

only became enthusiastic when demonstrations showed how it could leverage older content and suggest related stories. They also saw potential as an exploratory tool for examining the links between people mentioned in large collections such as official government Hansards and documents requested through freedom of information processes. Journalists are increasingly engaged in making their content discoverable by social media and search engines, so any tools that help automate the process were seen as positive. Finally, news product managers were most excited in recommending readers more content to keep them engaged and consuming. This was seen as a way to help differentiate the products from the many others vying for the readers' attention.

4. CONCLUSION

This paper reports on a four year joint project that explored how academic research could be applied to support a commercial news media business. We presented the COMPNEWS platform that was the vehicle for research contributions in named entity linking, quotation extraction and attribution, and event linking. Moreover, it allowed us to build prototype applications that demonstrated how NLP technology could be used to support the different stages of a news story: writing, displaying, promoting and analytics. We hope that this qualitative review provides some insight into the advantages and challenges of this type of endeavour.

5. ACKNOWLEDGEMENTS

Thanks must also go to the other members of Schwa Lab (especially Tim Dawborn), our Freelancer annotators, and supporters at Fairfax and the CMCRC. This work was supported by ARC Discovery grant DP1097291.

6. ADDITIONAL AUTHORS

Additional authors who worked on the project while at the University of Sydney and the CMCRC: Will Cannings, Tim O'Keefe, Matt Honnibal, David Vadas and Candice Loxley.

7. REFERENCES

[1] T. Dawborn and J. R. Curran. docrep: A lightweight and efficient document representation framework. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 762–771, Dublin, Ireland, August 2014. Dublin City University and Association for Computational Linguistics.

[2] B. Hachey, J. Nothman, and W. Radford. Cheap and easy entity evaluation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 464–469, Baltimore, Maryland, June 2014.

[3] B. Hachey, W. Radford, J. Nothman, M. Honnibal, and J. R. Curran. Evaluating entity linking with Wikipedia. *Artificial Intelligence*, 194:130–150, January 2013.

[4] J. Nothman. *Grounding event references in news*. PhD thesis, School of Information Technologies, University of Sydney, Sydney, Australia, 2014.

[5] J. Nothman, T. Dawborn, and J. R. Curran. Command-line utilities for managing and exploring annotated corpora. In *Proceedings of the Workshop on*

Open Infrastructures and Analysis Frameworks for Human Language Technologies, Dublin, Ireland, August 2014.

[6] J. Nothman, M. Honnibal, B. Hachey, and J. R. Curran. Event linking: grounding event reference in a news archive. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 228–232, Jeju, Korea, July 2012.

[7] T. O'Keefe. *Extracting and Attributing Quotes in Text and Assessing them as Opinions*. PhD thesis, School of Information Technologies, University of Sydney, Sydney, Australia, 2014.

[8] T. O'Keefe, J. R. Curran, P. Ashwell, and I. Koprinska. An annotated corpus of quoted opinions in news articles. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 516–520, Sofia, Bulgaria, August 2013. Association for Computational Linguistics.

[9] T. O'Keefe, S. Pareti, J. R. Curran, I. Koprinska, and M. Honnibal. A sequence labelling approach to quote attribution. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 790–799, Jeju, Korea, July 2012.

[10] G. Pink, W. Radford, W. Cannings, A. Naoum, J. Nothman, D. Tse, and J. R. Curran. SYDNEY-CMCRC at TAC 2013. In *Proceedings of the Text Analysis Conference*, Gaithersburg, MD USA, November 2013. National Institute of Standards and Technology.

[11] W. Radford. *Linking Named Entities to Wikipedia*. PhD thesis, School of Information Technologies, University of Sydney, Sydney, Australia, 2015.

[12] W. Radford, W. Cannings, A. Naoum, J. Nothman, G. Pink, D. Tse, and J. R. Curran. (Almost) Total Recall – SYDNEY-CMCRC at TAC 2012. In *Proceedings of the Text Analysis Conference*, Gaithersburg, MD USA, November 2012. National Institute of Standards and Technology.

[13] W. Radford and J. R. Curran. Joint apposition extraction with syntactic and semantic constraints. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 671–677, Sofia, Bulgaria, August 2013. Association for Computational Linguistics.

[14] W. Radford, B. Hachey, M. Honnibal, J. Nothman, and J. R. Curran. Naive but effective NIL clustering baselines – CMCRC at TAC 2011. In *Proceedings of the Text Analysis Conference*, Gaithersburg, MD USA, November 2011. National Institute of Standards and Technology.

[15] W. Radford, B. Hachey, J. Nothman, M. Honnibal, and J. R. Curran. Document-level entity linking: CMCRC at TAC 2010. In *Proceedings of the Text Analysis Conference*, Gaithersburg, MD USA, November 2010. National Institute of Standards and Technology.