Denilson Barbosa

Vita – November 2026

+1 (780) 492 2285
denilson@ualberta.ca
www.ualberta.ca/~denilson
denilsonbarbosa

Personal

Education

2005/03 **PhD in Computer Science**, *University of Toronto*

1999/05 **MSc in Computer Science**, *Universidade Federal de Minas Gerais, Brazil*

1996/12 **BSc in Computer Science**, *Universidade Federal de Pernambuco, Brazil*

Employment History

2022/07 – present **Professor**, *Department of Computing Science*, University of Alberta

2011/07 – 2022/06 **Associate Professor**, *Department of Computing Science*, University of Alberta

2014/07 – 2015/04 **Visiting Scientist**, *Max Planck Institute of Informatics*, Saarbrücken, Germany

2008/07 – 2011/06 **Assistant Professor/Ingenuity New Faculty**, *Department of Computing Science*, University of Alberta

2008/05 – 2008/07 **Visiting Professor**, *Faculty of Computer Science, Free University of Bozen-Bolzano*, Italy

2005/01 – 2008/06 **Assistant Professor**, *Department of Computer Science*, University of Calgary

Honors and Awards

2019 Outstanding PC member; 2019 WSDM Conference (ACM)

2013 Best Reviews of 2013 Award; ACM Computing Reviews

2010 Distinguished Service Award in Teaching; Department of Computing Science, University of Alberta

2010 **Best Paper Award**—2010 IEEE International Conference on Data Engineering

2006 Alberta Ingenuity Fund New Faculty Award

Fully Refereed Publications

Trainees under my formal supervision are highlighted in bold.

Journal Articles

CEAI'26 **J. G. Mathew**, **S. Taher**, **A. Kundu**, and D. Barbosa. LLMs do not grade essays like humans. *Computers and Education: Artificial Intelligence* 11, 100666 (2026). https://doi.org/10.1016/j.caeai.2026.100666

ESE'24 H. Sahar, A.A. Bangash, A. Hindle, and D. Barbosa. IRJIT: A simple, online, information retrieval approach for just-in-time software defect prediction. *Empir Software Eng* 29, 131 (2024). https://doi.org/10.1007/s10664-024-10514-z

WWWJ'22 **V. Ranganathan**, and D. Barbosa. HOPLoP: Multi-hop Link Prediction over Knowledge Graph Embeddings. *World Wide Web* 25 (2), 1037–1065 (2022), Springer Science + Business Media.

JATT'22 T. Firoozi, O. Bulut, C. Demmans Epp, and **A. Naeimabadi**, D. Barbosa. The Effect of Fine-tuned Word Embedding Techniques on the Accuracy of Automated Essay Scoring Systems Using Neural Networks. *Applied Testing Technology* 23 (Special Issue 1) 21–29, 2022.)

TKDD'21 A. Rossi, D. Barbosa, D. Firmani, A. Matinata, and P. Merialdo. Knowledge graph embedding for link prediction: a comparative analysis. *ACM Trans. Knowl. Discov. Data*, Article 14 (49 pp.), April 2021.

SemWeb'17 **Z. Guo**, D. Barbosa. Robust Named Entity Disambiguation with Random Walks. *Semantic Web*. Preprint, pp. 1-21, 2017.170273.

TIST'15 **H. Sepehri-Rad**, D. Barbosa. Identifying Controversial Wikipedia Articles using Editor Collaboration Networks. *ACM Trans. Inf. Sys. and Tech.* 6(1): 5:1-5:24, 2015.

TKDE'13 **M. Moraes**, C. Heuser, V. Moreira, D. Barbosa. Pre-Query Discovery of Domain-specific Query Forms: A Survey. *IEEE Trans. Data and Knowledge Engineering*, 25(8); pp. 1830-1848, 2013.

TWEB'12 **Y. Merhav**, **F. Mesquita**, D. Barbosa, W. G. Yee, O. Frieder. Extracting Information Networks from the Blogosphere. *ACM Transactions on the Web*, 6(3): 33 pp., 2012.

TKDE'11 N. Augsten, D. Barbosa, M. Böhlen, and T. Palpanas. Efficient Top-k Approximate Subtree Matching in Small Memory. *IEEE Trans. Data and Knowledge Engineering*, 23 (August 2011): 1123-1137, 2011.

TISSEC'11 **G. Leighton** and D. Barbosa. Access Control Policy Translation, Verification, and Minimization within Heterogeneous Data Federations. *ACM Trans. Inf. Syst. Secur.* 14(3): 27 pp., 2011.

SP&E'06 D. Barbosa, and A. O. Mendelzon. Declarative Generation of Synthetic XML Data. *Software: Practice and Experience* 36(10), pp. 1051-1079, 2006. John Wiley & Sons.

WWWJ'05' D. Barbosa, L, Mignet, and P. Veltri. Studying the XML Web: Gathering Statistics from an XML Sample. *World Wide Web* 8(4), pp. 413-438, December 2005, Springer Science + Business Media.

In Proceedings

ACL'24 **P. Yao** and Denilson Barbosa. Accurate and Nuanced Open-QA Evaluation Through Textual Entailment. In *Findings of the Association for Computational Linguistics*: ACL 2024, pages 2575–2587. Association for Computational Linguistics.

Neurips'24 **P. Yao**, **J. G. Matthew**, **S. Singh**, D. Firmani, and D. Barbosa. A Bayesian Approach Towards Crowdsourcing the Truths from LLMs. In Neurips workshop on Bayesian Decision-making and Uncertainty: from probabilistic and spatiotemporal modeling to sequential experiment design. 2024

NAACL'24 **N. Hervieux**, **P. Yao**, S. Brown, and D. Barbosa. Language Resources From Prominent Born-Digital Humanities Texts are Still Needed in the Age of LLMs. In *4th International Conference on Natural Language Processing for Digital Humanities*, Association for Computational Linguistics. 2024.

COLING'24 **A. Shahriar**, and D. Barbosa. Improving Bengali and Hindi Large Language Models. In *Joint International Conference on Computational Linguistics, Language Resources and Evaluation* (LREC-COLING 2024). ELRA and ICCL. 2024

MDM'24 **H. S. Mohammed**, M. A. Nascimento and D. Barbosa. Effective Trajectory Imputation using Simple Probabilistic Language Models. 25th IEEE International Conference on Mobile Data Management (MDM), doi: 10.1109/MDM61037.2024.00027. 2024

ACLW'24 **P. Yao**, **K. Guzhva**, and D. Barbosa. Semantic Graphs for Syntactic Simplification: A Revisit from the Age of LLM. *TextGraphs-17: Graph-based Methods for Natural Language Processing*. Association for Computational Linguistics. 2024

UNLP'24 **T. Ustyianovych**, and D. Barbosa. Instant Messaging Platforms News Multi-Task Classification for Stance, Sentiment, and Discrimination Detection. In *Third Ukrainian Natural Language Processing Workshop* (UNLP) @ LREC-COLING 2024 ELRA and ICCL. 2024

EACL'23 **P. Yao**, **M. Kosmajac**, **A. Waheed**, **K. Guzhva**, **N. Hervieux**, and D. Barbosa. NLP Workbench: Efficient and Extensible Integration of State-of-the-art Text Mining Tools. In *Proc. of the 17th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*. 2023.

