|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Ben Hachey – Curriculum Vitae** | | | (Last updated: 16 April 2014) | |
| t: +61 416 550 174  e: ben.hachey [at] gmail.com | | School of Information Technologies  University of Sydney, NSW 2006, Australia | | |
|  | |  | | |
| Employment | | | | |
|  | **September 2013 – Present** | | | **University of Sydney** |
|  | DECRA Research Fellow, School of Information Technologies | | | |
|  | **Role:** Early career researcher funded by Australian Research Council.  **Software:** The project will improve the coverage and reliability of collaborative resources (e.g., Wikipedia), building high-performance tools for linking text to entity KBs and using them to develop comprehensive relation extraction and fact-driven summarisation. | | | |
|  |  | | | |
|  | **December 2011 – September 2013** | | | **Thomson Reuters** |
|  | Senior Research Scientist, Research & Development | | | |
|  | **Role:** Initiate/lead R&D in large-scale language technology for professional content.  **Software:** Induction of financial risk taxonomies, identifying extended entity types in legal text, linking entities across content, text-driven automation of master data management.  **Other:** Consult on, e.g., linked data, Wikipedia, NER in queries, biography generation, patent summarisation, data science and evaluation. Manage machine-learning intern. | | | |
|  |  | | | |
|  | **October 2008 – October 2011** | | | **Macquarie University** |
|  | Research Fellow, Centre for Language Technology | | | |
|  | **Role:** R&D in language technology for financial analytics and news curation.  **Software:** Tera-scale text mining tools for fraud detection, algorithmic decision-making, and automated news curation. | | | |
|  |  | | | |
|  | **November 2002 – September 2004** | | | **University of Edinburgh** |
|  | Research Associate, Institute for Communicating and Collaborative Systems | | | |
|  | **Role:** Drove research agenda on several information extraction and summarisation projects.  **Software:** State-of-the-art active learning tools for information extraction (optioned to Cognia Corporation). | | | |
|  |  | | | |
| Education | | | | |
|  | **January 2009** | | | **University of Edinburgh** |
|  | PhD, Informatics (supervisor: Claire Grover) | | | |
|  | **Thesis:** Towards Generic Relation Extraction | | | |
|  | *\*\*\*Edinburgh-Stanford Link PhD Studentship\*\*\** | | | |
|  |  | | | |
|  | **Autumn 2002** | | | **University of Edinburgh** |
|  | MSc, Speech and Language Processing (supervisor: Mirella Lapata) | | | |
|  | **Thesis:** Recognising Clauses Using Symbolic and Machine Learning Approaches | | | |
|  |  | | | |
|  | **Spring 2001** | | | **University of Minnesota** |
|  | B.A., Computer Science and Linguistics (Spanish minor) | | | |
|  | Double major with coursework emphasis on Computational Linguistics, Syntax and AI. | | | |
|  |  | | | |
| Supervision | | | | |
|  | **PhD:** Andrew Chisholm; Joel Nothman, (2013, associate); Will Radford (2013, associate) | | | |
|  | **Honours/MSc:** Ka Shing Law (2010, Honours); Gareth Robinson (2010, Honours); Will Radford (2009, Honours, first class); Tamsin Maxwell (2007, MSc, distinction,) | | | |
|  |  | | | |
| Teaching | | | | |
|  | **Guest/Lecturer:** *Statistical Natural Language Processing* (2010-11, Honours, USydney) | | | |
|  | **Teaching Assistant:** *XML Document Management* (2005-07, Masters, UEdinburgh, with Henry Thompson); *Processing Natural and Formal Languages* (2005-06, Undergraduate, UEdinburgh, with Bonnie Webber); *Introduction to Computational Linguistics* (2003, Masters, UEdinburgh, with Ewan Klein); *Introduction to Syntax* (2001, Undergraduate, UMinnesota, with Hooi Ling Soh) | | | |
|  |  | | | |
| Research Interests | | | | |
|  | Computational linguistics; Machine learning; Information extraction; Data integration. | | | |
