BEN HACHEY – CURRICULUM VITAE

T: +1 612 913 0057

E: BEN.HACHEY [AT] GMAIL.COM

RESEARCH & DEVELOPMENT, THOMSON REUTERS

610 OPPERMAN DRIVE, ST PAUL, MN 55123

(Last updated: 15 August 2013)

EMPLOYMENT

December 2011 – Present　　　　　　　　**Thomson Reuters**
Snr Research Scientist, Research & Development
Role: Initiate/lead R&D in large-scale language technology for professional content.
Software: Tools for mining taxonomies of risk terms, identifying extended entity types in legal text, linking entities across content, maintaining entity authorities.
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).

June 2001 – August 2001　　　　　　　　**BCL Technologies**
Programmer Intern, Spoken Language User Interface Toolkit
Software: Designed and implemented C++ text normalisation code.

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: Joel Nothman, (PhD, CS, associate); Will Radford (PhD, CS, associate)
MSc: Tamsin Maxwell (2007, Natural Language Engineering, distinction, co-supervisor)
Honours: Ka Shing Law (2010, CS, co-supervisor); Gareth Robinson (2010, CS, co-supervisor); Will Radford (2009, CS, first class, co-supervisor)

TEACHING

Lecturer: *Statistical Natural Language Processing* (2011, Honours/Masters, U of Sydney)
Guest Lecturer: *Statistical Natural Language Processing* (2010, Honours/Masters, U of Sydney)
Teaching Assistant: *XML Document Management* (2005, 2006, 2007, Masters, U of Edinburgh, with Henry Thompson); *Processing Natural and Formal Languages* (2005, 2006, Ugrad, U of Edinburgh, with Bonnie Webber); *Introduction to Computational Linguistics* (2003, Masters, U of Edinburgh, with Ewan Klein); *Introduction to Syntax* (2001, Ugrad, U of Minnesota, with Hooi Ling Soh)

RESEARCH INTERESTS

Web-scale text mining; Information extraction/aggregation; Automatic summarisation; Computational linguistics and social media; Un/semi-supervised learning for NLP

FUNDING

Computable News (2010-2011, Fairfax/CMCRC, author & project lead)
EASIE (2004-2007, Edinburgh-Stanford Link, named student)

SERVICE

Organisation: NAACL Workshop on Computational Linguistics in a World of Social Media (2010); Workshop on Text Mining for Biology and Medicine (2008).
Journal Reviewing: Journal of Artificial Intelligence Research; Language Resources and Evaluation; Information Fusion; Information Processing and Management; Artificial Intelligence and Law; Natural Language Engineering.
Conference PCs: EMNLP (2013, 2011, ***Best reviewer award 2010***); ACL (2013, 2011); ACL Student Session (2011); NAACL Student Session (2012); EACL (2012, 2009); COLING (2010); Australasian Language Technology (2011, 2010, 2009, 2008); ICAIL (2013, 2011); AIRS (2010); SIGIR (2010, 2009).
Workshop PCs: Language Technology for Cultural Heritage, Social Sciences, and Humanities (ACL 2013, ACL 2011, ECAI 2010, EACL 2009, LREC 2008); Web Science and Information Exchange in the Medical Web (WWW 2011); Applying Human Language Technology to Law (ICAIL 2011); Active Learning for Natural Language Processing (NAACL 2010, NAACL 2009); Mining User-Generated Content for Security (2009); CLUK Research Colloquium (2007).
Student Volunteer: ACL (2006, 2005).
Internal: Reading group (2010-2011, ə-Lab), Fire steward (2005-2008, U of Edinburgh).

CONSULTING

July – September 2008   **Capital Markets CRC Limited**
Consultant, Financial Text Mining
Advised on setup of text mining infrastructure.

July 2006   **ITI Scotland Limited**
Consultant, Text Mining Review
Reviewed portability of automated text curation software to different application domains (e.g., scientific literature, resumes, law and legal opinion).