EMNLP'22 **P. Yao**, **T. Renwick**, and D. Barbosa. WordTies: Measuring Word Associations in Language Models via Constrained Sampling. In *Findings of the Association for Computational Linguistics: EMNLP 2022*. 2022.

ACL'22 D. Teodorescu, J. Matalski, D. Lothian, and D. Barbosa, C. Demmans Epp. Cree Corpus: A Collection of nêhiyawêwin Resources. In *Proc. of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2022.

EGOS'22 L. M. Lefsrud, C. G. Gutierrez, E. Stroulia, and J. Gehman, D. Barbosa. Energy to contest? Emotional and multimodal contestation of energy markets. In *Proc. of the 38th EGOS Colloquium*. 2022.

DH'22 G. Rockwell, **N. Hervieux**, H. Zafar, and K. Land, A. MacDonald, D. Barbosa, L. Frizzera, M. Ilovan, S. Brown. Web Services for Voyant: LINCS, Voyant and NSSI. In *Proc. of DIGITAL HUMANITIES 2022*. 2022.

WWW'21 **P. Yao**, D. Barbosa. Typing Errors in Factual Knowledge Graphs: Severity and Possible Ways Out. In *WWW'21: The Web Conference 2021*. 8 pp. ACM, 2021.

CAI'21 **T. Renwick**, D. Barbosa. Detection and Identification of Obfuscated Obscene Language with Character Level Transformers. In *32nd Canadian Conference on Artificial Intelligence*. 8 pp. 2021.

CASCON'21 C. G. Gutierrez, **A. Whittaker**, K. Patenio, L. Lefsrud, J. Gehman, Joel, D. Barbosa, and E. Stroulia. Analyzing and Visualizing Twitter Conversations. In *CASCON X EVOKE 2021*. 8pp. 2021.

CIKM'20 **E. Macdonald** and D. Barbosa. Neural relation extraction on Wikipedia tables for augmenting knowledge graphs. In 29th ACM Int. Conf. on Information & Knowledge Management, pp. 2133–2136. ACM, 2020.

SAC'20 **V. Ranganathan**, S. Suresh, Y. Mathur, N. Subramanyam, and D. Barbosa. Grcluster: a score function to model hierarchy in knowledge graph embeddings. In *SAC '20: The 35th ACM/SIGAPP Symposium on Applied Computing*, pp. 964–971. ACM, 2020.

ICSE'20 **C. Pang**, A. Hindle, and D. Barbosa. Understanding DevOps education with grounded theory.. In *Proc. of the 41st Int. Conf. on Software Engineering: Software Engineering Education and Training.*, pp. 8. IEEE, 2020.

ECNLPIR'20 R. Chahardoli, D. Rafiei, D. Barbosa. Relation Extraction With Synthetic Explanations And Neural Networks. In *2020 European Conference on Natural Language Processing and Information Retrieval*. 4 pp. 2020.

EMNLP'19 F. Mesquita, M. Cannaviccio, J. Schmidek, P. Mirza, and D. Barbosa. KnowledgeNet: a benchmark dataset for knowledge base population. In *Proc. of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp 749–758. ACL, 2019.

NAACL'19 **P. Xu** and D. Barbosa. Connecting language and knowledge with heterogeneous representations for neural relation extraction. In *Proc. of the 2019 Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp 3201–3206. ACL, 2019.

KDD'19 workshop **D. Caminhas**, **D. Cones**, **N. Hervieux**, and D. Barbosa. Detecting and correcting typing errors in DBpedia. In *1st Int. Workshop on Challenges and Experiences from Data Integration to Knowledge Graphs co-located with SIGKDD*. Volume 2512 of CEUR Workshop Proc.. CEUR-WS.org, 2019.

WWW'18 **M. Cannaviccio**, D. Barbosa, and P. Merialdo. Towards Annotating Relational Data on the Web with Language Models. In *Proc. of the 2018 World Wide Web Conference*. pp 1307-1316. ACM, 2018.

NAACL'18 **P. Xu**, and D. Barbosa. Neural Fine-Grained Entity Type Classification with Hierarchy-Aware Loss. In *Proc. of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics*. 10 pp. ACL, 2018.

WebDB'18 **M. Cannaviccio**, L. Ariemma, D. Barbosa, and P. Merialdo. Leveraging Wikipedia Table Schemas for Knowledge Graph Augmentation. In *Proc. of the 21st Int. Workshop on Web and Databases*. 9 pp. ACM, 2018.

CAI'18 **P. Xu**, and D. Barbosa. Matching Résumés to Job Descriptions with Stacked Models. In Proc. of the 31st Canadian Conference on Artificial Intelligence. 6 pp. Springer, 2018.

DBPedia workshop'17 **M. Cannaviccio**, D. Barbosa, and P. Merialdo. Lector+ Entry for the TextExt - DBpedia Open Extraction Challenge. In *CEUR-WS.org - CEUR Workshop Proc.*, 2017.

WSDM'16 M. Yahya, D. Barbosa, K. Berberich, Q. Wang, and G. Weikum. Relationship Queries on Extended Knowledge Graphs. In *Proc. of the 9th ACM Int. Conf. on Web Search and Data Mining*. ACM, 2016.

WebDB'16 **M. Cannaviccio**, D. Barbosa, and P. Merialdo. Accurate fact harvesting from natural language text in wikipedia with Lector. In *Proc. of the 19th Int. Workshop on Web and Databases*, San Francisco, CA, USA, June 26, 2016, 9. ACM, 2016.

SAC'16 **I. Polato**, D. Barbosa, A. Hindle, and F. Kon. Hadoop energy consumption reduction with hybrid HDFS. In *Proc. of the 31st Annual ACM Symposium on Applied Computing*, Pisa, Italy, April 4-8, 2016, 406–411. ACM, 2016.

ICDE'15 D. Barbosa, H. Wang, and C. Yu. Inferencing in information extraction: techniques and applications. In *Proc. of the IEEE 31st Int. Conf. on Data Engineering*. IEEE, 2015.

GSCC'15 **I. Polato**, D. Barbosa, A. Hindle, and F. Kon. Hadoop Branching: Architectural Impacts on Energy and Performance. In *6th Int. Green and Sustainable Computing Conference*, 2015, 1–4. IEEE Computer Society, 2015.

WWW'15 workshop **Z. Guo** and D. Barbosa. Entity recognition and linking on tweets with random walks. In *Proc. of the the 5th workshop. on making sense of microposts co-located w/ 24th WWW Conf.*, pp. 57–58. ACM, 2015.

ASONAM'14 J. Fagnan, O. R. Zaiane, and D. Barbosa. Using triads to identify local community structure in social networks. In *2014 IEEE/ACM Conf. on Adv. in Soc. Net. Analysis and Mining*, pp. 108–112, 2014.

CIKM'14 **Z. Guo** and D. Barbosa. Robust Entity Linking via Random Walks. In *Proc. of the 23rd ACM Int. Conf. on Information and Knowledge Management*, pp. 499-508. ACM, 2014.

