENTS ON

PHP

Version 4.3.0 was released late last year, and it's a good idea to upgrade if you're running version

COMMENTS ON

MySQL

The production release is still in the "3.23" series as the MySQL folks are very careful with their

COMMENTS ON

Gcc

The Gnu Compiler Collection supports Fortran, Java and ADA compilation as well as Kernighan and Richie C, ANSI C, Objective C and C++.

If you're looking to install a first C compiler on your system, remember that you'll need to download a pre-build binary distribution rather than the source code release.

Looking for other Open Source software?

Did you realise that more than 50,000 projects are hosted at SourceForge.net? Here are ten that are likely to be relevant to our readers:

phpMyAdmin

A tool written in PHP intended to handle the administration of MySQL over the WWW.

Alvaro's Messenger

Written in Tcl, this is an MSN messenger client for linux/unix/windows/mac.

SquirrelMail

A PHP4-based Web email client. It includes built-in pure PHP support for IMAP and SMTP.

Fink

Written in Perl, an attempt to bring the full world of Unix Open Source software to Darwin and Mac OS X.

Tiki

A PHP-based CMS system based on a system where the users can edit the pages as they navigate.

POPFile

A Perl-based email classification system. Sort incoming mail, or simply use it to eliminate spam.

PHP Nuke

An "Advanced Content Management System" which we're starting to hear a lot about.

Expat XML parser

James Clark's library in C; required if you're going to use the XML::Expat module in Perl.

phpdev

A bundle of PHP/Apache/MySQL/PERL/phpMyAdmin/PHP-GTK preconfigured to run on Windows.

Webmin

A Java and Perl web-based system administration tool for Unix servers and services.



Apple get it right to the core

Imagine an operating system that's as robust and powerful as Unix, with the ease of use and excellent peripheral support one normally associates with the Mac or Windows world. It's here - it's Mac's OS X.

Mac's OS "Ten" is built on the groundrock of BSD Unix, with their Aqua Graphic User Interface (GUI) running on top of it - and they've done an excellent job. Many applications have been ported, and those which haven't or are still under way can be run through Classic Mode and an OS9 folder.

We've noticed a move amongst the big names towards Mac Titanium laptops for their "travelling system", and where these trendsetters have the confidence to step, others will follow. It's a powerful package, everything you would expect from Unix (Perl, Apache, PHP, Java are our favourites) built in, server as well as client. Those who have used Microsoft Office v.X, and Explorer on both a PC and a Mac are insistent the applications

run better on the Mac. The traditional Mac strengths haven't gone away; with the OS comes Sherlock, iMovie, iPhoto, iTunes and more, and you can install Illustrator, Photoshop, InDesign, FileMaker, Dreamweaver and so on. (Applications such as QuarkXpress and FrameMaker will run on the Classic, but are not yet OS X-capable.) AOL software allows you to dial in worldwide with ease, and the Unix base turns even a laptop into a thumping good server at home or on the road.

Hardware integration within the operating system is excellent too, even with third-party products. There's no hassle with drivers as you print, run a wireless, burn your own CDs/DVDs, dial in from hotel rooms, watch your favourite dvd, download photos directly from your USB digital camera ... well, there's just no hassle.

OS X comes as user-friendly or as geekish as you need. Any level of administrator access is allowed. In other words, if you know how to do it, it can be done.

CODING TIPS

Consider the use level of code before you write it

Give some thought to writing a piece of code. Your coding techniques might be very different from a single-use utility to produce a unique report to code that's going to be run on a public web site accessed by many different users.

In order to qualify these considerations, think of three use levels.

Level 1 - Code that is required to work on a single instance of a data set. It may not be a correct solution for all possible data sets, but it works for the specific need. For example, you're wanting to extract 800 addresses from HTML web pages, with names, postcodes, counties, etc. entered by different people in different ways. This would be Level 1 code, and consider that you may have to manually adjust a few results (such as a confusion between the county of Isle of Skye and S [Sheffield] and SO [Southampton] postcodes, especially in cases where the original data was SO - "S-zero").

Level 2 - Code designed for use by its author on multiple data sets. It may not be a correct solution for everyone, nor according to the full specification of the input, but it does work for all the data that's thrown at it. Special cases (such

as someone with the surname "Wiltshire" specified without a "Mr" or a first initial in front of it) could cause problems.

