

Curriculum Vitae

Mark James Carman BA & BE(EE), Ph.D

Senior Lecturer, Faculty of Information Technology, Monash University
Melbourne, Australia

August 24, 2017

1 Career History

Jan 2015 - Present: Senior Lecturer (Tenured Position at Level C),
Faculty of Information Technology
Monash University, Melbourne, Australia

Nov 2010 - Dec 2014: Lecturer (Tenured Position at Level B),
Faculty of Information Technology,
Monash University, Melbourne, Australia.

Aug 2007 - Oct 2010: Postdoctoral Researcher in the Faculty of Informatics,
University of Lugano (Universita' della Svizzera Italiana), Switzerland.

Jan 2005 - Nov 2006: Visiting PhD Student/Scholar at the Information Sciences Institute,
University of Southern California, Los Angeles, USA.

Dec 2001 - Dec 2004: PhD Student at ITC-irst and the Department of ICT,
University of Trento, Italy.

Jan 2001 - Nov 2001: Assistant Researcher in the Automated Reasoning Systems Division,
ITC-irst, Trento, Italy.

Jan 2000 - Dec 2000: Research Technologist, E-Commerce Division,
Telstra Research Laboratories, Sydney, Australia.

2 Education and Qualifications

2.1 Education

Ph.D (Dottorato di Ricerca) in Information and Communication Technologies,

University of Trento, Italy, July 2006.

Thesis title: *Learning Semantic Definitions of Information Sources on the Internet*

Bachelor of Engineering (Electrical & Electronic) and Bachelor of Arts

with First Class Honours in Engineering

University of Adelaide, Australia, 1999.

2.2 Languages Spoken

Dr. Carman is tri-lingual:

- *English: Native speaker*
He was born and raised in South Australia.
- *Italian: Highly fluent*
Dr. Carman lived and worked in Trento from 2001 to 2004 and in Lugano from 2007 to 2010. He became an Italian citizen in 2014 and regularly spends time in Italy with relatives and friends.
- *German: Fluent*
Dr. Carman studied German at high school and university, achieving a Mittelstufe II certification in 1998. He studied for one semester at the University of Stuttgart in 1998/99 and majored in German studies (as part of his double degree in Engineering/Arts) at the University of Adelaide in 2000.

2.3 Awards and Fellowships

2011: Faculty of Information Technology Early Career Researcher Award, Monash University, AUD 10,000

2009: Offered ERCIM Fellowship (European Research Consortium for Mathematics & Informatics)

2006: Nominated by the University of Trento for an ACM Doctoral Dissertation Award

1997: The E.V. Clarke Prize for Electrical & Electronic Engineering, University of Adelaide

1997: Dean's Merit Certificate for 2nd Year Results in Engineering, University of Adelaide

1996: Dean's Merit Certificate for 1st Year Results in Engineering, University of Adelaide

1995: Australian Students' Prize (on High School completion)

3 Research

3.1 Research Interests and Expertise

Dr. Carman is an expert in **Data Science**, **Data Mining** and the analysis of **Big Data**. His primary research interests are **Information Retrieval** and **Machine Learning**. His specific areas of expertise include:

- *Regression & Rank Learning*: machine learning algorithms for classification and regression with large data quantities, especially structured learning problems such as rank learning for Web Search.

- *Statistical Modeling of Text*: machine learning and topic modelling techniques applied to various problems in information retrieval.
- *Personalized Search & Recommender Systems*: Models for recommending items to users and personalising web search results to user interests.
- *Analysis of User-Generated Data*: Application of Machine Learning techniques (including sentiment analysis) to the modeling of blogs, tags, tweets and query-log data.
- *Relational Learning and Automated Planning* applied to problems in data integration, Web Services and grid computing.

His research and interests span theoretical studies (e.g. investigating statistical properties of information retrieval measures), through to practical applications (e.g. technology for assisting police during digital forensic investigations).

Dr. Carman has authored a large number of publications in prestigious venues, including full papers at SIGIR, KDD, IJCAI, CIKM, ECIR, WSDM, HT, CoNLL, EACL, HCOMP and ICDAR, and articles in TOIS, IR, JMLR, ML, PR, JAIR, CS&L, JASIST, DI and CSUR. According to Google Scholar, his publications have:

- a citation count over 1150
- an h-index of: 17

Major contributions of his research career have included developing **state-of-the-art techniques** for:

- learning Web search ranking functions and transferring knowledge across collections [6, 19, 27, 28],
- fast Machine Learning algorithms allowing systems to scale up to very large datasets [2, 5, 26, 8],
- clustering high-dimensional data using density-based and subspace-clustering techniques [4, 22],
- improving quality-control for crowd-sourcing applications [16, 44],
- accelerating digital forensic investigations and analysis of the Dark Web [7],
- personalising Web search results to the interests of individual searchers [30, 35],
- characterising, detecting and generating sarcasm in text [1, 15, 21, 24, 25, 45, 47, 49],
- optical-character recognition error correction for Indic languages [14],
- efficiently and accurately evaluating Named Entity extraction systems [3, 18],
- recommending products and services based on rating and click data (aka collaborative filtering) [32],
- modelling the category structure in Wikipedia for document classification and clustering [20, 48],
- ranking of weblogs based on content and expressed sentiment [9, 37, 10],
- analysing and leveraging tag data in information retrieval [39, 50, 33],
- routing queries to appropriate collections within search engines [41, 29],
- learning semantic descriptions of online services [13, 42],
- planning for the automated composition of web services [54, 56], and
- managing data repositories on a data grid [43].