WWW'14 workshop **Y. A. Sekhavat**, **F. di Paolo**, D. Barbosa and P. Merialdo. Knowledge Base Augmentation using Tabular Data. In 7th Workshop on Linked Data on the Web. CEUR-WS.org, 2014.

WWW'14 workshop **Z. Guo** and D. Barbosa. Entity linking with a unified semantic representation. In *23rd International World Wide Web Conference*, Companion Volume, pp. 1305-1310. ACM, 2014

LREC'14 **J. Schmidek**, D. Barbosa. Improving Open Relation Extraction via Sentence Re-Structuring. In *Proc. of the Ninth Int. Conf. on Language Resources and Evaluation*. European Language Resources Association (ELRA), Reykjavik, Iceland, 2014.

ACL'13 **H. He**, D. Barbosa, and G. Kondrak. Identification of Speakers in Novels. In *Proc. of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1312-1320. Association for Computational Linguistics, 2013.

NAACL'13 Y. Xu, M.-Y. Kim, K. Quinn, R. Goebel, and D. Barbosa. Open Information Extraction with Tree Kernels. In *Proc. of the 2013 Conf. of the N. American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 12 pp. ACL, 2013.

EMNLP'13 **F. Mesquita**, **J. Schmidek**, D. Barbosa. Effectiveness and Efficiency of Open Relation Extraction. In *EMNLP 2013: Conference on Empirical Methods in Natural Language Processing*. pp 447–457. ACL, 2013.

ICDE'13 D. Barbosa, H. Wang, and C. Yu. Shallow information extraction for the knowledge web. In IEEE 29th Int. Conf. on Data Engineering, 1264 – 1267. IEEE, 2013.

WISE'13 **M. Moraes**, C. A. Heuser, V. Moraes, D. Barbosa. Automatically Training Form Classifiers. In *Proc. of the 14th Int. Conf. on Web Information System Engineering*. 14 pp. Springer, 2013.

WikiSym'12 **H. Sepehri-Rad**, and D. Barbosa. Identifying controversial articles in Wikipedia: A comparative study. In *7th International Symposium on Wikis and Open Collaboration*, 10 pp.; ACM, 2012.

HT'12 **H. Sepehri-Rad**, A. Makazhanov, D. Rafiei and D. Barbosa. Leveraging editor collaboration patterns in wikipedia. In *Proc. of the 23rd ACM conference on Hypertext and social media*, pp. 13-22. ACM, 2012.

EACL'12 workshop **S. Husby** and D. Barbosa. Topic Classification of Blog Posts Using Distant Supervision. In *Proc. of the Workshop on Semantic Analysis in Social Media*, pp. 28-36. Association for Computational Linguistics, 2012.

NAACL'12 workshop **M. Bronzi**, **Z. Guo**, **F. Mesquita**, D. Barbosa and P. Merialdo. Automatic Evaluation of Relation Extraction Systems on Large-scale. In *Proc. of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction*, pp. 19-24. Association for Computational Linguistics, 2012.

SIGMOD'12 **T. Sliwkanich**, **D. Schneider**, **A. Yong**, **M. Home**, and D. Barbosa. Towards scalable summarization and visualization of large text corpora (abstract only). In *Proc. of the ACM SIGMOD Int. Conf. on Management of Data*, pp. 863, 2012. **(Best Undergraduate Poster Award)**

ICWSM'11 **F. Mesquita** and D. Barbosa. Extracting Meta Statements from the Blogosphere. In *Proc. of the Fifth International AAAI Conference on Weblogs and Social Media*, 2011.

DEXA'11 C. E. Pires, P. Vieira, M. Saraiva and D. Barbosa. Generating Synthetic Database Schemas for Simulation Purposes. In *Proc. of the 22nd Int. Conf. on Database and Expert Systems Applications*, Springer-Verlag, 2011.

WWW'11 **H. Sepehri-Rad** and D. Barbosa. Towards identifying arguments in Wikipedia pp.. In *Proc. of the 20th Int. Conf. on World Wide Web*, pp. 117-118, ACM, 2011.

ICDE'10 N. Augsten, D. Barbosa, M. Böhlen, and T. Palpanas. TASM: Top-k Approximate Subtree Matching. In *26th Int. Conf. on Data Engineering* (ICDE); pp. 353-364, IEEE Computer Society, 2010 **Best Paper Award**).

SACMAT'10 **G. Leighton** and D. Barbosa. Access Control Policy Translation and Verification within Heterogeneous Data Federations. In *Proc. of the 15th ACM Symposium on Access Control Models and Technologies*, pp. 173–182, ACM, 2010

NIPS'10 workshop A. Celikyilmaz, D. Hakkani-Tur, **H. He**, G. Kondrak and D. Barbosa. The Actor-Topic Model for Extracting Social Networks in Literary Narrative. In *Proc. of the 2010 Neural Information Processing Systems in Machine Learning for Social Computing Workshop* (NIPS 2010 - MLSC), 7 pp.. December 2010.

ICWSM'10 **F. Mesquita**, **Y. Merhav** and D. Barbosa. Extracting Information Networks from the Blogosphere: State-of-the-Art and Challenges. In *4th Int'l AAAI Conference on Weblogs and Social Media–Data Challenge Track*, 8 pp., AAAI, 2010.

DESWeb'10 **P. Cappellari**, D. Barbosa and P. Atzeni. A Framework for Automatic Schema Mapping Verification Through Reasoning. In Proc. of the *2010 Int. Workshop on Data Engineering meets the Semantic Web* (with ICDE), pp. 245-250, IEEE Computer Society, 2010.

ICSM'10 **M. Hasan**, E. Stroulia, D. Barbosa, M. Alalfi. Analyzing Natural-Language Artifacts of the Software Process. In *26th IEEE Int. Conf. on Software Maintenance* (Early Results Track), 5 pp., IEEE Computer Society, 2010.

SIGIR'10 **Y. Merhav**, **F. Mesquita**, D. Barbosa, W. G. Yee, O. Frieder. Incorporating Global Information into Named Entity Recognition Systems using Relational Context. In *33rd annual ACM SIGIR conference*, 2 pp., ACM, 2010.

CIKM'10 **V. Ganev**, **Z. Guo**, **D. Serrano**, B. Tansey, D. Barbosa, E. Stroulia. Exploring and Visualizing Academic Social Networks. In *19th ACM Int. Conf. on Information and Knowledge Management* (CIKM), 2 pp., ACM, 2010.

MEDES'09 **V. Ganev**, **Z. Guo**, **D. Serrano**, B. Tansey, D. Barbosa, E. Stroulia. An environment for building, exploring and querying academic social networks. In *International ACM Conference on Management of Emergent Digital EcoSystems*, pp. 282-289, ACM, 2009.

XSym'09 **G. Leighton**, D. Barbosa. Optimizing XML Compression. In *Database and XML Technologies, 6th International XML Database Symposium*, pp. 91-105, Springer-Verlag, 2009.

LID'09 **P. Cappellari** and D. Barbosa. Towards Automatic Schema Mapping Verification Through Reasoning. In *Int. Workshop on Logic in Databases*, pp. 43-56, Rotschild University, 2009.

ODBASE'07 A. S. da Silva, D. Barbosa, J. M. B. Cavalcanti, and **M. Sevalho**; Labeling Data Extracted from the Web. In *Proc. of the 6th Int. Conf. on Ontologies, DataBases, and Applications of Semantics*; pp. 1099-1116, Springer, 2007.