INDUSTRY OUTREACH

I am a strong advocate of the use of language technology in the commercial world. I have contributed insight and advice to technology managers at Thomson Reuters, Royal Bank of Scotland, Wotnews, Sirca, Ernst and Young, Nuix and Fairfax Digital.

TALKS

Bootstrapping Named Entity Recognition: Active Learning and Annotator Performance, Thomson West, Eagan, MN, USA (2005)
Bootstrapping Named Entity Recognition: Active Learning and Annotator Performance, Stanford University (2005)

TECHNICAL SKILLS

Various programming languages (e.g., Python, Perl, Java, C/C++, XSLT)
Various data management tools (e.g., XML, Tokyo Cabinet, PostgreSQL, Django)
System/corpus design and development (e.g., IE, Summarisation, QA)
Various language/speech and machine learning tools (e.g., LT TTT, NLTK, Weka, R)
Spanish (minor, U of Minnesota), Ojibwe (1 year, U of Minnesota)
French, German, Italian (travel-oriented independent study)

CERTIFICATES/TRANSFERABLE SKILLS

IP in IT and Patent Searching (2006, U of Edinburgh)
Managing Research Projects (2006, U of Edinburgh)
Time Management (2006, U of Edinburgh)
Communicating Science to Non-specialists (2005, U of Edinburgh)
European Master's in Language and Speech (2002, ISCA & EACL)
XML Document Management (2002, U of Edinburgh)

Journal Articles

Evaluating Entity Linking with Wikipedia. **B. Hachey**, W. Radford, J. Nothman, M. Honnibal and J. R. Curran. *Artificial Intelligence*, 193: 130-150 (2013). (citations: 9)

Datasets for Generic Relation Extraction. **B. Hachey**, C. Grover and R. Tobin. *Journal of Natural Language Engineering*, 18(1): 21-59 (2011). (citations: 4)

Extractive Summarisation of Legal Texts. **B. Hachey** and C. Grover. *Artificial Intelligence and Law: Special Issue on E-government*, 14(4):305-345 (2006). (citations: 32)

Refereed Conference Papers

Event Linking: Grounding Event Reference in a News Archive. J. Nothman, M. Honnibal, **B. Hachey** and J. Curran. In *Proceedings of the 50th Annual Meeting of the Association for Compntational Linguistics*, Korea (2011). (citations: 3)

Graph-based Named Entity Linking with Wikipedia. **B. Hachey**, W. Radford and J. Curran. In *Proceedings of the 12th International Conference on Web Information System Engineering*, Sydney (2011). (citations: 5)

Multi-Document Summarisation Using Generic Relation Extraction. **B. Hachey**. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Singapore (2009). (citations: 11)

Generic Relation Identification: Models and evaluation. **B. Hachey**. In *Australasian Language Technology Workshop*, Sydney, Australia (2009).

Tracking Information Flow in Financial Text. W. Radford, **B. Hachey**, J. R. Curran and M. Milosavljevic. In *Australasian Language Technology Workshop*, Sydney, Australia (2009). (citations: 3)

Investigating the Effects of Selective Sampling on the Annotation Task. **B. Hachey**, B. Alex and M. Becker. In *9th Conference on Computational Natural Language Learning*, Ann Arbor, MI, USA (2005). (citations: 32)

Sequence Modelling for Sentence Classification in a Legal Summarisation System. **B. Hachey** and C. Grover. In *20th Annual ACM Symposium on Applied Computing*, Santa Fe, NM, USA (2005). (citations: 21, ***ACM SIGAPP Student Travel Award***)

Automatic Legal Text Summarisation: Experiments with summary structuring. **B. Hachey** and C. Grover. In *10th International Conference on Artificial Intelligence and Law*, Bologna, Italy (2005). (citations: 15)

Sentence Extraction for Legal Text Summarisation. **B. Hachey** and C. Grover. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence (Posters)*, Edinburgh, Scotland (2005).