3.2 Research Supervision

Dr. Carman is a prolific and dedicated research supervisor. He has supervised two PhD student to completion and currently supervises seven other PhD students, two of whom have already submitted their theses for examination. He is versatile in his supervision methods, doing some of his supervision via Skype, with three of his students based in Mumbai (India), one based in Kuala Lumpur (Malaysia), the remainder spread across three different university campuses in Melbourne.

Completed PhD students:

- Muhammad Ibrahim, Monash University (11/01/2017)
Thesis title: *Scalability and Performance of Random Forest based Learning-to-Rank for Information Retrieval*
- Ye Zhu, Monash University (31/05/2017)
Thesis title: *Efficient Identification of Arbitrarily Shaped and Varied Density Clusters in High-dimensional Data*

PhD students with theses currently under examination:

- Ramakrishna Bairi, IITB-Monash Research Academy
Thesis title: *Adaptive Organization of Digital Documents using Knowledge Graphs*
- Aditya Joshi, IITB-Monash Research Academy
Thesis title: *Investigations in Computational Sarcasm*

Dr. Carman currently supervises the following PhD projects:

- Estimating user expertise in user-generated content environments.
Monash University (Malaysia)
- Accelerating digital forensic investigations.
Monash University (Caulfield)
- Improving the reliability of crowd-sourced data.
Monash University (Caulfield)
- Unsupervised transfer learning techniques for learning to rank.
RMIT University (Melbourne)
- Improving reliability of optical character technology for Indic scripts.
IITB-Monash Academy (Mumbai)

In addition to PhD student supervisions, Dr. Carman has supervised to completion 2 Honours students and 5 Masters minor-thesis students. He has also examined a large number of Honours and Minor thesis projects.

Dr Carman has examined 1 Masters-by-Research thesis. He has been on progress review committees for a large number of PhD students at Monash, and was asked to be an auxiliary member of a Spanish PhD thesis examination panel.

3.3 Research Grants

Mark Carman has received competitive research project funding from research bodies in Australia, Asia and Switzerland.

2014: Victorian Government Department of State Development, Business and Innovation, Technology Voucher Program (with QSR International Pty Ltd): *Large-scale Qualitative Data Analysis*, AUD 64,793

- 2014: National ICT Australia Limited (NICTA) Collaborative Research Project: *Improved Statistical Models of Document Corpora and User-Click Behaviour with Application to Personalised Patent Search*, AUD 60,000
- 2012: Co-applicant for US Air Force Asian Office of Aerospace Research and Development (AOARD) grant FA2386-12-1-4030: *Dynamic dimensionality selection for Bayesian classifier ensembles*, USD 270,000
- 2010: Co-applicant for Swiss National Science Foundation (SNSF) grant 130208: *Mining Conversational Content for Topic Modelling and Author Identification (ChatMiner)*, CHF 156,540

3.4 Research Project Participation

Prior to gaining a faculty position at Monash, Dr. Carman also participated in a number of European, Swiss and US projects:

- 2008/9: ADIR+ project, Hasler Stiftung
- 2007/8: ADIR project, Swiss National Science Foundation
- 2005/6: CALO project, Defense Advanced Research Projects Agency, DoD
- 2004: KnowledgeWeb project, European Research Council
- 2002/3: Edamok project, Provincia Autonoma di Trento
- 2001: DataGrid Project, European Research Council

4 Teaching

Dr. Carman is passionate about teaching. He has over 8 years experience teaching a variety of different undergraduate and postgraduate courses at Monash University and the University of Lugano (Universita' della Svizzera Italiana). He has taught courses ranging from introductory level through to advanced and consistently receives strong positive feedback from students regarding the content and delivery of the courses he teaches and has been nominated on multiple occasions for teaching awards.

4.1 Courses taught

The following is a list of courses taught by Dr. Carman, in which he filled the role of both Lecturer and Examiner. For most courses taught at Monash University he also managed a number of tutors and took one of the tutorials for the course himself. Dr. Carman regularly revises and updates the material for each of the courses he teaches, emphasising the practical aspects of each course while making the content as exciting, current and coherent as possible.