|  |  | | | |
| Funding | | | | |
|  | Wikilinks (2013-2016, Australian Research Council, principal investigator) | | | |
|  | Computable News (2010-2011, Fairfax/CMCRC, author & project lead) | | | |
|  | EASIE (2004-2007, Edinburgh-Stanford Link, named student) | | | |
|  |  | | | |
| Service | | | | |
|  | **Reviewing:** Journal of artificial intelligence research, Language resources and evaluation, Information processing and management, Natural language engineering, etc. | | | |
|  | **Conference PCs:** ACL, EMNLP, COLING, NAACL, SIGIR, etc. | | | |
|  | **Organisation:** NAACL workshop on computational linguistics in a world of social media; Finding the hidden knowledge: text mining for biology and medicine. | | | |
|  |  | | | |
| Technical Skills | | | | |
|  | Various programming languages (e.g., Python, Perl, Java, C/C++, XSLT) | | | |
|  | Various data management tools (e.g., MongoDB, Tokyo Cabinet, PostgreSQL, Django) | | | |
|  | System/corpus design and development (e.g., IE, Summarisation, QA) | | | |
|  | Evaluation tools (e.g., <https://github.com/wikilinks/conll03_nel_eval>) | | | |
|  | Various language/speech and machine learning tools (e.g., LT TTT, NLTK, Weka, R) | | | |
|  | Spanish (minor, UMinnesota), Ojibwe (1 year, UMinnesota) | | | |
|  |  | | | |

|  |  |  |
| --- | --- | --- |
| Selected Publications | | (Citation counts from [Google Scholar](http://scholar.google.com.au/citations?user=Oh3A0P0AAAAJ). Current h-index: 12.) |
|  | Evaluating entity linking with Wikipedia. **B. Hachey**, W. Radford, J. Nothman, M. Honnibal and J. R. Curran. *Artificial Intelligence*, 193: 130-150*,* 2013. (Citations: 19) | |
|  |  | |
|  | Event linking: grounding event reference in a news archive. J. Nothman, M. Honnibal, **B. Hachey** and J. Curran. In *ACL*, pages 228-232, 2011. (Citations: 3) | |
|  |  | |
|  | Graph-based named entity linking with Wikipedia. **B. Hachey**, W. Radford and J. Curran. In *WISE*, pages 213-226, 2011. (Citations: 8) | |
|  |  | |
|  | Document-level entity linking: CMCRC at TAC 2010. W. Radford, **B. Hachey**, J. Nothman, M. Honnibal and J. R. Curran. In *Proceedings of the Text Analysis Conference*, 2010. (Citations: 16) | |
|  |  | |
|  | Multi-document summarisation using generic relation extraction. **B. Hachey**. In *EMNLP*, pages 420-429, 2009. (Citations: 12) | |
|  |  | |
|  | Tracking information flow in financial text. W. Radford, **B. Hachey**, J. R. Curran and M. Milosavljevic. In *ALTW*, pages 11-19, 2009. (Citations: 3) | |
|  |  | |
|  | Extractive summarisation of legal texts. **B. Hachey** and C. Grover. *Artificial Intelligence and Law*, 14(4):305-345, 2006. (Citations: 33) | |
|  |  | |
|  | Investigating the effects of selective sampling on the annotation task. **B. Hachey**, B. Alex and M. Becker. In *CoNLL*, pages 144-151, 2005. (Citations: 35) | |
|  |  | |
|  | Optimising selective sampling for bootstrapping named entity recognition. M. Becker, **B. Hachey**, B. Alex and C. Grover. In *ICML Workshop on Learning with Multiple Views*, pages 5-12, 2005. (Citations: 29) | |
|  |  | |
|  | A rhetorical status classifier for legal text summarisation. **B. Hachey** and C. Grover. In *ACL Text Summarization Branches Out Workshop*, pages 35-42, 2004. (Citations: 13) | |
|  |  | |
|  | Use of ontologies for cross-lingual information management in the web. **B. Hachey**, C. Grover, V. Karkaletsis, A. Valarakos, M.-T. Pazienza, M. Vindigni, E. Cartier and J. Coch. In *Workshop on Ontologies and Information Extraction*, 2003. (Citations: 11) | |
|  |  | |
|  | The Embra system at DUC 2005: query-oriented multi-document summarization with a very large latent semantic space. **B. Hachey**, G. Murray and D. Reitter. In *Proceedings of the Document Understanding Conference*, 2005. (Citations: 25) | |
|  |  | |