WIDM'07 **F. Mesquita**, D. Barbosa, **E. Cortez**, and Altigran S. da Silva; FleDEx: Flexible Data Exchange. In *Proc. of the 9th ACM Int. Workshop on Web Information and Data Management*; pp. 25-32, ACM, 2007.

WWW'07 **J. Park**, and D. Barbosa. Adaptive Record Extraction From Web pp.. In *Proc. of the 16th Int. Conf. on World Wide Web* (WWW), pp. 1335-1336, ACM, 2007.

XSym'06 D. Barbosa, **G. Leighton**, and **A. Smith**. Efficient Incremental Validation of XML Documents After Composite Updates. In *Proc. of the Fourth International XML Database Symposium*, pp. 107-121. Springer-Verlag, 2006.

VLDB'05 D. Barbosa, J. Freire, and A. O. Mendelzon. Designing Information-Preserving Mapping Schemes for XML. In *Proc. of the 31st Int. Conf. on Very Large Databases* (VLDB); pp. 109-102. ACM, 2005.

SIGMOD'05 M. P. Consens, D. Barbosa, **A. M. Teisanu**, and L. Mignet. Goals and Benchmarks for Autonomic Configuration Recommenders. In *Proc. of the 2005 ACM SIGMOD Int. Conf. on Management of Data* (SIGMOD); pp. 239-250. ACM, 2005.

ICDE'04 D. Barbosa, A. O. Mendelzon, L. Libkin, L. Mignet, and M. Arenas. Efficient Incremental Validation of XML Documents. In *Proc. of the 20th Int. Conf. on Data Engineering* (ICDE); pp. 671-682. IEEE Computer Society, 2004,

XSym'04 D. Barbosa, J. Freire, and A. O. Mendelzon. Information Preservation in XML-to-Relational Mappings. *Lecture Notes in Computer Science*, vol 3186 (Proc. of the Second International XML Database Symposium), pp. 66-81. Springer-Verlag, 2004.

XSym'03 D. Barbosa and A. O. Mendelzon. Finding ID Attributes in XML Documents. *Lecture Notes in Computer Science*, vol 2824 (Proc. of the First International XML Database Symposium, Berlin, Germany), pp. 180-194. Springer-Verlag, 2003.

SEDB'03 L. Mignet, D. Barbosa and P. Veltri. On the Characterization of the XML Web (extended abstract). In *Proc. of the Eleventh Italian Symposium on Advanced Database Systems*. Cetraro, Italy, 2003.

WWW'03 L. Mignet, D. Barbosa and P. Veltri. The XML web: a first study. In *Proc. of the 12th int. conf. on World Wide Web* (WWW); pp. 500-510. ACM, 2003.

SIGMOD'02 D. Barbosa, A. O. Mendelzon, J. Keenleyside and K. Lyons. ToXgene: a template-based data generator for XML. In *Proc. of the 2002 ACM SIGMOD Int. Conf. on Management of Data*, pp. 616, 2002, ACM.

WebDB'02 D. Barbosa, A. O. Mendelzon, J. Keenleyside and K. Lyons. ToXgene: a template-based data generator for XML. In *Proc. of the Fifth Int. Workshop on the Web and Databases*; pp. 49-54. Madison, WI, USA, 2002.

WIIW'01 D. Barbosa, A. Barta, A. O. Mendelzon, G. Mihaila, F. Rizzolo and P. Rodriguez-Gianolli. ToX - the Toronto XML Engine. In *Proc. of the Workshop on Information Integration on the Web*. Rio de Janeiro, Brazil, 2001.

SIBGRAPI'99 D. Barbosa, J. P. Kitajima and W. Meira Jr. Parallelizing MPEG Video Encoding using Multiprocessors. In *Proc. of the XII Brazilian Symposium on Computer Graphics and Image Processing*. Campinas, SP, Brazil. IEEE Computer Society, 1999.

ICPCS'99 D. Barbosa, J. P. Kitajima and W. Meira Jr. Real-time MPEG Encoding in Shared-Memory Multiprocessors. In *Proc. of the 2nd Int. Conf. on Parallel Computing Systems*. Ensenada, Baja California, Mexico, 1999.

VECPAR'98 A. Macêdo, E. S. Silva, D. Barbosa, M. A. Cristo, J. P. Kitajima, B. Ribeiro-Neto, G. Navarro and N. Ziviani. Experimental Analysis of a Parallel Quicksort-based Algorithm for Suffix Array Generation. In *Proc. of the 3rd International Meeting on Vector and Parallel Processing*. Porto, Portugal, 1998.

Other Publications

Book

2013 R. T. Ng, P. C. Arocena, D. Barbosa, G. Carenini, L. Gomes-Jr, S. Jou, R. A. Leung, E. Milios, R. J. Miller, J. Mylopoulos, R. A. Pottinger, F. Tompa, E. Yu. *Perspectives on Business Intelligence*. Synthesis Lectures; Morgan and Claypool, 2013. DOI:10.2200/S00491ED1V01Y201303DTM034.

Book Chapter

RSO'26 L. M. Lefsrud, H. Ali, C. A. Gutierrez Gutierrez, J. Gehman, D. Barbosa, and E. Stroulia. Energy to Contest? Emotional and Multimodal Contestation of Institutional Infrastructure. In *Beyond Microfoundations and Macrofoundations of Institutions: A Multi-Level Theory of Institutions* (D. Harmon, P. Haack, T. R. Hannigan, and H. Etchanchu, eds.). Emerald Publishing Limited, 2026. https://doi.org/10.1108/S0733-558X20260000104006

ANAA'13 **D. Serrano**, E. Stroulia, D. Barbosa, and V. Guana. SociQL: A Query Language for the Social Web. In *Advances in Network Analysis and its Applications* (E. Kranakis, ed.). 30 pp. Mathematics in Industry; Springer Verlag. 2013.

Lightly Refereed and/or Invited

2026 C. Kwan, V. Sati, N. Fwangje, X. Fan, D. Barbosa, and R. Goebel. BJPWH at COLIEE 2026 Task 1: Combining BM25 and Dense Retrieval with Year-Based Filtering for Legal Case Retrieval. In *Proc. of the Workshop on the Thirteenth International Competition on Legal Information Extraction and Entailment* (COLIEE 2026), 2026.

2009 D. Barbosa, P. Bohannon, J. Freire, C.-C. Kanne, I. Manolescu, V. Vassalos, M. Yoshikawa. XML Storage. In *Encyclopedia of Database Systems*, pages 3627-3634, Springer-Verlag, 2009.

2009 D. Barbosa, I. Manolescu, J. X. Yu. XML Benchmarks. In *Encyclopedia of Database Systems*, pages 3576-3579, Springer-Verlag, 2009.

2009 D. Barbosa, I. Manolescu, J. X. Yu. Application Benchmark. In *Encyclopedia of Database Systems*, pages 99-100, Springer-Verlag, 2009.

2009 D. Barbosa, I. Manolescu, J. X. Yu. Microbenchmark. In *Encyclopedia of Database Systems*, page 1737, Springer-Verlag, 2009.