- *Introduction to Data Science (FIT5145)*, Master of Data Science, Monash University
 - Years taught: 2016, 2017
 - Main topics covered: *roles, resources, technologies and methods involved in data science projects; data cleaning, analysis, management and storage*
 - Median student evaluation of teaching score (1-low to 5-high): 3.80
- *Computer models for business decision making (FIT2017)*, B. of Business Information Systems, Monash

- Years taught: 2014, 2015, 2017
- Main topics covered: *linear programming, non-linear programming, inventory modelling, decision making under uncertainty, basic forecasting*
- Median student evaluation of teaching score (1-low to 5-high): 3.86, 4.19, 3.91
- *IT Research Methods (FIT4005/5125)*, Bachelor of Computer Science (Hons) / Master of IT, Monash
 - Years taught: 2012, 2013, 2016, 2017
 - Main topics covered: *experiment design, probability theory, statistical hypothesis testing*
 - Median student evaluation of teaching score (1-low to 5-high): 2.5, 3.75, 4.5, 4.13/4.75
- *Computer science project (FIT3036)*, Bachelor of Computer Science (Honours), Monash University
 - Years taught: 2017
 - Main topics covered: *designing, developing and testing a computer science project*
 - Median student evaluation of teaching score (1-low to 5-high): TBA
- *Introduction to Data Science (FIT1043)*, Bachelor of Computer Science, Monash University
 - Years taught: 2016
 - Main topics covered: *roles, resources, technologies and methods involved in data science projects; data cleaning, analysis, management and storage*
 - Median student evaluation of teaching score (1-low to 5-high): 3.31
- *Intelligent Systems (FIT5047)*, Master of Information Technology, Monash University
 - Years taught: 2013, 2014
 - Main topics covered: *artificial intelligence, propositional and first order logic, planning, supervised and unsupervised machine learning, Bayesian networks*
 - Median student evaluation of teaching score (1-low to 5-high): 4.0, 4.05
- *Business Intelligence Modelling (FIT5097)*, Master of Information Technology, Monash University
 - Years taught: 2012, 2013
 - Main topics covered: *linear & non-linear programming, inventory modelling, queue models, decision theory, forecasting*
 - Median student evaluation of teaching score (1-low to 5-high): 4.21, 3.91
- *Advanced topics in algorithms and discrete structures (FIT4010)*, Honours in Computer Science, Monash
 - Years taught: 2011, 2012, 2013
 - Main topics covered: *propositional satisfiability solvers, non-linear programming, constrained optimisation, Markov chain Monte Carlo*
 - Median student evaluation of teaching score (1-low to 5-high): 4.25, 4.5
- *Object and semi-structured database (FIT5168)*, Master of Information Technology, Monash University
 - Years taught: 2011, 2012
 - Main topics covered: *semi-structured databases, XML, schema languages, transformation & query languages*
 - Median student evaluation of teaching score (1-low to 5-high): 4.12, 4.61
- *Programming in C*, Bachelor in Informatics, Università della Svizzera italiana
 - Years taught: 2009
 - Main topics covered: *basic C syntax and data structures, dynamic memory allocation*

4.2 Examples of Student Feedback

In 2013 Dr. Carman received teaching award nominations for both of FIT5047 and FIT5097, with anonymous comments such as the following for FIT5047:

Dr. Carman was always very lively during the lectures. He tried to keep the lectures interesting even though on occasions the lecture material was quite mathematical and dry. The problems discussed in class helped me relate to real-world scenarios and gave me a deeper insight into the topic of “intelligent systems”.

And for FIT5097:

Good use of “real world” examples and ability to draw connections between different topics in the course and good use of the whiteboard in lecture theatres. Detailed and timely responses to questions from students and helpful feedback over email when required. Broad interest in different subject material and ‘cross promotion’ of other available courses.

4.3 Courses Assisted

Prior to joining Monash, Dr. Carman assisted teaching of a number of courses at the University of Lugano and the University of Southern California:

2010: Lecturing: *Information and Knowledge Management II*, Bachelor degree, U. Lugano

2010: Lecturing: *Intelligent Information Access*, Masters degree, U. Lugano

2008: Project assistance: *Software Atelier IV*, Bachelor degree, U. Lugano

2008: Guest lecturer: *Intelligent Information Access*, Masters degree, U. Lugano

2008: Guest lecturer: *Information and Knowledge Management II*, Bachelor degree, U. Lugano

2006: Guest lecturer: *Information Integration on the Web*, Masters degree, U. Southern California

5 Professional Activities

Mark Carman is a highly active participant in the Information Retrieval and Data Mining communities. He regularly reviews for the premier international conferences (SIGIR, KDD, CIKM, IJCAI, AAAI, ACML, etc.). He also organises conferences, and is an associate editor for the most important journal in Information Retrieval, namely TOIS.

5.1 Editorial Duties

Jan 2016 – Present: **Associate Editor**

ACM Transactions on Information Systems (TOIS)

TOIS is considered the most prestigious journal in Information Retrieval.

5.2 Conference Organising

Dec 2016: **Conference Co-chair** and **Program Co-chair**

21st Australasian Document Computing Symposium (ADCS 2016),

Melbourne, Australia

2003: Local Arrangements Organiser

Third International Summer School on AI Planning,
Madonna di Campiglio, Trentino, Italy

5.3 Program Committees

Below is a list of some of the major international conferences for which Dr. Carman has been a program committee member. The list includes *the most important international conferences in Information Retrieval (namely SIGIR, WSDM, and CIKM), Data Mining (KDD, WWW and EMNLP) and Artificial Intelligence (IJCAI and AAAI)*, among many others. For brevity, local Australian conferences and workshops are not listed, nor are multiple tracks at the same conference.

