

BEN HACHEY – CURRICULUM VITAE

(Last updated: 10 January 2011)

T: +61 414 917 683
E: BEN.HACHEY [AT] MQ.EDU.AU

DEPARTMENT OF COMPUTING, MACQUARIE UNIVERSITY
SYDNEY NSW 2109, AUSTRALIA

EMPLOYMENT

-
- October 2008 – Present** **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 Cognition 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

Current: Joel Nothman (PhD, CS, associate supervisor); Will Radford (PhD, CS, associate supervisor)
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)
Undergraduate: Sean Gardiner, George Karpenkov, Sophie Liang, Bin Zhou (2010, independent studies project)

TEACHING

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

FUNDING

Computing News (2010-2011, Fairfax/CMCRC, author & steering group member)
EASIE (2004-2007, Edinburgh-Stanford Link, named student)

RESEARCH INTERESTS

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

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 and Law; Journal of Natural Language Engineering

Conference PCs: ACL (2011); ACL Student Session (2011); ICAIL (2011); AIRS (2010); EMNLP (2010, ***Best reviewer award***); COLING (2010); SIGIR (2010, 2009), EACL (2009)

Workshop PCs: Language Technology for Cultural Heritage, Social Sciences, and Humanities (NAACL 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); Australasian Language Technology (2010, 2009, 2008); Active Learning for Natural Language Processing (NAACL 2010, NAACL 2009); Mining User-Generated Content for Security (2009); CLUK Research Colloquium (2007)

Student Volunteer: ACL (2005, 2006)

Internal: Reading group (2010-present, 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, RBS, 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)

MEMBERSHIPS

Association for Computational Linguistics

PERSONAL

United States citizen

Currently hold PR visa for Australia (Subclass 121)

Journal Articles

Datasets for Evaluation of Multi-type Relation Extraction across Domains. **B. Hachey**, C. Grover and R. Tobin. *Journal of Natural Language Engineering* (to appear).

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: 21)

Refereed Conference Papers

Multi-Document Summarisation Using Generic Relation Extraction. **B. Hachey**. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Singapore (2009). (citations: 5)

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: 29)

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: 17, ***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: 9)

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: 5)

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: 17)

Refereed Workshop Papers

Tracking Information Flow between Primary and Secondary News Sources. W. Radford, **B. Hachey**, J. 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. Curran. In *Mining User-Generated Content for Security*, Venice, Italy (2009). (citations: 1)

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. Curran and M. Milosavljevic. In *Australasian Language Technology Workshop*, Sydney, Australia (2009). (citations: 3)

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: 4)

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).

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: 17)

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: 6)

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: 10)

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: 19)

Refereed Poster/Demonstration Papers

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

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*, Valencia, Spain (2004). (citations: 4)

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 (HLT/NAACL)*, Edmonton, Canada (2003).

Unrefereed Workshop Papers

CMCRC at TAC10: Document-level entity linking with graph-based reranking. W. Radford, **B. Hachey**, J. Nothman, M. Honnibal and J. Curran. In *Proceedings of the Text Analysis Conference*, Gaithersburg, MD, USA (2010).

Datasets for Comparative Evaluation of Relation Extraction in the Biomedical and News Domains. **B. Hachey**, C. Grover and R. Tobin. In *Finding the Hidden Knowledge: Text Mining for Biology and Medicine*, Glasgow, Scotland (2008).

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: 18)

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: 6)

Theses

Towards Generic Relation Extraction. **B. Hachey**. PhD Thesis, University of Edinburgh (2009). (citations: 2)

Recognising Clauses using Symbolic and Machine Learning Approaches. **B. Hachey**. Master's Thesis, University of Edinburgh (2002). (citations: 10)