Level 3 - Robust code that is written according to the specification of the input. My "Wiltshire" example would work, and a file that someone has called "html.html" wouldn't cause problems to a web page handling program written to Level 3.

It's always best to write your code at Level 3 if you practically can; other programmers have usually designed their data formats for reasonably easy analysis. More often than not, coding to this level is not a big problem. On a few occasions, though, Level 3 coding can require some considerable investment in writing and isn't necessary, even if the computer purists will point you in its direction. You should aim to achieve this level if you're using it on a public-facing web server; otherwise, there's a chance of your users finding a security hole they can enter.

Level 3 code would be far too complicated, perhaps impossible, if used on the above address example, or when matching a spoken language, or if you're doing character or voice recognition, or translations.

Mouse mats for sale

Can you remember your \d's from your [[:graph:]]'s? If not, you might like to purchase a Regular Expression mousemat - just £4.95 (inclusive of VAT and shipping to a UK address).

<http://www.wellho.net/regex/mousemat.php4>



Setting up Mac OS X to look at the local /etc/hosts file

Using your first Mac on your intranet?

Looking to get name resolution working? It'll work fine if you have NIS, NISplus or DNS set up and configured on your network. But you'll

find that access to the local /etc/hosts file is disabled by default.

To turn it on, use the "sudo" command to give you administrator access, then tell the lookup daemon.

```
niutil -create . /locations/lookupd/hosts
```

```
niutil -createprop . /locations/lookupd/hosts LookupOrder CacheAgent FFAgent NIAgent DNSAgent DSAgent
```

The following public courses are scheduled at our Melksham, Wiltshire training centre.

Dates are through October 2003; see <http://www.wellho.net> for further dates. If you have four or more trainees, we'll run a private course for you at our training centre or at your office, and tailored to meet individual company needs.

PERL

Perl is a programming language used in data manipulation, web interaction and systems administration. Its wide range of facilities make it ideal for large and small jobs, and it can act as an excellent glueware between otherwise unconnectable applications.

Learning to Program in Perl

A foundation course for those who haven't programmed before. Covers the fundamentals of programming in Perl, and the major features of the language that are required in most applications.

☒ 5-day course, runs:

Mon 17 Feb to Fri 21 Feb 2003

Mon 19 May to Fri 23 May 2003

Mon 18 Aug to Fri 22 Aug 2003

Perl Programming

For those with programming experience. Covers the fundamentals of Perl by comparing the language with other languages, then proceeds on to the major features of the language that are used in most applications of Perl.

☒ 4-day course, runs:

Mon 31 Mar to Thu 3 Apr 2003

Mon 30 Jun to Thu 3 Jul 2003

Tue 7 Oct to Fri 10 Oct 2003

Using Perl on the Web

Using Perl on a web server through CGI, and other web-based interfaces such as automated FTP, clients, servers, modPerl, etc. An advanced course that assumes fundamental Perl knowledge.

☒ 2-day course, runs:

Mon 17 Mar and Tue 18 Mar 2003

Mon 7 Jul and Tue 8 Jul 2003

Mon 20 Oct and Tue 21 Oct 2003

Perl for Larger Projects

This advanced course covers topics that you'll need to know if you're using Perl for a large project. Content includes variable scoping, namespaces, object orientation, POD, XML, SQL database access, and more.

Scheduled Courses

Prices

1-day courses	£ 290.00	470.00 euros
2-day courses	£ 490.00	800.00 euros
3-day courses	£ 690.00	1,130.00 euros
4-day courses	£ 890.00	1,460.00 euros
5-day courses	£ 1,090.00	1,790.00 euros

For multiple students from the same company booked on the same course, simply calculate the total cost using the formula:

£200.00 for each student per day plus £90.00

(e.g. 2 students / 4 days = £1,690.00)

Prices are valid for all courses offered in this schedule and on our web site <http://www.wellho.net> for bookings confirmed by 31st August 2003.

☒ 3-day course, runs:

Wed 19 Mar to Fri 21 Mar 2003

Wed 9 Jul to Fri 11 Jul 2003

Wed 22 Oct to Fri 24 Oct 2003

Perl Extra

Perl Extra is only available to individual trainees who have attended one of our public Perl courses, and require further advanced training in a combination not offered by our other advanced training.

☒ Dates to be mutually convenient to trainee and trainer.