- 2018: 27th International World Wide Web Conference (WWW 2018)
- 2017: 9th Asian Conference on Machine Learning (ACML 2017)
40th Int. ACM Conference on Research & Development in Information Retrieval (SIGIR'17-short)
10th International Conference on Web Search and Data Mining (WSDM 2017)
26th International World Wide Web Conference (WWW 2017)
- 2016: 8th Asian Conference on Machine Learning (ACML 2016)
25th ACM Conference on Information and Knowledge Management (CIKM'16)
22nd ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD'16)
39th Int. ACM Conference on Research & Development in Information Retrieval (SIGIR'16)
25th International Joint Conference on Artificial Intelligence (IJCAI'16)
38th European Conference on Information Retrieval (ECIR 2016)
- 2015: 7th Asian Conference on Machine Learning (ACML 2015)
24th ACM Conference on Information and Knowledge Management (CIKM'15)
21st ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD'15)
38th Int. ACM Conference on Research & Development in Information Retrieval (SIGIR'15)
International Conference on the Theory of Information Retrieval (ICTIR'15)
37th European Conference on Information Retrieval (ECIR 2015)
- 2014: 6th Asian Conference on Machine Learning (ACML 2014)
Conference on Empirical Methods on Natural Language Processing (EMNLP 2014)
37th Int. ACM Conference on Research & Development in Information Retrieval (SIGIR'14)
36th European Conference on Information Retrieval (ECIR 2014)
- 2013: 36th Int. ACM Conference on Research & Development in Information Retrieval (SIGIR'13)
35th European Conference on Information Retrieval (ECIR 2013)
- 2012: 26th Conference on Artificial Intelligence, AI and the Web track (AAAI-12 AIW)
35th Int. Conf. on Research & Development in Information Retrieval (SIGIR'12-Poster)
34th European Conference on Information Retrieval (ECIR 2012)
- 2011: 25th Conference on Artificial Intelligence, AI and the Web track (AAAI-11 AIW)
International Conference on the Theory of Information Retrieval (ICTIR'11)
20th ACM Conference on Information and Knowledge Management (CIKM'11-Poster)
5th International AAAI Conference on Weblogs and Social Media (ICWSM'11)

2010: 33rd Int. ACM Conference on Research & Development in Information Retrieval (SIGIR'10)
24th Conference on Artificial Intelligence, AI and the Web track (AAAI-10 AIW)
4th International AAAI Conference on Weblogs and Social Media (ICWSM'10)

2009: 32nd Int. ACM Conference on Research & Development in Information Retrieval (SIGIR'09)

2008: 22nd Conference on Artificial Intelligence, AI and the Web track (AAAI-08 AIW)

5.4 Journal Reviewing

Dr. Carman regularly reviews for a number of prestigious international journals, some of which are listed below:

- Machine Learning (ML): *2017*
- Artificial Intelligence (AI): *2017*
- Pattern Recognition (PR): *2016 to 2017*
- Communications in Statistics – Theory and Methods: *2017*
- IEEE Transactions on Knowledge and Data Engineering (TKDE): *2009 to 2017*
- ACM Transactions on the Information Systems (TOIS): *2015 to 2017*
- Information Retrieval (IR): *2016*
- Data Mining and Knowledge Discovery (DAMI): *2010 to 2015*
- Information Processing and Management (IP&M): *2007 to 2015*
- ACM Transactions on the Web (TWEB): *2015*
- Informs Journal on Computing (IJOC): *2014*
- Foundations and Trends in Information Retrieval (FnTIR): *2012*

5.5 Funding Proposal Reviewing

Dr. Carman has been asked to review funding proposal submitted to the Austrian Science Fund (FWF).

5.6 Invited talks

2016 *Investigating performance and scalability issues for rank learning with regression tree ensembles*, The Artificial Intelligence Research Institute (IIIA), Barcelona, Spain

2016 *Investigating performance and scalability issues for rank learning with regression tree ensembles*, Computer Science Colloquium, University of Groningen, Netherlands

2006 *Learning Semantic Descriptions of Online Information Sources*, Information Sciences Institute, University of Southern California, USA

2003 *Planning for Web Services*, Digital Enterprise Research Institute, University of Innsbruck, Austria

5.7 University Positions and Committees

Since joining Monash University in late 2010, Dr. Carman has been an active member of the academic community within the Faculty of IT. He has started reading groups, participated in and managed various committees, and directed undergraduate degree courses. Some of the roles that he has assumed, have included:

Jan 2016 - present: **Course Director** for the Bachelor of Computer Science Advanced (Honours), and the Bachelor of Informatics and Computation Advanced (Honours).
Tasks include: managing student enrolment, course progression issues including credit assignment, handling course accreditation process with accreditation bodies.

2016: **Member** of the Faculty Undergraduate Programs Committee.
Tasks included: monitoring and managing changes to undergraduate courses in the Faculty of IT.

2015: **Member** of the Honours Curriculum Review Committee.
Tasks included: reviewing entry requirements and assessment policies for Honours and Minor theses programs.

2011-2015: **Chair** of the Faculty of IT Early Career Researcher committee
Task included: representing the interests of early career researchers, managing faculty funds for early career researchers, running career building and research writing workshops.

2011-2015: **Member** of the Faculty Research Committee.
Tasks included: monitoring research output and research initiatives within the Faculty.

