

database marketing

B2B MARKETERS STRUGGLE WITH SOCIAL MEDIA

New survey finds B2B social media marketing mired in confusion

INTRODUCING A FRESH APPROACH TO MATCHING

Postcode Anywhere's new Cleanse+ software may be a game-changer



HAS THE NEW PAF LICENCE DELIVERED?

A year since it was unveiled, how has the latest PAF licence delivered against the objectives set for it?



DQM ON THE MAP

DATA QUALITY MANAGEMENT FINALLY HAS ITS DAY IN THE SUN

CUSTOMER ANALYSIS

GETTING THE MOST OUT OF YOUR SOFTWARE - AND ITS USERS

NEW MEDIA, OLD ISSUES

DIGITAL MEDIA AND THE EMPEROR'S NEW CLOTHES

Welcome to the adaptive cleansing engine...

Abandoning traditional key-based and fuzzy address matching techniques, Postcode Anywhere has developed a new 'adaptive cleansing engine' approach that does away with the requirement to write thousands of lines of rules – with very impressive results, finds **James Lawson**.

Never a glamorous subject, address matching is nevertheless a critical part of customer data management and processing. Extra accuracy can translate directly into greater income and reduced costs, not to mention multiple other benefits. Addressing experts Postcode Anywhere have taken a fresh approach to the matching challenge, and this month we get a sneak preview of their new application, Cleanse+.

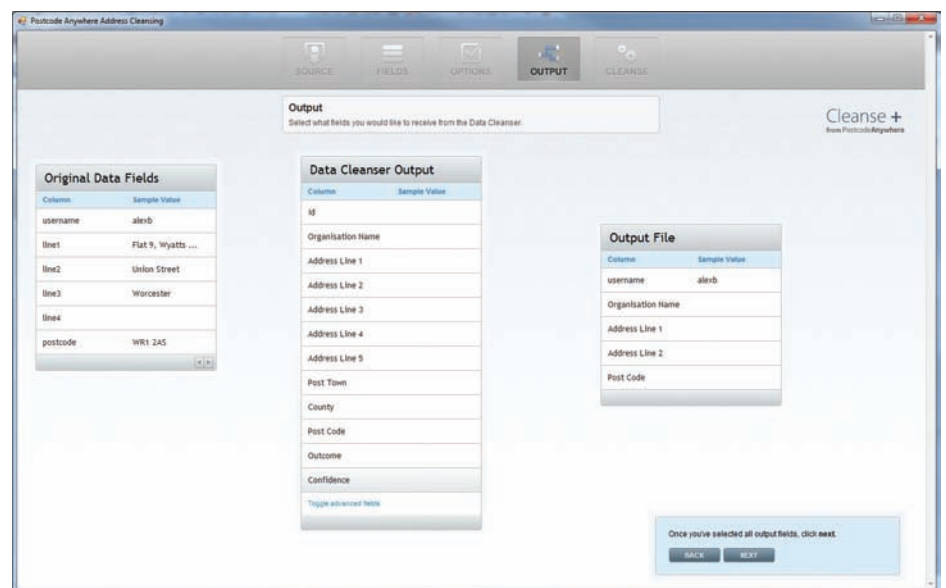
REPLACING KEYS

There have been various approaches to address matching over the years. Key-based matching is probably the most common: this involves generating a set of "keys" (character strings) that represent the addresses in question. These keys can then be compared using a variety of techniques in order to decide whether there is a match between them or not.

These techniques are usually governed by complex rule sets which are built up manually over many years using the experience gained from real world processing. Each has advantages and disadvantages of its own, the arguments about which would easily fill this magazine and a few others besides.

The vendor has chosen to consign match keys, fuzzy matching and so forth to the dustbin of history and has taken a new tack by employing what it calls an "adaptive cleansing engine". This involves using a complex self-learning mathematical model that looks at address structure and the patterns within addresses, then builds its own rules as it goes rather than relying on manually-coded rules to govern it. The software can take multiple possible versions of a final record right through processing and only make the choice right at the end.

This method obviates the need to painstakingly write thousands of lines of rules based on years of experience; the vendors claim that rule-based address cleansing software requires well over 100,000 rules to be coded into



it before it starts to be effective. It also dispenses with the inflexibility of rule-based systems: any change to address formats requires rule changes and these systems can typically only cope with addresses in one or a handful of countries due to the variation in (and lack of) global addressing standards.

One example is the traditional need to re-format records prior to PAF and other matching in order to improve the chances of a successful match. This type of solution should be format-agnostic, as long as you initially train it on the sort of data you will subsequently want to process.

The downside is that the learning process is entirely "black box". It also hoovers up vast amounts of processing power, though the vendor relies on the massively scalable capabilities of the

"cloud" to take care of that. However, though users have no influence over the learning process, the software is certainly not completely hands-off.

It's possible to create custom look-up lists to be used by the software on certain jobs, while users can also inspect the rules the software builds up to check they make sense. Over time, the software has the capability to build custom lists itself and can also add to or change those built for it using its learning capability.

