

## **Mimotek Structuriser software helps UK newspapers enhance press cuttings**

Amsterdam/London, February 2007.

The Newspaper Licensing Agency's (NLA) has announced that market acceptance of its eClips database has reached over 80%. Introduced during the second half of 2006, the eClips service uses innovative technology to deliver a central database of digital clippings to users via their press cuttings agencies.

The service does not use the traditional approach of scanning newspaper pages or articles that have been cut out manually. Instead, the cuttings are generated from PDF pages that come directly from the newspapers' editorial systems, and they are delivered as high-quality PDF files.

The NLA selected Mimotek's Structuriser as the tool to be used for the production of a majority of the digital clippings. Mimotek Structuriser is software that segments PDF pages by recognising the different content elements (headlines, by-lines, body text, images etc). The Structuriser software generates each press clipping in PDF format and also extracts the corresponding text from the page to be loaded into a database that is used for searching. Branded A4 clippings are generated on request by adding metadata, branding elements and page thumbnails to the basic clipping. This process is largely automatic, an operator being required only to check, and if necessary correct, the page segmentation.

eClips offers significant advantages over scanned clippings. Most obvious is the convenience of doing everything digitally, as opposed to scanning paper and emailing or faxing clippings. But in addition, the eClips cuttings are of much higher quality since there is no degradation resulting from the scanning process. The text and graphics are represented in the PDF in such a way that they can be viewed and printed crisply at any magnification. This means that, when fitting a large article on an A4 sheet, it is possible to reduce the scaling significantly further than would be possible with scanned text, with the text remaining legible. Lastly, the text that is used for searching has been extracted directly from the page PDFs, limiting errors introduced by using OCR on a scan and significantly increasing the search accuracy.

A unique feature of Mimotek's technology is that the logical structure of the segmented page is stored in the PDF file itself, making use of PDF's tagged structure capability. This means that once a page has been processed, its contents can easily be extracted for a variety of further uses including online interactive newspapers, transforming newspaper content for reading on portable devices and personalised press clippings digests. So the segmented pages provide a basis for a wide range of future developments.

The NLA is currently generating more than 3000 digital press clippings each day from all the UK national newspapers and a selection of 12 leading regional newspapers. More titles will be added during 2007.

Andrew Hughes, Managing Director of the NLA eClips project says, “The Mimotek Structuriser software has enabled us to take full advantage of the digital feeds that have been made available by the publishers. The software also provides major benefits to our customers, offering a significant increase in quality for viewing and printing. And the fact that the text is extracted from the same PDF page ensures much more accurate searches of the clippings database.”

#### About NLA:

The Newspaper Licensing Agency (NLA) was established by the UK’s national newspapers in 1996 to manage copyright collection. The NLA authorises paper and digital copying on behalf of national, regional and international newspapers, including 1300 in the UK, for over 7,000 licensees.

#### About Mimotek

Mimotek bv ([www.mimotek.com](http://www.mimotek.com)) is an award-winning Anglo-Dutch software company that specialises in providing tools for processing the content of PDF documents. Since its launch in February 2002, the company has built on the technical experience of its employees (the majority of whom had previously worked for Adobe Systems) to establish itself as a leading supplier of PDF tools to the publishing and business communities. In addition to providing off-the-shelf tools, Mimotek also develops customised solutions for its clients.

#### **For further information, please contact:**

Andrew Hughes  
Managing Director, Digital  
Newspaper Licensing Agency  
Direct tel: +44 (0)20 7332 6040  
[ahughes@nla.co.uk](mailto:ahughes@nla.co.uk)

Micha Moses  
Managing Director  
Mimotek  
Direct tel: + 31-6- 21 27 47 54  
[moses@mimotek.com](mailto:moses@mimotek.com)