2013-2014: **Member** of two Interview Panels for new Lecturer/Senior Lecturer Positions.
Tasks included: interviewing and ranking of applicants.

2011-2014: **Coordinator** of the Faculty of IT Machine Learning Reading Group.
Task: Organising fortnightly meetings to promote interactions between researchers within the vibrant machine learning community of the faculty.

6 Technology Transfer & Press Coverage

6.1 Technology Transfer

Most of Dr. Carman's publications provide practical solutions to real world problems. Examples of the types of technology developed include methods for improving blog search, personalising Web search retrieval results, improving classification performance on large datasets, identifying Twitter users under the influence of alcohol, aiding police to accelerate digital investigations, etc..

In addition to disseminating results via publication, Dr. Carman works closely with industry wherever possible and also releases the code associated with his research. Some examples of the types of technology transfer he is involved in are:

2016 Crawl and analysis of the Dark Web, with knowledge and support of the Australian Federal Police and the Australian Justice Minister.

- 2015 Development of a prototype system for automated skill key-phrase extraction from job advertisements and its demonstration to Australia’s largest employment search website.
- 2015 Analysis of wifi logs from a large suburban shopping mall in order to extract commercially relevant knowledge on the aggregate behaviour of shoppers.
- 2014 Benchmarking commercial and open-source Named Entity Extraction systems for use by a qualitative data analysis software vendor in order to develop new product functionality.
- 2013 Release of software developed for fast estimation of large scale classifiers using the Weighted Naive Bayes model.
- 2011 Talk to industry leaders on behalf of the Monash Faculty of IT on the importance of Web personalisation technology.
- 2006 Release of software for service discovery and semantic characterisation developed during PhD thesis as open source.
- 2001 Report and prototype development with CERN for a Data Grid infrastructure.

6.2 Press Coverage

Dr. Carman’s work and/or expertise has been discussed in the press:

- **MIT Technology Review:** A paper co-authored by Mark Carman was discussed in the article “*How Vector Space Mathematics Helps Machines Spot Sarcasm*” (October 13, 2016).¹
- **New Scientist:** Mark Carman was quoted as a subject expert in the article “*AI reads your tweets and spots when you’re being sarcastic*” by Edd Gent (4 August 2016).²
- **The Indian Express:** A paper co-authored by Mark Carman was discussed in the article “*IIT-Bombay team creates program to detect drunk text message writers*” by Mihika Basu (21 August 2015).³

7 Refereed Publications

Below is a selection of refereed publications in reverse chronological order, divided into journal articles followed by conference papers, workshop papers and posters.

7.1 Journal Articles

- [1] Aditya Joshi, Pushpak Bhattacharyya, and Mark James Carman. “Automatic Sarcasm Detection: A Survey”. In: *ACM Computing Surveys* (TO APPEAR, Accepted 12 July 2017) (2017).
- [2] N. Zaidi, Geoffrey I Webb, M. Carman, F. Petitjean, W. Buntine, H. Hynes, and H. De Sterck. “Efficient Parameter Learning of Bayesian Network Classifiers”. In: *Machine Learning* (Online 26 January 2017) (2017), pp. 1–41.
- [3] Zahraa S. Abdallah, Mark Carman, and Gholamreza Haffari. “Multi-domain evaluation framework for named entity recognition tools”. In: *Computer Speech & Language* 43 (2017), pp. 34–55.

¹<https://www.technologyreview.com/s/602639/how-vector-space-mathematics-helps-machines-spot-sarcasm/>

²<https://www.newscientist.com/article/2100007-ai-reads-your-tweets-and-spots-when-youre-being-sarcastic/>

³<http://indianexpress.com/article/cities/mumbai/iit-bombay-team-creates-program-to-detect-drunk-text-message-writers/>

- [4] Ye Zhu, Mark James Carman, and Kai Ming Ting. “Density-ratio based clustering for discovering clusters with varying densities”. In: *Pattern Recognition* 60 (2016), pp. 983–997.
- [5] Nayyar A. Zaidi, Geoffrey I. Webb, Mark J. Carman, François Petitjean, and Jesús Cerquides. “ALRn: accelerated higher-order logistic regression”. In: *Machine Learning* 104.2 (2016), pp. 151–194.
- [6] Muhammad Ibrahim and Mark Carman. “Comparing Pointwise and Listwise Objective Functions for Random Forest based Learning-to-Rank”. In: *ACM Transactions on Information Systems (TOIS)* 34.4 (2016), 20:1–20:38.
- [7] Janis Dalins, Campbell Wilson, and Mark Carman. “Monte-Carlo Filesystem Search - A crawl strategy for digital forensics”. In: *Digital Investigation* 13 (2015), pp. 58–71.
- [8] Nayyar A. Zaidi, Jesús Cerquides, Mark J. Carman, and Geoffrey I. Webb. “Alleviating Naive Bayes Attribute Independence Assumption by Attribute Weighting”. In: *Journal of Machine Learning Research* 14 (2013), pp. 1947–1988.
- [9] Shima Gerani, Mark Carman, and Fabio Crestani. “Aggregation Methods for Proximity-Based Opinion Retrieval”. In: *ACM Transactions on Information Systems (TOIS)* 30.4 (Nov. 2012), 26:1–26:36.
- [10] Mostafa Keikha, Fabio Crestani, and Mark James Carman. “Employing document dependency in blog search”. In: *Journal of the American Society for Information Science and Technology (JASIST)* 63.2 (2012), pp. 354–365.
- [11] Mark Baillie, Mark James Carman, and Fabio Crestani. “A multi-collection latent topic model for federated search”. In: *Information Retrieval* 14.4 (2011), pp. 390–412.
- [12] Craig Knoblock, José Luis Ambite, Mark Carman, Matthew Michelson, Pedro Szekely, and Rattapoom Tuchinda. “Beyond the Elves: Making Intelligent Agents Intelligent”. In: *AI Magazine* 29.2 (2008).
- [13] Mark J. Carman and Craig A. Knoblock. “Learning Semantic Definitions of Online Information Sources”. In: *Journal of Artificial Intelligence Research (JAIR)* 30 (2007), pp. 1–50.