Sentence Classification Experiments for Legal Text Summarisation. **B. Hachey** and C. Grover. In *17th Annual Conference on Legal Knowledge and Information Systems*, Berlin, Germany (2004). (citations: 7)

Adaptive, Multilingual Named Entity Recognition in Web Pages. G. Petasis, V. Karkaletsis, C. Grover, **B. Hachey**, M.-T. Pazienza, M. Vindigni, J. Coch. In *Proceedings of the 16th European Conference on Artificial Intelligence (Posters)*, Valencia, Spain (2004). (citations: 4)

Automatic Summarisation of Legal Documents. C. Grover, **B. Hachey**, I. Houghson and C. Korycinski. In *9th International Conference on Artificial Intelligence and Law*, Edinburgh, Scotland (2003). (citations: 24)

Demonstration of the CROSSMARC System. V. Karkaletsis, C. D. Spyropoulos, D. Souflis, C. Grover, **B. Hachey**, M.-T. Pazienza, M. Vindigni, E. Cartier and J. Coch. In *Proceedings of the Human Language Technology Conference (Demonstrations)*, Edmonton, Canada (2003).

Refereed Workshop Papers

Tracking Information Flow between Primary and Secondary News Sources. W. Radford, **B. Hachey**, J. R. Curran and M. Milosavljevic. In *NAACL Workshop on Computational Linguistics in a World of Social Media*, Los Angeles, CA, USA (2010).

Automating Financial Surveillance. M. Milosavljevic, J.-Y. Delort, **B. Hachey**, B. Arunasalam, W. Radford and J. R. Curran. In *Mining User-Generated Content for Security*, Venice, Italy (2009). (citations: 2)

Domain-Neutral Relation Characterisation: Evaluation on disease-treatment data. **B. Hachey**. In *ISMB BioLINK SIG Meeting: Linking Literature, Information and Knowledge for Biology*, Vienna, Austria (2007). (citations: 1)

Comparison of Similarity Models for the Relation Discovery Task. **B. Hachey**. In *ACL Workshop on Linguistic Distances*, Sydney, Australia (2006). (citations: 6)

Dimensionality Reduction Aids Term Co-occurrence based Multi-Document Summarisation. **B. Hachey**, G. Murray and D. Reitter. In *ACL Workshop on Task-Focused Summarization and QA*, Sydney, Australia (2006). (citations: 4)

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*, Bonn, Germany (2005). (citations: 27)

The HOLJ Corpus: Supporting summarisation of legal texts. C. Grover, **B. Hachey** and I. Hughson. In *COLING Workshop on Linguistically Interpreted Corpora*, Geneva, Switzerland (2004). (citations: 7)

A Rhetorical Status Classifier for Legal Text Summarisation. **B. Hachey** and C. Grover. In *ACL Text Summarization Branches Out Workshop*, Barcelona, Spain (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*, Bucharest, Romania (2003). (citations: 11)

Summarising Legal Texts: Sentential tense and argumentative roles. C. Grover, **B. Hachey** and C. Korycinski. In HLT-NAACL *Workshop on Text Summarization*, Edmonton, Canada (2003). (citations: 24)

Shared Task Papers

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*, Gaithersburg, MD, USA (2010). (citations: 7)

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*, Vancouver, BC, Canada (2005). (citations: 24)

Selective Sampling for Information Extraction. **B. Hachey**, M. Becker, C. Grover and E. Klein. In *Proceedings of the PASCAL Challenge Workshop on Evaluating Machine Learning for Information Extraction*, Southampton, UK (2005).

Grounding Gene Mentions with Respect to Gene Database Identifiers. **B. Hachey**, H. Nguyen, M. Nissim, B. Alex and C. Grover. In *BioCreAtIvE Workshop (Critical Assessment of IE Systems in Biology)*, Granada, Spain (2004). (citations: 7)

Theses

Towards Generic Relation Extraction. **B. Hachey.** PhD Thesis, University of Edinburgh (2009). (citations: 8)

Recognising Clauses using Symbolic and Machine Learning Approaches. **B. Hachey.** Master's Thesis, University of Edinburgh (2002). (citations: 14)