# search insider

In this month's Search Insider, David Green decides to test the claims of Bright Station's WebTop searching tool and its related WebCheck search utility.

In 1999 Dialog Corporation sold its information services division to the Thomson Corporation for considerably less than it paid to acquire the Dialog and DataStar services from Knight Ridder. Free from its burdening debt, the streamlined company renamed itself Bright Station with three main areas of focus: ecommerce, internet ventures and web solutions. It is from this last division that the company's WebTop search engine (www.webtop.com), and related WebCheck search utility have sprung.

Top of the web

After three months of intensive development by just three people, the WebTop search engine was launched with an index of around 50 million documents. Today its index has grown to over half a billion documents which, according to Search Engine Watch, makes it the world's second largest search engine (trailing behind Google). It currently indexing 10 million pages a day. This
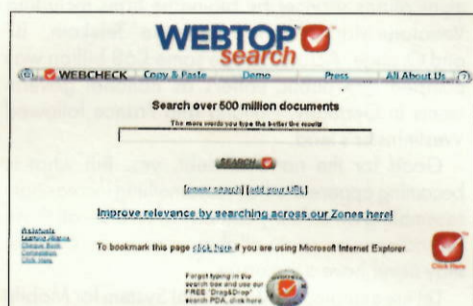
explosive growth has been partially enabled by Bright Station snapping up second-hand servers when Boo.com went bust.

Whereas Northern Light categorises results into folders, WebTop categorises its index in a small number of 'zones'. Most of the web sites in its index are categorised under 'WebZone', but there are a number of other zones, including:
- NewsZone which covers over 1500 news sources (provided through a licensing agreement with Moreover.com).
- CompanyZone which covers most public companies listed on the world's stock exchanges. The web sites of over 15,000 public companies are indexed and updated every day.
- WAPZone which covers over 500 web sites that focus on WAP developments and protocols. (Bright Station is a member of the WAP Forum, which thrashes out such protocols.)

WebTop is a natural language search (NLS) engine. Like other NLS engines, it seeks to identify the key concepts contained in the user's query to determine its search statement. It is based on a technology that Bright Station calls 'ActivQ'. This is essentially a distillation and hybridisation of two Dialog technologies, Muscat and InfoSort.

A direct competitor to Autonomy and Verity, Muscat is based on the concept of 'user determined relevance'. It incorporates intelligent agents, which attach weightings to words to define a concept-based search. In a nutshell, InfoSort indexes and categorises documents by examining the semantic and syntactical relationships in that document and using a 'rule base' to decide how it should be categorised.

Concept to content

Combined, these two technologies deliver the ActivQ 'concept to content' methodology that drives the WebTop search engine. WebTop started life by assimilating the old EuroFerret search engine. Consequently users can search using several Roman alphabet European languages. Indeed, in June the company signed an agreement with Netscape for WebTop to be the preferred search engine for Netscape's Netherlands portal.

Quite why Bright Station boasts that you can 'type and search' or 'copy and paste' into the search box is beyond me. You can do this on any nat-ural language search engine (and if you were on a different type of search engine you'd be typing in a structured search statement anyway). Still, I put this to the test to determine whether WebTop perhaps delivered more accurate results.

Put to the test

In the test, WebTop went head to head with AltaVista. Copying the sentence 'WebTop is a natural language search (NLS) engine' into the search boxes for each service produced surprisingly different results.

WebTop failed to pick up its own URL in its first results, but did list a page from Bright Station's web site. AltaVista produced some excellent results on both WebTop and natural language search engines. WebTop's first page of news headlines also picked up plenty of irrelevant headlines about languages, whereas AltaVista's news headlines were exclusively focused on

search engines – demonstrating a better understand of the key concept of my sentence. However, WebTop's Company Zone was better than AltaVista as it correctly listed the web site for Bright Station. AltaVista's Business folder failed to list anything other than other categories in its index.

WebCheck

WebCheck is the (free) nifty little search utility companion to the WebTop search engine. The beta product was previewed at last year's Online Information exhibition, with WebCheck formally launched in March this year.

Its key feature is a 'drag and drop' function that allows users to highlight text and drag it into the WebCheck icon (a silver sphere which floats around your screen). WebCheck then interrogates the WebTop index.

Combined with the natural language searching ability of the WebTop search engine, this 'drag and drop' function is useful in that it saves users from distilling complex concepts into a structured Boolean search statement.

Still, other search utilities such as Bullseye, Mata Hari and Copernic offer greater sophistication in their functionality: view results by site type, relevance or in 'concept clusters', document summaries, save searches, highlighting search terms on page and so on.