7.2 Conference Papers

- [14] Rohit Saluja, Devaraj Adiga, Parag Chaudhuri, Ganesh Ramakrishnan, and Mark Carman. “Error Detection and Corrections in Indic OCR using LSTMs”. In: *Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition (ICDAR 2017)*. 2017, (TO APPEAR).
- [15] Aditya Joshi, Samarth Agrawal, Pushpak Bhattacharyya, and Mark Carman. “Expect the unexpected: Harnessing Sentence Completion for Sarcasm Detection”. In: *Proceedings of the 15th International Conference of the Pacific Association for Computational Linguistics (PACLING 2017)*. 2017, (TO APPEAR).
- [16] Yuan Jin, Mark Carman, Dongwoo Kim, and Lexing Xie. “Leveraging Side Information to Improve Crowd-sourcing”. In: *Proceedings of the 5th AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2017)*. AAAI. 2017, (TO APPEAR).
- [17] Wern Han Lim, Mark James Carman, and Sze-Meng Jojo Wong. “Estimating Relative User Expertise for Content Quality Prediction on Reddit”. In: *Proceedings of the 28th ACM Conference on Hypertext and Social Media*. ACM. 2017, pp. 55–64.
- [18] Gholamreza Haffari, Tuan Dung Tran, and Mark Carman. “Efficient Benchmarking of NLP APIs using Multi-armed Bandits”. In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. Valencia, Spain: Association for Computational Linguistics, 2017, pp. 408–416.
- [19] Pengfei Li, Mark Sanderson, Mark Carman, and Falk Scholer. “On the Effectiveness of Query Weighting for Adapting Rank Learners to New Unlabelled Collections”. In: *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, CIKM’16*. ACM. 2016, pp. 1413–1422.

- [20] Ramakrishna B Bairi, Mark J Carman, and Ganesh Ramakrishnan. “Beyond Clustering: Sub-DAG Discovery for Categorising Documents”. In: *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, CIKM’16*. ACM, 2016, pp. 801–810.
- [21] Aditya Joshi, Vaibhav Tripathi, Pushpak Bhattacharyya, and Mark J Carman. “Harnessing Sequence Labeling for Sarcasm Detection in Dialogue from TV Series ‘Friends’”. In: *Proceedings of the 20th ACM SIGNLL Conference on Computational Natural Language Learning (CoNLL 2016)*. 2016, pp. 146–155.
- [22] Kai Ming Ting, Ye Zhu, Mark Carman, Yue Zhu, and Zhi-Hua Zhou. “Overcoming key weaknesses of Distance-based Neighbourhood Methods using a Data Dependent Dissimilarity Measure”. In: *Proceedings of the 22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2016)*. 2016, pp. 1205–1214.
- [23] Wern Han Lim, Mark James Carman, and Sze-Meng Jojo Wong. “Estimating Domain-Specific User Expertise for Answer Retrieval in Community Question-Answering Platforms”. In: *Proceedings of the 21st Australasian Document Computing Symposium*. ADCS ’16. ACM, 2016, pp. 33–40.
- [24] Aditya Joshi, Vaibhav Tripathi, Kevin Patel, Pushpak Bhattacharyya, and Mark Carman. “Are Word Embedding-based Features Useful for Sarcasm Detection?” In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP’16)*. Association for Computational Linguistics, 2016, pp. 1006–1011.
- [25] Aditya Joshi, Abhijit Mishra, Balamurali A. R, Pushpak Bhattacharyya, and Mark James Carman. “A Computational Approach to Automatic Prediction of Drunk-Texting”. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics ACL 2015, Volume 2: Short Papers*. 2015, pp. 604–608.
- [26] N.A. Zaidi, M.J. Carman, J. Cerquides, and G.I. Webb. “Naive-Bayes Inspired Effective Pre-Conditioner for Speeding-Up Logistic Regression”. In: *IEEE International Conference on Data Mining (ICDM)*. 2014, pp. 1097–1102.
- [27] Muhammad Ibrahim and Mark Carman. “Undersampling Techniques to Re-balance Training Data for Large Scale Learning-to-Rank”. In: *10th Asia Information Retrieval Societies Conference (AIRS2014)*. 2014, pp. 444–457.
- [28] Muhammad Ibrahim and Mark Carman. “Improving scalability and performance of random forest based learning-to-rank algorithms by aggressive subsampling”. In: *Proceedings of the 12th Australasian Data Mining Conference (AusDM)*. 2014, pp. 91–99.
- [29] Ilya Markov, Mark Carman, and Fabio Crestani. “Towards Risk-Aware Resource Selection”. In: *10th Asia Information Retrieval Societies Conference (AIRS2014)*. 2014, pp. 148–159.
- [30] Morgan Harvey, Fabio Crestani, and Mark James Carman. “Building user profiles from topic models for personalised search”. In: *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management, CIKM’13*. 2013, pp. 2309–2314.
- [31] Morgan Harvey, Mark James Carman, and David Elswailer. “Comparing Tweets and Tags for URLs”. In: *Advances in Information Retrieval, 34th European Conference on IR Research, (ECIR 2012)*. 2012, pp. 73–84.
- [32] Morgan Harvey, Mark James Carman, Ian Ruthven, and Fabio Crestani. “Bayesian latent variable models for collaborative item rating prediction”. In: *Proceedings of the 20th ACM Conference on Information and Knowledge Management, CIKM 2011*. 2011, pp. 699–708.
- [33] Morgan Harvey, Mark James Carman, and Ian Ruthven. “Improving Social Bookmark Search Using Personalised Latent Variable Language Models”. In: *Proceedings of the fourth ACM international conference on Web search and data mining*. WSDM ’11. Hong Kong, China, 2011, pp. 485–494.

