

<http://www.nla.gov.au/pub/gateways/issues/102/story06.html>

Number 102 | December 2009

## Crowdsourcing Improves Historical Newspapers

In 2008, through the [Australian Newspapers Digitisation Program](#) (ANDP) and in collaboration with Australian state and territory libraries, the National Library of Australia made available an exciting service that provides free, online access to selected out-of-copyright Australian newspapers. The [Australian Newspapers](#) service has been available for over a year and gives users the ability to search and browse over 8.4 million articles from over 830 000 newspaper pages dating from 1803, with more pages being added. This month, the early editions of *The Sydney Morning Herald* will be made available.

The Australian Newspapers service is truly innovative in the way it delivers digitised newspaper content and engages with the online user community. The ANDP has embraced web2.0 technology to provide a cutting-edge service that allows users to interact, contribute and add value to the newspaper content. The newspaper pages have been digitised and converted into electronically translated, searchable text through the use of Optical Character Recognition (OCR) software. While OCR works well for documents with a consistent typeface and standard format, the nature of historical newspapers, with varying fonts and print quality and little white space between text, means that the electronically translated text for historical newspapers is usually poor and inaccurate.

Appealing to the public to help correct and improve the data by crowdsourcing is a relatively new idea and has achieved amazing results. There is no other equivalent newspaper service in the world that allows users to tag, add comments and correct the electronically translated text. The goal that has been given to the public is to improve and correct all of the poor-quality text of Australian historical newspapers, thereby improving the searching and retrieval of articles for everyone. The public has embraced the call and has made a real difference towards realising what once seemed to the Library to be a huge and insurmountable task.

By October 2009 over 6000 members of the public had already enhanced the data significantly by correcting over 7 million lines of text in 320,000 articles, and adding 200,000 tags and 4,600 comments to articles. One exceptional user has corrected over 285,000 lines of text in over 7,000 articles. The Library has not yet actively appealed for the public to help—the ability to correct text was simply made available in the service and users have caught on.

The volunteers have reported enjoying the interactions and feeling that they are playing an important part in recording Australian history. Correctors say they are motivated to correct text for 'the common good' and that it is a worthy cause for a non-profit institution. They also report loving the task and getting 'hooked in' and 'addicted' to text correction.

The Library wants to thank and congratulate all the text correctors in the community for their hard work and encourages new users to join the crowd.

Rose Holley, Manager, Australian Newspapers Digitisation Program.