JAVA 2

Java 2 is a wide-ranging, portable, network-aware language that you may wish to run on servers, on clients, or on stand-alone hardware.

Java Programming for the Web

This course covers the world of Java, and the programmer interfaces used on Java web-based applications.

☒ 5-day course, runs:

Mon 28 Apr to Fri 2 May 2003

PHP

PHP is a language that you embed in web pages to run on your server. You should already know something of the web, HTML, SQL and programming. We offer two courses: a technology day that covers enough of those prerequisites to get you started, and a PHP programming course that covers both the language and its web use.

Technology for PHP

The structure of the web, web pages through HTML and relational database access through SQL covering the prerequisites needed for our PHP programming course.

☒ 1-day course, runs:

Mon 3 Mar 2003

Tue 27 May 2003

Tue 26 Aug 2003

PHP Programming

This course covers programming in the PHP language, and also state, cookies, interfacing to SQL and XML handling, and emailing from the server that you'll need to write effective server side code.

☒ 3-day course, runs:

Tue 4 Mar to Thu 6 Mar 2003

Wed 28 May to Fri 30 May 2003

Wed 27 Aug to Fri 29 Aug 2003

MYSQL

We study the SQL language and how it's used from PHP, Perl and Java, and administration including user account and security controls, and database design so that you'll be putting together well thought out, maintainable, expandable databases.

The MySQL Relational Database

Our MySQL course covers the structure of a relational database, the SQL language, database design, installation and administration of MySQL, and its use from open source languages.

☒ 2-day course, runs:

Mon 10 Mar and Tue 11 Mar 2003

Mon 16 Jun and Tue 17 Jun 2003

Mon 8 Sep and Tue 9 Sep 2003

TCL, TCL/TK and EXPECT

The Tcl language has a wide range of extensions, such as Expect and Tk, making it effective for applications such as writing Graphic User Interfaces and automating repetitive operator tasks.

Tcl and Expect Basics

This course teaches you Tcl, from fundamentals through most of the major features, and also the Expect and [Incr-Tcl] extensions, but not the Tk extension.

☒ 3-day course, runs:

Mon 9 Jun to Wed 11 Jun 2003

Mon 13 Oct to Wed 15 Oct 2003

Tcl/Tk and Expectk

This course covers the Tk extension to Tcl which provides the Graphic User Interface. Learn how to build widgets, how to pack them into frames, and how to use them to provide interactive, GUI-based programs. A knowledge of Tcl is a prerequisite of this course.

☒ 2-day course, runs:

Thu 12 Jun and Fri 13 Jun 2003

Thu 16 Oct and Fri 17 Oct 2003

PUBLIC SECTOR

The National Health Service, Local Government and other public sector bodies are working hard to meet government requirements to have their data available online in the next few years.

Getting Your PS Data Online

This one-day course takes you behind the scenes of web-based applications and overviews the various technologies for you, setting you in the right direction for your online provision.

☒ 1-day course, runs:

Wed 12 Mar 2003

Thu 8 May 2003

REGULAR EXPRESSIONS

Regular expressions are probably the most powerful pattern matching tools that is available. In just one or two short lines of code, the knowledgeable programmer can write code that would be 10 times as long using any other technique.

Regular Expressions

This course refreshes any knowledge of Regular Expressions that you have from first principles, and shows you how to make effective use of them for recognising, capturing and altering patterns.

☒ 1-day course, runs:

Thur 13 Mar 2003

Fri 9 May 2003

Still some growing pains, but expect a very mature product

Do you have teenage children? If you do, in more tense moments you may wryly think to yourself "if only I had done x when they were young".

Perl is a teenager too. Written initially by Larry Wall, it hit the world in 1988 when the Internet was a dream, before the @ character had a special significance in email addresses and before Object Oriented languages used the dot to say "run this method on this object". If only Larry had known how things would develop, he would have done things somewhat differently.

More recent languages (such as PHP, Java and Ruby) have taken account of developments, but no doubt they'll have their teenage problems in due course if they survive as long.

How then can Perl be a language for the future? By making more brave decisions that again turn out to be farsighted. The next release of Perl – Perl 6 – will be a rewrite that will not be source-code compatible with Perl 5. Even the "." operator that we use on nearly every line changes. But there will be good conversion tools and Perl 6 should be a language fit for the third millennium.