- [34] Giacomo Inches, Mark James Carman, and Fabio Crestani. “Investigating the Statistical Properties of User-Generated Documents”. In: *9th International Conference on Flexible Query Answering Systems, (FQAS 2011)*. 2011, pp. 198–209.
- [35] Mark James Carman, Fabio Crestani, Morgan Harvey, and Mark Baillie. “Towards query log based personalization using topic models”. In: *Proceedings of the 19th ACM Conference on Information and Knowledge Management, (CIKM 2010)*. Toronto, 2010, pp. 1849–1852.
- [36] Morgan Harvey, Mark James Carman, and Ian Ruthven. “Ranking social bookmarks using topic models”. In: *Proceedings of the 19th ACM Conference on Information and Knowledge Management, (CIKM 2010)*. Toronto, 2010, pp. 1401–1404.
- [37] Shima Gerani, Mark Carman, and Fabio Crestani. “Proximity based Opinion Retrieval”. In: *33rd Annual International ACM SIGIR Conference on Research and Development on Information Retrieval (SIGIR 2010)*. Geneva, Switzerland, 2010, pp. 403–410.
- [38] Morgan Harvey, Mark Baillie, Ian Ruthven, and Mark Carman. “Tripartite Hidden Topic Models for Personalised Tag Suggestion”. In: *Advances in Information Retrieval, 32nd European Conference on IR Research (ECIR 2010)*. Milton Keynes, UK, 2010, pp. 432–443.
- [39] Mark J. Carman, Mark Baillie, Robert Gwadera, and Fabio Crestani. “A Statistical Comparison of Tag and Query Logs”. In: *32nd Annual International ACM SIGIR Conference on Research and Development on Information Retrieval (SIGIR 2009)*. Boston, USA, 2009, pp. 123–130.
- [40] Shima Gerani, Mark J. Carman, and Fabio Crestani. “Investigating Learning Approaches for Blog Post Opinion Retrieval”. In: *Advances in Information Retrieval, 31st European Conference on IR Research, ECIR 2009*. Toulouse, France, 2009, pp. 313–324.
- [41] Mark Baillie, Mark J. Carman, and Fabio Crestani. “A Topic-based Measure of Resource Description Quality for Distributed Information Retrieval”. In: *Advances in Information Retrieval, 31st European Conference on IR Research, ECIR 2009*. France, 2009, pp. 485–496.
- [42] Mark James Carman and Craig A. Knoblock. “Learning Semantic Descriptions of Web Information Sources”. In: *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI-07)*. Vol. 7. Hyderabad, India, 2007, pp. 2695–2700.
- [43] Mark James Carman, Floriano Zini, Luciano Serafini, and Kurt Stockinger. “Towards an Economy-Based Optimisation of File Access and Replication on a Data Grid”. In: *CCGRID '02: Proceedings of the 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid*. 2002, pp. 340–345.

7.3 Workshop Papers

- [44] Yuan Jin, Mark J Carman, and Lexing Xie. “A Little Competition Never Hurt Anyone’s Relevance Assessments”. In: *3rd International Workshop on Gamification for Information Retrieval (GamifIR 2016)*, Pisa, Italy. 2016.
- [45] Aditya Joshi, Prayas Jain, Pushpak Bhattacharyya, and Mark Carman. “‘Who would have thought of that!’: A Hierarchical Topic Model for Extraction of Sarcasm-prevalent Topics and Sarcasm Detection”. In: *Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics (ExProM)*. 2016, pp. 1–10.
- [46] Aditya Joshi, Pushpak Bhattacharyya, and Mark Carman. “Political Issue Extraction Model: A Novel Hierarchical Topic Model That Uses Tweets By Political And Non-Political Authors”. In: *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2016)*. Association for Computational Linguistics, 2016, pp. 82–90.