The matching process is not the only difference either. Rather than extracting and then reloading address data, the vendor has gone for the "in-database" approach whereby data is cleaned in-situ by connecting directly to a database table. Aside from mailing list processing, this is very much the modern

"The engine uses a complex self-learning mathematical model that looks at address structure and the patterns within addresses, then builds its own rules as it goes."

standard as it avoids complex refreshes and also suits 24/7 systems where there is simply no time available for offline servicing.

It's on-demand too. The vendor originally made its name through innovation in its use of web services to deliver address lookup and many other customer data-related information services. This new tool is no different, with the software and associated reference files running centrally.

Users send input files down the line and verified files come back. Given today's high bandwidths, remote processing is unlikely to be a limitation for most users but the company is also planning an installed version that will run within its Enterprise Server application and so deal more effectively with the largest batch jobs.

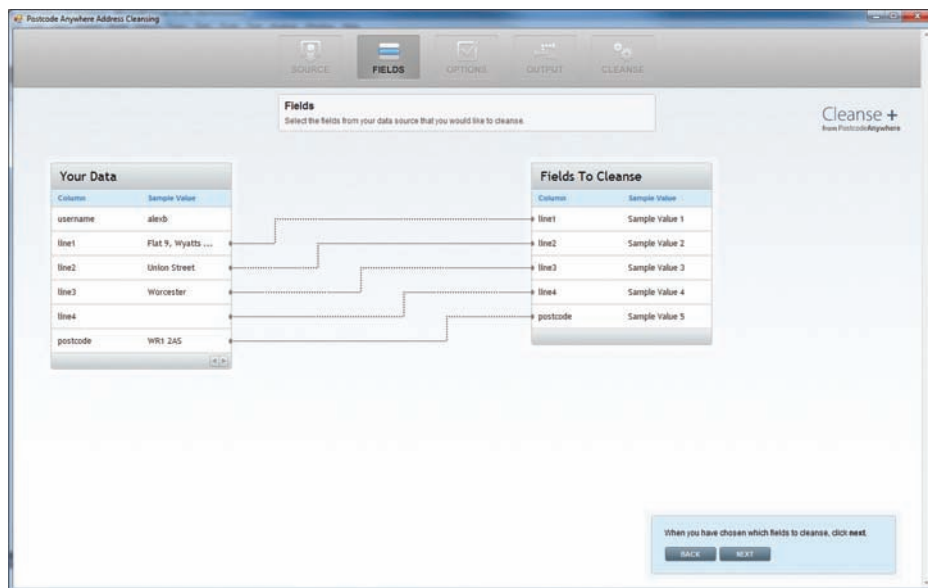
Though the version tested employed a straightforward browser-based interface, the release version of Cleanse+ will use a small desktop client rather than just a browser, allowing the company to interact with local data sources and databases as desired using standards like ODBC, and also enhancing security compared to browser-based applications.

The demo version of the interface was clear and clean, with the minimum of options. These include whether to retain vanity addresses, the number of output address lines and the fields required in the output file.

EXTREME SIMPLICITY

In contrast to the complexity under the bonnet, using the software in its current incarnation is an exercise in extreme simplicity. It offers auto-field mapping – also based on self-learning – which worked flawlessly in the demo version. There's an option to preview the input records, then it goes straight into a batch PAFing routine, matching test records against a standard, non-optimised PAF reference file and displaying the results.

Unlike other PAF matching modules, there are



very few configuration options indeed. Just like its current addressing products, the vendors say that this is to avoid confusion and complication. They also plan to use their ability to log which addresses have been captured or processed in order to alert clients when updates are needed. This will be an integral part of the Plus suite, with the future goal being fully automatic updating with no client intervention.

As well as the PAF file, the software has already had many thousands of addresses run through it to aid the learning process and the test matches were spot on. However, without running full test datasets, it's impossible to estimate the matching performance of this or any other application. As ever, it's up to buyers to decide whether it will achieve the results they want using their own test data.

This application is not yet a full competitor to all existing solutions, partly because it runs remotely over the web and so is necessarily limited in the volume of data it can batch

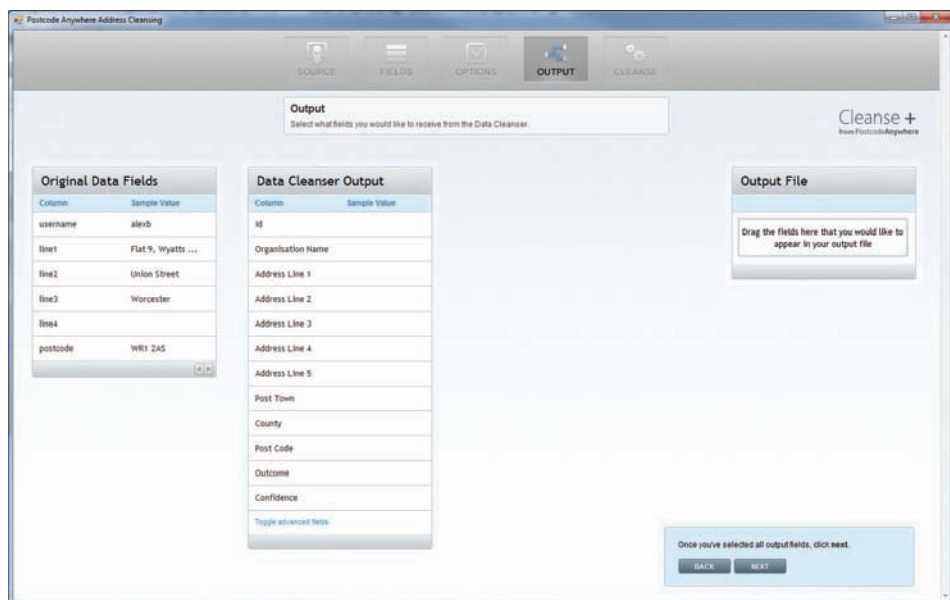
process, but mostly because of the need for further tools beyond basic PAF matching to handle the various other matching applications: deduplication, multi-file merge/purge and so forth. According to the vendors, name checking against CACI's Ocean and suppression matching will arrive shortly, along with other tools in the Plus suite that will handle data capture, updating and so on.