Reality check

When the company unveiled WebCheck, it wasn't just launching a product. No, it was launching 'a revolution'. So there.
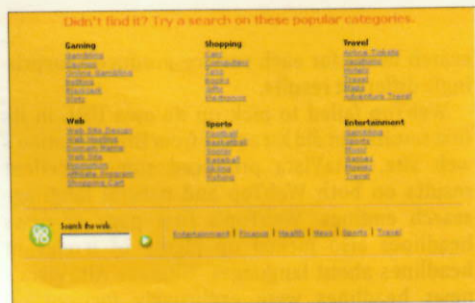
Seeking to differentiate WebCheck, Bright Station refers to it, not as a search utility, but as a 'Personal Desktop Assistant (PDA)'.

Privacy, people and profiling

One word of warning. Each copy of WebCheck has its own unique ID. Every time you use WebCheck, it builds a behavioural profile of your interests which Bright Station then offers to advertisers seeking to target particular types of people, or people interested in particular subjects. Bright Station swear they will never match this data with the user information required to download WebCheck. Hope so. Come to think of it, the silver WebCheck icon does look like that spherical camera of Channel 4's 'Big Brother' fame, peering back at me through the screen. What a thought.

David Green is a council member of the Institute of Information Scientists. He writes here in a personal capacity.
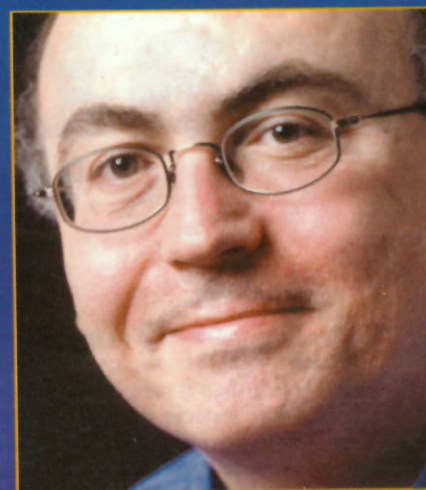
Flood on ... 3G

So you know all that fuss about the third generation mobile auctions – when the telcos and mobile firms coughed up billions to win licences from European governments for the right to get in on the ground floor of the future?

It may already be unravelling. And you, the consumer, will be expected to come to the rescue.

Some £22 billion was lashed out on that part of the UK radiowave spectrum allocated for so-called third-generation UMTS (universal mobile telecommunications service) by big-name firms including Vodafone-Mannesmann, Deutsche Telekom, BT and Orange. Across Europe some £68 billion was pumped into public coffers as national governments in Germany, Holland and France followed Westminster's lead.

Good for the national debt, yes. But what is becoming apparent is that – in something increasingly resembling an unfunny comedy of errors – all those big companies spent all that money on jam that may never have a tomorrow.

Let me explain. GSM – Global System for Mobile communications, the current standard that your mobile phone uses (in Europe) – is fine for what it is, but it's limited.

UMTS offers much more rapid rates of data transfer, but also more bandwidth: the idea being you could use your phone as a super-duper Space Age videophone cum record player cum personal data assistant.

Smashing. Only two small problems. One, GSM has succeeded mainly because you as the con-

> **"The reason GSM is good is that the technology is not that fancy – so it works well"**

sumer actually only pays a fraction of the real cost of the phone and the calls. The companies have made their money out of expanding the market tremendously through subsidising it as a loss-leader.

Third generation is going to be too expensive to follow that model, so, secretly, firms are looking at dropping the subsidy and obliging you to pay full price for phones – which could be as much as £600 for the high-end ones.

And if the main reason to buy a 3G phone is for the fancy extra services, you can't expect them for free. The trouble is, how much to charge? Vendors are now bandying around the figure of $2000 a year revenue for each punter to make UMTS profitable for them. Fancy that as your annual phone bill?

Second, the reason GSM is good is that it's limited – the technology is not that fancy, so it works well. Now it turns out that the only way you can get to watch that pop video on your Buck Rogers phone is if you were on a hill, in line of sight of the transmitter, and no-one else was using another UMTS phone for 50 kilometres. In everyday life in a big city, it looks like GSM could give UMTS a pretty good run for its money after all.

My newspaper, *Computing*, has a terrific story about this you can find if you look up '3G' at vnunet.com. I'll whet your appetite by lifting a quote from the story: "I suspect we don't have the time, the manpower, or perhaps even the money to deliver [3G's] full potential."

Looks like we'll all be hanging on to our old mobiles for a spell longer!

Gary Flood is deputy editor of one of IWR's sister titles for IT professionals, Computing. To talk back, email: editor@iwr.co.uk