2007 D.Barbosa, **G. Elliott**; Generating Synthetic XML Documents by Example. Demo at the *2007 Conference of the IBM Center for Advanced Studies on Collaborative research*, 2007.

Technical Reports

2009 **G. Leighton**, D. Barbosa. Optimizing XML Compression. CoRR abs/0905.4761, 2009.

2007 **F. Mesquita**, D.Barbosa, **E. Cortez**, and A. S. da Silva. A Lightweight Framework for Exchanging Web Data. Technical Report 2007-876-28, University of Calgary; July, 2007.

2004 D.Barbosa, M. P. Consens, L. Mignet. Experimental Evaluation of Autonomic Indexing. Technical Report CSRG-495, University of Toronto, July 2004.

2004 D.Barbosa and A. O. Mendelzon. CASCON 2003 Workshop Report–XML Query Processing. IBM Technical Report TR-74.193-21, January 2004.

2003 D.Barbosa. ToXgene Template Specification Language. Technical Report CRG-474. University of Toronto, July 2003.

2003 D.Barbosa and A. Mendelzon. CASCON 2002 Workshop Report–XML Data Management. IBM Technical Report TR-74.188-14, January 2003.

Theses

2005 Incremental Validation of XML Documents and Mappings. Ph.D. Thesis, Department of Computer Science, University of Toronto, 2005, 149 pages, ISBN:0-494-02722-3.

1999 Compressão de Vídeo Digital MPEG em Paralelo. M.Sc. Thesis (in Portuguese), Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, 1999.

Software and Other Scholarly Works

NAACL'24 **Orlando Information Extraction Dataset**: language resources from prominent born-digital humanities texts, released with our NLP4DH 2024 paper. With **N. Hervieux**, **P. Yao**, and S. Brown. Borealis, 2024. `https://doi.org/10.5683/SP3/RCVANO`

EACL'23 **NLP Workbench**: Efficient and Extensible Integration of State-of-the-art Text Mining Tools. A web-based platform for text mining that allows non-expert users to obtain semantic understanding of large-scale corpora using state-of-the-art models, including entity linking, sentiment analysis, semantic parsing, and relation extraction. Its extensible design lets researchers and developers replace an existing model or integrate a new one, and its microservice architecture facilitates allocation of acceleration hardware and parallelization of computation. Released under the MIT license.
`https://github.com/U-Alberta/NLPWorkbench`

WedBD'16 **Lector+** Lector is an extraction tool, jointly developed with colleagues at Roma Tre University, able to extract facts from English Wikipedia article text, learning phrases that are commonly used to describe relationships between named entities in the text. It reaches an estimated precision of 95%.
`https://github.com/miccia4/LectorPlus`

CIKM'14 **The Walking NED** (WNED) is a named entity disambiguation systems based on random walks. It remains one the most accurate named entity disambiguation tools out there and the best pre-neural networks, being cited often as a baseline.
`https://github.com/U-Alberta/wned`

EMNLP'13 **EXEMPLAR** is an open relation extraction system that supports binary and n-ary relations. It was used by researchers at Stanford University as the basis for a PhD thesis leading to improvements on their own NLP pipeline. `https://github.com/U-Alberta/exemplar`

MEDES'09 **ReaSoN**: a portal for visualizing, exploring, querying, and integrating academic social networks, extracted from bibliographic and citation databases. It had over 350K publications, 1.2M citations, and 379K authors. `http://hypatia.cs.ualberta.ca/reason/`

SP&E SIGMOD'02 WebDB'02 ToXGene is a declarative, template-based tool for the generation of arbitrarily large collections of complex, semantically-correlated synthetic XML documents. ToxGene has had over 3800 downloads and has been used in over 50 universities and over 200 companies and research labs around the world (as of 2005, when I stopped tracking). It is also the data generator of **XBench** (Yao et al., ICDE 2004), the most comprehensive XML benchmark to date developed at the University of Waterloo; **TPoX** V1.0 (Nicola et al., SIGMOD 2008), an industry-standard benchmark developed by IBM (`http://tpox.sourceforge.net/`); and the **MemBeR** (`http://tilcara.science.uva.nl:8080/exist/benchmarks/index.xq`) repository for XML benchmarks, developed by INRIA, The University of Amsterdam, and The University of Antwerp.

MSc work **A parallel MPEG video encoder**. This tool was developed as part of my MSc work, using VECPAR'98 C++ and PThreads, on a Sun multiprocessor with 32 CPUs. It was subsequently transferred to the Electrical and Computer Engineering Department at **Carnegie Mellon University**, where it is being used in the development of a high-level multiprocessor simulator.

MSc work **A Parallel Suffix-Array Indexer**: This work concerned the design and tuning of a Quicksort-based parallel text indexing algorithm using a shared-nothing, message passing architecture. This software was **transferred to a commercial Internet search engine** in Brazil, subsequently acquired by Google.

Invited Talks, Panels, Tutorials, and Visits

Note: For personal and family reasons I have been severely limiting travel in recent years.

2015/10 **Knowledge Extraction From the Web**, University of Regina

2015/06 **Knowledge Extraction From the Web**, Dalhousie University

2015/06 **Knowledge Extraction From the Web**, Federal University of Pernambuco, Recife, Brazil

2015/06 **Knowledge Extraction From the Web**, University of the Peloponnese, Tripoli, Greece

2015/05 **Knowledge Extraction From the Web**, Roma Tre University, Italy

2015/04 **Inferencing in Information Extraction: Techniques and Applications**, *IEEE 31st International Conference on Data Engineering*, Seoul, South Korea
Refereed

2014/11 **Knowledge Extraction From the Web**, Ludwig Maxmilliam University, Munich, Germany

2014/09 **Knowledge Extraction From the Web**, Technical University Darmstad, Germany

2013/04 **Shallow Information Extraction for the Knowledge Web**, *IEEE 29th International Conference on Data Engineering*, Brisbane, Australia
Refereed

2013/02 **Information Extraction for Social Media Analysis**, *2013 UFRJ/EMC Summer School on Big Data*, Rio de Janeiro, Brazil
Invited

2012/06 **Information Extraction for Social Media Analysis**, *6th AAAI Conference on Weblogs and Social Media*, Dublin, Ireland
Refereed

2012/08 **Towards Summarizing and Making Sense of the Blogosphere**, Pontifical Catholic University, Rio de Janeiro, Brazil

2012/06 **Towards Summarizing and Making Sense of the Blogosphere**, Roma Tre University, Italy

2012/03 **Towards Summarizing and Making Sense of the Blogosphere**, *MoMiNIS Seminar Series at Dalhousie University*, Halifax

2010/07 **Google Faculty Summit**, Mountain View, USA
Invited

2010/07 **Top-k Approximate Subtree Matching**, Roma Tre University, Italy

2010/07 **A Framework for Automatic Schema Mapping Verification Through Reasoning**, Roma Tre University, Italy

2010/06 **Building, Exploring and Querying Social Networks**, Roma Tre University, Italy

2009/11 **Building, Exploring and Querying Social Networks**, National Institute for Web Science and Technology, Federal University of Minas Gerais, Belo Horizonte, Brazil

2009/11 **Building, Exploring and Querying Social Networks**, Federal University of Pernambuco, Recife, Brazil

2008/07 **Evaluating the Effectiveness of Autonomic Data Management Tools**, Free University of Bozen-Bolzano, Italy

2008/03 **Towards Schema-Free Exchange and Update of XML Data**, University of Waterloo

2008/02 **Semantics and Information Loss in XML Data Exchange**, INRIA Lille Nord Europe, Lille, France

2006/09 **XML Updates**, *Panel at the 4th International XML Database Symposium*, Seoul, South Korea, Invited

2005/04 **Some Issues in Autonomic Data Management**, University of Alberta

Select Student Supervision

I have also supervised, on average, 2-3 undergraduate students every year (omitted for brevity). In the list below, **PDF** stands for post-doctoral fellow.