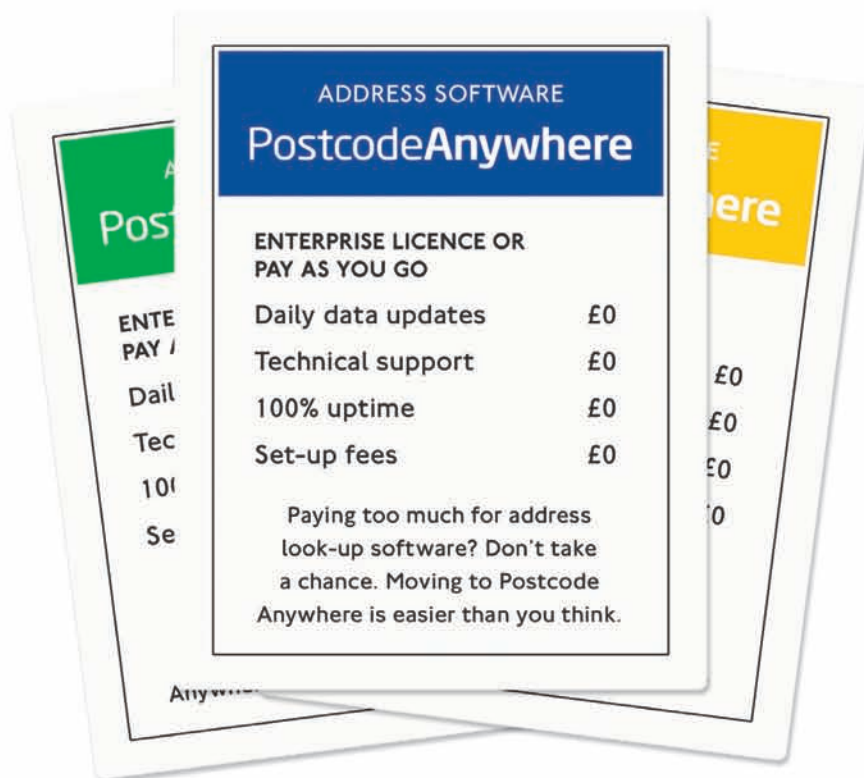
DIFFERENT APPROACH

It's worth noting that, though innovative, this isn't the first time that probabilistic techniques have replaced deterministic (or rule-based) techniques in record matching. There's a history of research in this area in the computer science fields of speech recognition and natural language processing, as well as specifically in name and address record matching.

The self-learning approach seems to be gaining supporters. US matching specialist Netrics, acquired by MDM vendor Tibco early last year, also uses a mathematical model, "that mimics human perception of similarity, identifying hidden relationships in the data." Its matching engine also includes a self-learning capability that continually improves its matching over time by evaluating other manual matches made by business users. Maybe this is the way we'll all be matching records in a few years?

Certainly this is an impressive start and Postcode Anywhere's innovation is to be applauded, and not just for the techniques it applies. It brings a welcome simplicity to the matching process that is well suited to online processing services in particular. How this approach will suit Marketing Service Providers is less easy to predict. Once again, the company is adopting the latest technology, and rattling the cages of established vendors in the process. It should be encouraged for its investment in advancing the addressing industry. ■





Premium addressing software doesn't have to come at London prices

Address auto-fill is an essential tool for fast, reliable data entry, whether on your contact database or website. But why make do with old technology that costs the earth? Switching to Postcode Anywhere is easier than you think.

- Official Royal Mail PAF® data updated automatically every day via internet
- No set-up fees, no service upgrade charges and full backwards compatibility
- Free technical support, from a team of specialists in the UK
- Effortless integration with any IT system, including bespoke software

Lloyds TSB, National Savings & Investments, British Gas, Pizza Hut, Rolls Royce, Oxfam and over 8,000 others have already moved to Postcode Anywhere. To make it easier for you too, we're offering 10% off your first credit pack or user licence.

Just quote SWITCH11 when you call or check out online. Call **0800 047 0495** now or visit postcodeanywhere.com/switch

QUOTE

“

We're saving £26,000 a year, a staggering £500 every week, since moving to Postcode Anywhere.

Seatwave