



## Comment Article

### VNUNet – The value of automated web reading

By Bob Tarzey, Service Director, Quocirca Ltd  
Feb 2009

As client-server technology developed through the early 1990s, it became clear that there was a problem. While workstations and PCs could make users more productive, and their use of computers more enjoyable, many of the most important applications, and the data associated with them, resided on proprietary mainframe computers. To make client-server really useful it needed to be linked in with these legacy applications.

Building a server application to run on a mainframe was tricky. Specialist skills were needed that the client-server vendors did not have, the mainframe was a jealously guarded and regulated environment, and some applications were so old that nobody really knew how they worked anyway. But they did work, and produced data that populated the green screens of visual display units that older users of IT will remember.

The solution that the client-server vendors came up with was screen scraping. This involved grabbing a screen produced by a mainframe application and, knowing the co-ordinates of the useful data, working through the screen image and extracting the bits required and displaying them in a fancy new client application.

All well and good. But the problem was that many mainframe applications were a bit more complicated than this, requiring a level of user interaction before the final output was produced. To solve this, more complex mainframe adapters were built that could manage the required interaction between the user and the mainframe, and eventually display the results on the client.

This potted history has been repeated recently in a slightly different guise with the web. There are a number of reasons for automating the reading of web sites: search companies want to know what content is where; price comparison sites need to keep themselves up to date; and the web can provide a wealth of competitive information about constantly changing markets.

Static web sites are one thing, and the search companies' crawlers mine such content on a daily basis. Extracting useful information from dynamic web sites is another matter, as it presents the same problem faced by the client-server vendors on the mainframe. Enter web-scraping. As with screen-scraping, web scraping could only go so far, perhaps a one-off query to produce a dynamic screen with no further interaction required. Anyone who has booked a flight online knows that the reality is more complex than this.

The best way to determine the cost and availability of a given flight is to proceed with the booking process as far as you can without actually committing to a purchase. This requires a number of interactions: departure point and destination, preferred dates, composition of party etc.

This data is often requested through a series of interactions that entice you down the path to purchase. And, of course, it is not just air travel. It is the same with hotel bookings, offers on a retailer's web site, the latest prices of hardware from IT dealers and so on. Getting to pricing information can be a highly interactive experience and, most importantly, prices are dynamic. Just making a booking for an airline ticket might mean that the price is higher for the next customer.

One way to get a comparative price for different products and services is to use price comparison web sites like Kelkoo, Pricerunner and Momondo. The best of these sites deploy a technology similar to the mainframe adapters developed by client-server vendors. The important thing is that they can do this interactively; there is no need for a component to be installed on the target web site using technology from vendors that specialise in web data extraction.

Momondo, for example, uses a product from Kapow Technologies to return near real-time



## Comment Article

prices to users on its site which specialises in finding cheap travel options. Other Kapow customers include BAE, Vodafone and Intel which use its products to keep an eye on competitors. Ubiquick and QL2 specialise in this area too.

AdvertisementAnother is Lixto, which it is not focused on price comparison web sites per se, but provides customers with up-to-the-minute information on their competitors. Lixto's main markets are all highly competitive, and include travel, automotive and consumer electronics, where prices change on a regular basis.

For an individual to monitor the range of prices manually from a huge array of agents and dealers is near impossible, and Lixto can automate the task. As Lixto points out, such information is not just important to ensure

competitive pricing, but to prevent under-pricing and losing unnecessary margin.

The company lists Fujitsu Siemens Computers and SAP among its customers. Many are from Lixto's home market of Germany, but it hopes to change this as it expands in Europe. The firm launched in the UK in February 2008, and has signed a number of new customers in the 12 months since.

What goes around comes around. Some commentators suggest that today's computing paradigm is not that different to the old mainframe: huge arrays of server blades piled high and virtualised into vast clouds of compute power accessed by users running web browsers with little local intelligence. They may have a point, and technology that kept the mainframe alive in the client-server era has been adapted to help mine the data banks of the 21st century.

### About Quocirca

Quocirca is a primary research and analysis company specialising in the business impact of information technology and communications (ITC). With world-wide, native language reach, Quocirca provides in-depth insights into the views of buyers and influencers in large, mid-sized and small organisations. Its analyst team is made up of real-world practitioners with first hand experience of ITC delivery who continuously research and track the industry and its real usage in the markets.

Through researching perceptions, Quocirca uncovers the real hurdles to technology adoption – the personal and political aspects of an organisation's environment and the pressures of the need for demonstrable business value in any implementation. This capability to uncover and report back on the end-user perceptions in the market enables Quocirca to advise on the realities of technology adoption, not the promises.

Quocirca research is always pragmatic, business orientated and conducted in the context of the bigger picture. ITC has the ability to transform businesses and the processes that drive them, but often fails to do so. Quocirca's mission is to help organisations improve their success rate in process enablement through better levels of understanding and the adoption of the correct technologies at the correct time.

Quocirca has a pro-active primary research programme, regularly surveying users, purchasers and resellers of ITC products and services on emerging, evolving and maturing technologies. Over time, Quocirca has built a picture of long term investment trends, providing invaluable information for the whole of the ITC community.

Quocirca works with global and local providers of ITC products and services to help them deliver on the promise that ITC holds for business. Quocirca's clients include Oracle, Microsoft, IBM, O2, T-Mobile, HP, Xerox, EMC, Symantec and Cisco, along with other large and medium sized vendors, service providers and more specialist firms.

Details of Quocirca's work and the services it offers can be found at

<http://www.quocirca.com>