2025 – present **PDF**, *Dr. J. G. Mathew (PhD, Sapienza University of Rome)*

2020 – present **MSc**, *N. Hervieux*

2024 – 2026 **MSc**, *E. Vandermeer*

2024 – 2026 **MSc**, *N. Nawal (co-supervised with M. Abdalla)*

2023 – 2025 **MSc**, *H. S. Mohammed (co-supervised with M. A. Nascimento)*

2022 – 2025 **MSc**, *N. Khalifat*

2021 – 2024 **MSc**, *A. Shahriar*

2021 – 2024 **MSc**, *A. Kundu*

2021 – 2022 **MSc**, *S. A. A. Asli*

2020 – 2024 **MSc**, *K. Guzhva*

2020 – 2021 **PDF**, *Dr. A. E. dos Santos (PhD, University of Regina)*

2019 – 2025 **PhD**, *P. Yao*

2019 – 2023 **MSc**, *G. Alexander*

2019 – 2021 **MSc**, *V. Ranganathan*

2019 – 2021 **MSc**, *S. Ravishankar (course-based)*

2018 – 2020 **MSc**, *T. Renwick*

2017 – 2021 **MSc**, *A. Hejazizo (co-supervised with L. Mou)*

2017 – 2020 **MSc**, *E. Macdonald*

2017 – 2019 **MSc**, *D. Caminhas*

2017 – 2019 **MSc**, *W. Ni*

2016 – 2018 **MSc**, *P. Xu*

2015 – 2018 **PhD**, *M. Cannaviccio (degree from Roma Tre University; co-supervised with P. Merialdo)*

2014 **Visiting MSc student**, *M. Cannaviccio, from Roma Tre University*

2013 – 2014 **PDF**, *Dr. Y. Sekhavat (PhD, Memorial University of Newfoundland)*

2012 – 2019 **PhD**, *C. Pang (co-supervised with A. Hindle)*

2011 **Visiting PhD student**, *M. Bronzi, from Roma Tre University*
Research Scientist at Mila (Montreal)

2011 **Visiting PhD student**, *M. Moraes, from Federal University of Rio Grande do Sul (Brazil)*

2011 **Visiting MSc student**, *C. Bigonha, from Federal Univeristy of Minas Gerais (Brazil)*

2010 – 2018 **PhD**, *Z. Guo*

2010 – 2013 **MSc**, *A. Anandan*

2010 – 2011 **MSc**, *S. Husby (course-based)*

2009 – 2011 **MSc**, *H. He (co-supervised with G. Kondrak)*

2009 – 2010 **Visiting PhD student**, *Y. Merhav, from Illinois Institute of Technology*

2009 **Visiting PhD student**, *E. Cortez, from Federal University of Amazonas (Brazil)*

2008 – 2016 **PhD**, *H. Sepehri-Rad*

2008 – 2015 **PhD**, *F. Mesquita*

2008 – 2011 **PhD**, *G. Leighton*

2008 – 2011 **MSc**, *D. Serrano (co-supervised with E. Stroulia)*

2008 – 2011 **MSc**, *M. Hasan (co-supervised with E. Stroulia)*

2008 – 2010 **MSc**, *Z. Guo*

2008 – 2010 **MSc**, *J. Walny*

2008 – 2009 **PDF**, *Dr. P. Cappellari (PhD, Roma Tre University)*

2007 **Visiting MSc student**, *F. Mesquita, from Federal University of Amazonas (Brazil)*

HQP Awards

2019 P. Xu – Honorable mention for Departmental MSc Achievement Award

2017 M. Cannaviccio – 2017 TextExt - DBpedia Open Extraction Challenge

2012 T. Sliwkanich, D. Schneider, A. Yong, M. Home – 2012 ACM SIGMOD Undergraduate Student Competition

Funding

Research Networks

2020 – 2023 Co-Principal Investigator of the **Linked Infrastructure for Networked Cultural Scholarship (LINCS)**; CAD$5,000,000 (20%). Funded by the Canadian Foundation for Innovation (CFI), the provinces of Alberta, Ontario, Quebec and British Columbia, and industry.
LINCS is a consortium of Canadian Universities and Libraries (Alberta, Guelph, Ottawa, Saskatchewan, Victoria, York, Bucknell), consortia (Scholars Portal, Canadian Writing Research Collaboratory, ARC, Erudit, CRKN), industry (Parc/Xerox, Diffbot Inc, Noroon Collaboratory Inc, Born Digital), with partnerships in Europe and the United States to build an Open Knowledge Graph of Canadian Cultural Heritage materials.
`https://lincsproject.ca/`

2009 – 2014 Co-Principal Investigator and lead of the Data Quality Theme of the **NSERC Business Intelligence Network**; CAD$6,000,000 (6%) funded by NSERC with matching contribution from industry. The network involved 6 Canadian universities (Alberta, Carleton, Dalhousie, Toronto, UBC, and Waterloo) and several industrial partners, including IBM (Cognos team) and SAP (Business Objects team) and Palomino.

Alberta Innovates Advance Program

2026 – 2027 CAD$40,000: *AI in Education*.

MITACS Accelerate Grants

2022 – 2023 IRISNDT; CAD$40,000: *Machine Learning for Non-Destructive Testing*.

2020 – 2021 Analytica Advisors; CAD$15,000: *COVID19 and Disclosure of Risk in Energy Exposed Sectors*.

2020 – 2021 The Canadian Energy and Climate Nexus (CECN); CAD$100,000 (33%); *A Toolkit for Analyzing Online Conversations for Solutions Based Policy Development*.

2018 – 2019 MatchWorks; CAD$30,000: *Data Driven Assessment for Self-Determined Employment*.

2018 – 2019 Granify; CAD$75,000 (33%): *Semantic Information Processing to Improve e-Commerce Shopper Experience*.

NSERC Engage Grants

2019 Bioware/EA Games; CAD$25,000; *Text Analysis for Understanding Gamer Social Behavior*.

2017 HireGround Software Solutions; CAD$25,000; *Deep Learning and Word Embeddings for Information Retrieval*.

2016 Kira Systems Inc; CAD$25,000; *Document Structure Driven Named Entity Recognition*.

2013 Apropos Information Systems; CAD$25,000; *Information Extraction on Indigenous Knowledge*.

Other Industrial Research Support

2020 – 2022 Lead Researcher in the **Scotiabank Artificial Intelligence Research Initiative** at the University of Alberta; CAD$360,000.
Multiple projects on explainability in banking models, customer relations management, and detection of fraudulent transactions.

