NEW MEDIA AWARD:

KR's Multilingual Markup

by John Bryan

It's tough enough to get one newspaper system to talk to a World Wide Web site quickly and reliably. But when you're running Internet services for a whole newspaper group, it's like the United Nations when the translators are on strike: all confusion, all the time.

That's the fix Knight Ridder Digital of San Jose found itself in about a year ago, but it has begun to tear down its Tower of Babel.

"The various editorial systems never have had to do much except communicate to a printing press," explains Rohn Jay Miller, KRD's senior vice president for product and technology. "Now they're being asked to do much more." That means translating proprietary markup languages into something that uses XML to move printed words to a Web browser's screen. To do it, the standard much loved by geekdom is Perl scripting.

Thanks to the Java-based Xmultra application, staffers at Knight Ridder Digital no longer have to tweak Perl scripts endlessly to convert newspaper-system feeds to XML.

"Perl code was pretty flexible and had always worked pretty well," explains Miller. "It was kind of like one of those old telephone switchboards with the cables and plugs. The Perl programmer would plug this markup language into that Web output."

The problems with this approach are twofold—each time you change the markup code, you have to have somebody change the Perl script, and the basic life-force of the newsroom (outside of holding meetings) is to change designs as often as possible.

Perl scripts, then, simultaneously provide job security for Perl programmers and

blinding headaches for managers trying to keep up with newsroom tweaks.

Too easy for you? Try doing this with vaguely worded classified ads, which contain markup twice as complex as editorial. Had enough? Oh no, you haven't. Multiply these problems by 30 newspapers (in Knight Ridder's case), and you can see that the headache has morphed into full-blown head banging.

"If you [have] one or two publications, you can put up with that kind of overhead, but it was killing us," says Miller. "We had four people working full time doing nothing but maintenance on Perl scripts."

Thus was born Xmultra, a homebrew, Java-powered conversion engine. "We built our own because there was no choice," says Miller. Senior Developer Wayne Weber led a team that put it together in about a year.

Ridiculously simplified, Xmultra works like this: Instead of writing scripts for each and every markup situation Miller and his crew must confront, they write a configuration file that serves as a system-specific table—for Atex, SII, QuarkXPress, Unisys and so on.

When a newspaper inevitably changes its markup, instead of a software engineer having to change the Perl, a non-programming production person can simply make changes in the configuration file. It's faster and frees up the programmers for other projects.

"Instead of four people full-time, we think we ultimately can have one person monitoring editorial and classified feeds, and spending only a third of the time doing maintenance," Miller says.

Moreover, Xmultra's architecture makes it wicked fast—it can process 50 documents at a time and up to 600,000 per day. So far, it's been extremely stable, running 24/7 (the lone failure was during one of California's power outages last summer).

"Xmultra addresses a need all newspapers are facing right now—the need to increase revenues for their multichannel publishing operations, or find ways to make them more efficient," says Dave Beck, NAA director of media technology, touting the system's automation of once time-consuming tasks.

Miller now wants to take the show on the road, since the system's only hardware requirement is a server running Java. "We're thinking about moving this to the Real Cities network," he says, which would double the number of feeds Xmultra handles. After that, it could easily become a product for sale to other news organizations.

Which means Miller could at least transfer the head banging out of his department.

Bryan is a member of the Los Angeles Times' pagination project team. E-mail, ***John.Bryan@latimes.com****.*

'It was kind of like one of those old telephone switchboards with the cables and plugs. The Perl programmer would plug this markup language into that Web output.'

TechNews Volume 8, Number 1: January/February 2002