- [47] Anupam Khattri, Aditya Joshi, Pushpak Bhattacharyya, and Mark James Carman. “Your Sentiment Precedes You: Using an author’s historical tweets to predict sarcasm”. In: *6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA)*. 2015, p. 25.
- [48] Ramakrishna B. Bairi, Mark Carman, and Ganesh Ramakrishnan. “On the Evolution of Wikipedia: Dynamics of Categories and Articles”. In: *2015 ICWSM Workshop on Wikipedia, a Social Media: Research Challenges and Opportunities*. 2015.
- [49] Aditya Joshi, Anoop Kunchukuttan, Pushpak Bhattacharyya, and Mark James Carman. “SarcasmBot: An open-source sarcasm-generation module for chatbots”. In: *Fourth International Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM’15)*. 2015.
- [50] Cedric Mesnage and Mark Carman. “Piloted Search and Recommendation with Social Tag Cloud-Based Navigation”. In: *1st Workshop On Music Recommendation And Discovery (WOMRAD) at ACM RecSys*. Barcelona, Spain, 2010.
- [51] Mark J. Carman, Mark Baillie, and Fabio Crestani. “Tag Data and Personalized Information Retrieval”. In: *2008 ACM workshop on Search in social media (SSM 2008)*. Napa Valley, USA, 2008, pp. 27–34.
- [52] José Luis Ambite, Craig A. Knoblock, Kristina Lerman, Anon Plangprasopchok, Thomas Russ, Cenk Gazen, Steven Minton, and Mark Carman. “Exploiting Data Semantics to Discover, Extract, and Model Web Sources”. In: *First International Workshop on Semantic Aspects in Data Mining (SADM’08), Data Mining Workshops 2008*. 2008, pp. 771–779.
- [53] Mark James Carman and Craig A. Knoblock. “Inducing Source Descriptions for Automated Web Service Composition”. In: *Proceedings of the AAI 2005 Workshop on Exploring Planning and Scheduling for Web Services, Grid, and Autonomic Computing*. 2005.
- [54] Mark Carman, Luciano Serafini, and Paolo Traverso. “Web Service Composition as Planning”. In: *Proceedings of ICAPS’03 Workshop on Planning for Web Services*. Trento, Italy, 2003.
- [55] Mark Carman and Luciano Serafini. “Planning for Web Services the Hard Way”. In: *Workshop on Service Oriented Computing, International Symposium on Applications and the Internet (SAINT-2003)*. 2003.
- [56] M. Aiello, M. Papazoglou, J. Yang, M. Carman, M. Pistore, L. Serafini, and P. Traverso. “A request language for web-services based on planning and constraint satisfaction”. In: *VLDB workshop on Technologies for E-Services (TES)*. LNCS. 2002, pp. 76–85.

7.4 Demos, Posters & Doctoral Consortia

- [57] Aditya Joshi, Diptesh Kanojia, Pushpak Bhattacharyya, and Mark James Carman. “Sarcasm Suite: A Browser-Based Engine for Sarcasm Detection and Generation”. In: *Proceedings of the Thirty-First AAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*. 2017, pp. 5095–5096.
- [58] Aditya Joshi, Vaibhav Tripathi, Pushpak Bhattacharyya, Mark James Carman, Meghna Singh, Jaya Saraswati, and Rajita Shukla. “How Challenging is Sarcasm versus Irony Classification?: An Analysis From Human and Computational Perspectives”. In: *Australasian Language Technology Association (ALTA) Workshop 2016*, pp. 123–127.
- [59] Shima Gerani, Mostafa Keikha, Mark Carman, and Fabio Crestani. “Personal Blog Retrieval Using Opinion Features”. In: *Proceedings of the 33rd European Conference on Advances in Information Retrieval*. ECIR’11. 2011, pp. 747–750.
- [60] Giacomo Inches, Mark James Carman, and Fabio Crestani. “Statistics of Online User-Generated Short Documents”. In: *Advances in Information Retrieval, 32nd European Conference on IR Research (ECIR 2010)*. Milton Keynes, UK, 2010, pp. 649–652.

- [61] Mostafa Keikha, Mark Carman, and Fabio Crestani. “Blog Distillation using Random Walks”. In: *32nd Annual International ACM SIGIR Conference on Research and Development on Information Retrieval (SIGIR 2009)*. Boston, USA, 2009, pp. 638–639.
- [62] Mark J. Carman and Fabio Crestani. “Towards Personalized Distributed Information Retrieval”. In: *31st Annual International ACM SIGIR Conference on Research and Development on Information Retrieval (SIGIR 2008)*. Singapore, 2008, pp. 719–720.
- [63] Mark James Carman. “Learning Source Descriptions for Web Services”. In: *The Tenth AAAI/SIGART Doctoral Consortium*. 2005, pp. 1640–1641.
- [64] Mark James Carman. “Planning for Web Services”. In: *Proceedings of the ICAPS’03 Doctoral Consortium*. Trento, Italy, 2003, p. 5.