In London one evening a couple of months ago, Damien Conway, one of the group who's been working on the core of the new Perl, spoke to about 150 to 200 geeks, gathered in a Methodist Church Hall. Well House Consultants went along in some trepidation. We're delighted with Perl anyway, which is why we teach it, and to some extent we take the view "if it ain't bust, don't fix it". The geeks were perhaps wondering the same.

Time and again, Damien showed how some little niggle has been swept away and replaced by a piece of sheer elegance that glows "wow" all over, not only clearing the niggle but opening beautiful expanses of new possibilities implemented neatly, cleanly, and consistently.

Larry's done it again, it appears.

When will Perl 6 be with us? An informed source suggested that it's still some way off and that we would be very lucky if we see anything beyond testing code within the next year. Don't stop programming in Perl 5 just yet, and expect Perl 5.10 before Perl 6. But do stick with Perl – it's got a great future and will be with us for the foreseeable future.

Perl 6 Nugget

The "Intelligent comparison" operator

In Perl 6, the =~ operator becomes the intelligent match operator; there's no longer any need to use == to compare numbers and eq to compare strings. =~ will do the lot and many more things. Depending on the operator type, it will have 24 different tests it can perform.

Perl 6 Nugget

Vectorized operators

Put a ^ in front of an operator, and it's performed on each element of a list.

```
@average = (@in1 ^+ @in2) ^/ 2;
```

will create a list of averages from two incoming lists.

Perl 6 Nugget

Topicalisation

Establish the current topic, then all your operators are performed on it unless otherwise stated. Might sound a bit like \$_ (it is), but you ain't seen nothing' yet.

```
given $value {
  when < 10      { print "much too small"; }
  when m:!/<digit>/ { print "Contains non-digit"; }
  print "Might be possible";
  when < 20     {print "small but OK"; }
  when < 60    {print "not large"; }
  default      {print "magic!"; }
}
```

A report and further technical details of Perl 6 can be found at: <http://www.wellho.net/solutions/16533668477.html>.

Your contacts at Well House Consultants:

Administration:	Lisa Ellis	lisa@wellho.net	
Technical:	Graham Ellis	graham@wellho.net	
Phone:	01225 708 225	+44 1225 708	225
(International)			
Fax:	01225 707 126	+44 1225 707	126
(International)			
Web Site:	http://www.wellho.net		
Postal Address:	404, The Spa, Melksham, Wiltshire, UK SN12 6QL		

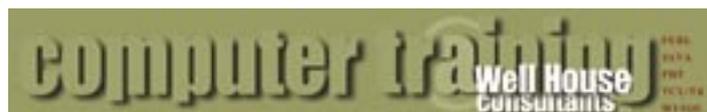
Our office hours are 08:00 to 18:00 Monday to Friday.

You'll usually reach us at weekend and in the evenings too, as we often divert the phone when we're not around. If you email us, we'll receive your email within 24 hours, 365 days a year.

This newsletter is published by Well House Consultants Ltd. and is distributed free. Please contact us if you're reading someone else's copy and would like to be added to the circulation list, or if you would like to be removed from our list.

All material in this newsletter is copyright Well House Consultants, Ltd., and may not be reproduced in any form or by any means electronic without prior written permission of Well House Consultants except in the case of brief quotations embodied in critical articles and reviews.

Well House Consultants has attempted throughout this publication to distinguish proprietary trademarks from descriptive terms by following the capitalisation style used by the manufacturer.



You'll see a new look to our training notes, our manuals, our stationary and our web site, reflecting the progress of Well House Consultants as a provider of high quality, specialist training in Open Source languages.

We introduced our first Perl course in 1996, when the main interest was in server-side programming of Web applications by innovators. That still remains an important part of our business, but Perl has grown much more rapidly in other uses. Tcl/Tk and MySQL also have substantial use away from Internet site applications, as does Java. Only PHP remains predominantly Internet-based, but even it finds uses away from web

sites these days.

"Computer Training" reads our new logo. Exactly. We're training specialists. The trainer who stands in front of you takes classes regularly, and knows how to present, and get across concepts from hashes to Object Orientation, and from database design principles to functions. He'll have played a major role in actually writing the course, and he'll have ongoing personal experience of the technologies he's teaching.

Have you ever stayed at a hotel where the bed is turned down and a chocolate placed on the pillow? "Wow", you say. We strive to be a "chocolate on the pillow" company and give you that wow factor.