2019 Grant from Amazon AWS, USD$40,000 (AWS credits).

2018 Grant from Servus Credit Union, CAD$50,000.

2017 Gift from **DiffBot.com**, USD$55,000.

- Gift from **NVIDIA**, CAD$2,000.

2013 – 2014 Grant from IBM Alberta Centre for Advanced Studies, CAD$40,000.

2011 Gift from **GE Intelligent Systems**, CAD$10,000.

2006 **IBM Faculty Award**, CAD$11,500.

Individual Research Support

2006 – 2030 **NSERC Discovery Grants**: CAD$693,000

- 2024 – 2030: CAD$250,000
- 2019 – 2023: CAD$115,000[1]
- 2013 – 2018: CAD$150,000
- 2008 – 2012: CAD$115,000
- 2005 – 2007: CAD$ 63,000

2007 – 2010 **Alberta Ingenuity New Faculty Award**, CAD$294,000.

2008 – 2012 Start-up grant at the University of Alberta, CAD$40,000.

2005 – 2006 Start-up grant at the University of Calgary, CAD$50,000.

Academic Service

Societies

2024 – 2025 **Member, Strategic Advisory Group**, *Net Zero Data Space, CEIMIA and the Centre for AI & Climate*

2023 – 2025 **President**, *Canadian Artificial Intelligence Association*
Acclaimed

2021 – 2023 **Vice-President**, *Canadian Artificial Intelligence Association*
Acclaimed

2019 – 2021 **Secretary**, *Canadian Artificial Intelligence Association*
As a member of the executive team I led the member consultation and transition to open access proceedings for the Canadian AI conference.

2020, 2021 **Elections Committee Member**, *CSCan-InfoCan*

2006 – 2012 **Associate Information Director**, *ACM SIGMOD*
Primarily responsible for publishing the SIGMOD Record online.

Select Reviewing Service

2019 – 2022 **Associate Editor**, *Distributed and Parallel Databases*, Springer

2015 – 2018 **Associate Editor**, *Transactions on Knowledge and Data Engineering*, IEEE

2015 – 2018 **Associate Editor**, *Computational Intelligence*, Elsevier

2012 – present **Reviewer**, *Computing Reviews*, ACM

2010 – 2014 **Associate Editor**, *SIGMOD Record*, ACM

Select Conference Organization

2016 **Program Committee Co-Chair**, *Applications Track of the 3rd IEEE International Conference on Data Science and Advanced Analytics*, Montreal
co-chair: G. Shroff, TATA

2015 **Program Committee Co-Chair**, *28th Canadian Conference on Artificial Intelligence*, Halifax
co-chair: E. Milios, Dalhousie

2012 **Program Committee Co-Chair**, *Second ACM SIGMOD Workshop on Databases and Social Networks*, Scottsdale, AZ, USA
co-chairs: E. Terzi, Boston University; and K. Lefevre, Google

2012 **Program Committee Co-Chair**, *3rd International Workshop on Data Engineering Meets the Semantic Web*, co-located with ICDE, Washington, DC, USA
co-chair: R. Pottinger, UBC

2011 **Program Committee Co-Chair**, *First ACM SIGMOD Workshop on Databases and Social Networks*, Athens, Greece
co-chairs: G. Miklau, U. Massachusetts Amherst; and C. Yu, Google

2008 **Web Chair**, *ACM SIGMOD Conference*, Vancouver

[1] From what I could gather from talking to applicants in this competition year from multiple universities, all applicants to the data management received drastic cuts compared to their previous grants.

2007 **Program Committee Co-Chair**, *Database and XML Technologies, 5th International XML Database Symposium*, co-located with VLDB, Vienna, Austria
co-chairs: A. Bonifati, INRIA; Z. Bellahsene, Université de Montpellier; E. Pustulka, FHNW University of Applied Sciences and Arts; and R. Unland, University of Duisburg-Essen

2007 **Publicity Chair**, *World Wide Web Conference*, Banff

2007 **Proceedings Chair**, *19th Scientific and Statistical Database Management Conference (SSDBM)*, Banff

Select Conference Program Committees

Recently, I have been serving as reviewer on 4-5 events each year, including focused workshops and national conferences (both Brazilian and Canadian) while being selective on the number of larger conferences I serve. The list below contains only international top-tier venues I have served as a regular member of the Program Committee.

EMNLP ACL Conference on Empirical Methods in Natural Language Processing (2020, 2021, 2022, 2023)

NAACL Annual Conference of the North American Chapter of the ACL (2016, 2018)

ACL Annual Meeting of the Association for Computational Linguistics (2020, 2022)

CIKM ACM International Conference on Information and Knowledge Management (2007, 2011, 2020)

ISWC International Semantic Web Conference (2012, 2013)

WWW The Web Conference (formerly the World Wide Web Conference) (2004, 2005, 2006, 2007, 2010, 2013, 2017, 2020)

WSDM ACM International Conference on Web Search and Data Mining (2017, 2019, 2021)

WebDB ACM/SIGMOD International Workshop on the Web and Databases (2006, 2007, 2009, 2013, 2014)

EDBT International Conference on Extending Database Technology (2010, 2012, 2014, 2018, 2020, 2021)

VLDB International Conference on Very Large Data Bases (2008, 2010, 2015, 2018)

SIGMOD ACM SIGMOD International Conference on Management of Data (2011, 2016, 2018)

ICDE IEEE International Conference on Data Engineering (2006, 2007, 2008, 2015, 2016)

Select External Reviews

Beyond what is listed below, I normally review 2 to 3 NSERC Discovery Grants (since 2006), 1 or 2 NSERC Alliance grants every year (since 2010), and 2 to 3 MITACS grant proposals (since 2012).

2025 **NSERC**, Member – CSE Research Communities Evaluation Committee

2024 **Toronto Metropolitan University**, Tenure Committee

2024 **Alberta Ministry of Education**, Data Science Program Reviewer

2024 **Maritime Provinces Higher Education Commission**, Program Reviewer

2024 **Deutsche Forschungsgemeinschaft (DFG)**, Grant evaluator

2023 **McMaster University**, Tenure Review

2023 **Hong Kong Research Grants Council**, Grant evaluator

2021 **Canadian Foundation for Innovation (CFI)**, Grant evaluator

2021 **University of Toronto**, Tenure and Promotion Reference

- **NSERC**, Reviewer – CREATE Program

- **NSERC**, Reviewer – Idea 2 Innovation Program

2021, 2022 **Prince Sattam Bin Abdulaziz University**, Tenure Evaluation

2020 **McMaster University**, Tenure And Promotion Evaluator

2020, 2019, 2016, 2017 **Swiss National Science Foundation**, Grant evaluator

2017 **NSERC**, Site Visit Committee Member – Industrial Research Chair program

- **NSERC**, Panel member – Canada Research Chair program

2016 **ANVUR**, *Italian National Agency for the Evaluation of Universities and Research Institutes*, Evaluator of Computer Science researchers and programs.

2016, 2014 **Canada Foundation for Innovation**, Grant evaluator

2016, 2014 **CINECA**, *Italian Consortium of Universities*, Faculty evaluator

2016, 2012 **FCT**, *Portuguese Foundation for Science and Technology*, Exact Sciences Panel Member

2010 **British Columbia Innovation Council**, Grant evaluator

2008 **Israel Science Foundation**, Grant evaluator

Select External Journal Reviews

2005–2008, 2010–2012, 2019–2021 IEEE Transactions on Data and Knowledge Engineering

2019 Information Processing and Management (Elsevier)

2018 IEEE Access

2005, 2007, 2008, 2011, 2017 Information Systems (Elsevier)

2017 Scholarly and Research Communication

2016 Artificial Intelligence (Elsevier)

2008–2010, 2014 The VLDB Journal (VLDB Endowment/Springer)

2012 Journal of Web Semantics (Elsevier)

2011 IEEE Internet Computing

2011 IEEE Transactions on Dependable and Secure Computing

2003–2006, 2010 ACM Transactions on Database Systems

2007, 2009 ACM Transactions on the Web

2009 Journal of Computer and Systems Sciences (Elsevier)

2005, 2008 Software: Practice and Experience (John Wiley & Sons)

2007, 2008 Data & Knowledge Engineering (Elsevier)

2008 Journal of Systems and Software (Elsevier)

2008 ACM Computing Surveys

MSc and PhD Examinations

At the U. of Alberta, excluding students supervised by me, between 2008 and 2021, I have participated in: (a) 12 PhD Supervisory Committees (5 ongoing); (b) 17 PhD examinations; and (c) 15 MSc examinations. Beyond those, I have also participated in the following examinations.

2021 - PhD, McMaster University, Canada

2021 - PhD, University of New South Wales, Australia

- MSc, McGill University
- MSc, Federal University of Minas Gerais, Brazil

2019 - MSc, Federal University of Minas Gerais, Brazil

2017 - PhD, Federal University of São Paulo (IME), Brazil

- MSc, University of Melbourne, Australia

2015 - PhD (Dr-Ing Dissertation), University of Saarland/Max Planck Inst. for Informatics, Germany

2014 - PhD, University of Sidney, Australia

2012 - PhD, Roma Tre University, Italy

- PhD, Dalhousie University
- MSc, Federal University of Minas Gerais, Brazil

2010 - MSc, University of Amsterdam, Netherlands

2008 - PhD, Roma Tre University, Italy

2007 - MSc, City University Dublin, Ireland

- MSc, Federal University of Amazonas, Brazil

2006 - MSc, University of Calgary

Administrative Service

Service to the University of Alberta

The following is a list of my extraordinary administrative service at the University of Alberta. Ordinary service includes being a member of the Graduate Student Admission committee. Since 2011 I have also been a member of the Tenure committee. Between 2014 and 2015 I was on sabbatical leave and relieved of service.

2024 – present **Chair of the General Faculties Council Appeal Board**, *University of Alberta*

2024 – present **SAIP Faculty Decision-Maker**, *College of Natural and Applied Sciences*

2023 **Associate Dean, Undergraduate**, *Faculty of Science*

2022 **Associate Chair, Graduate**, *Departmental*

2020 **Event Organizer**, *Reverse Expo*, February 14, with over 200 attendees from CS, Engineering and other faculties, guests from industry and government. Funded by by AltaML, BioWare, DeepMind, Intuit, Levven, Scotiabank, Servus Credit Union, and Startup Edmonton.

2020 **Committee Member**, *Faculty of Science*, Ad-hoc committee to liaise with a large Canadian Bank seeking to establish an AI lab in Alberta

2019 – 2021 **Industrial Relations Director**, *Departmental*

2019 – 2021 **Teaching Assistantship Coordinator**, *Departmental*

2019 **Event Organizer**, *Reverse Expo*, February 19, with over 200 attendees from CS, Engineering and other faculties, guests from industry and government. Funded by by Alberta Innovates, AltaML, BioWare, DeepMind, Google, Intuit, and Startup Edmonton.

2020 **Committee Member**, *Faculty of Science*, Ad-hoc committee to establish a partnership with Servus Credit Union

2018 **Event Organizer**, *Reverse Expo*, February 16, with over 130 attendees from CS and Engineering, guests from industry and government. Funded by by Alberta Innovates, AltaML, Deloitte, Google, Intuit, Servus Credit Union, and Startup Edmonton.

2017 **Event Organizer**, *Reverse Expo*, February 24, over 70 attendees from CS, industry and government. Funded by NSERC

2016 – 2017 **Committee Member**, *Department's Industry Liaison Committee*

2013 – 2014 **Science Internship Program Director**, *Departmental*

2012 – 2013 **IT Oversight Committee**, *Faculty of Science*

2010 – 2012 **Member of the Operations Executive Committee**, *Departmental*

2009 – 2011 **Member of the Undergraduate Curriculum Committee**, *Departmental*

Service to the University of Calgary

2006 – 2007 **Coach of the ACM ICPC Team**

- 2007 ACM ICPC World Finals: honorable mention
- 2007 ACM ICPC Rocky Mountain Regional (3 teams): **1st**, 14th, and 22nd
- 2007 Alberta Collegiate Programming Contest (3 teams): 1st, 5th, and 6th
- 2006 ACM ICPC Rocky Mountain Regional (3 teams): **1st**, 9th, and 13th
- 2006 Alberta Collegiate Programming Contest (3 teams): 2nd, 5th, and 6th

2005 – 2007 **Undergraduate Affairs Committee**, *Department of Computing Science*

Teaching and Course Development

Teaching at the University of Alberta

2024 – present **CMPUT461/561**, *Introduction to Natural Language Processing*, Regular expressions and automata; parts of speech and tagging; grammars and parsing; language models; text classification and sentiment analysis; evaluation; word embeddings; large language models.
Note: I modernized the content inherited from CMPUT497, and added new assignments and a course project.

2021 – present **CMPUT361**, *Introduction to Information Retrieval*, Boolean, ranked, and probabilistic information retrieval; text clustering and classification; Web search engines.
Note: I developed this course, and recently modernized all of its evaluations and assignments.

2019 – 2020 **CMPUT497**, *Topics in Natural Language Processing*, 2 offerings, **Co-taught** with G. Kondrak and C. Demmans Epp. Introduction to classical and probabilistic models for NLP tasks: POS tagging, grammars and parsing, sentiment analysis, information extraction.

2016 – 2021 **CMPUT391**, *Database Management Systems*, 8 offerings, Advanced SQL; query and transaction processing; non-relational data management; noSQL; distributed and parallel databases; Semantic Web
Note: I completely redesigned this course, modernizing its coverage and introducing several modern topics and tools, including cloud-based data management (AWS offerings).

2016 – 2021 **CMPUT361**, *Introduction to Information Retrieval*, 4 offerings, Boolean, ranked, and probabilistic information retrieval; text clustering and classification; Web search engines
Note: I introduced this course, basing the material on comparable courses taught in top schools in N. America.

2013 – 2019 **CMPUT696**, *Knowledge Graphs*, 4 offerings, Introduction to knowledge graphs; uses in question answering and other NLP tasks; building and curating knowledge graphs.

2012 **CMPUT174**, *1 offering*, Foundations of Programming I (Python), Data structures, such as control flow, iterations, lists, sets and maps, with an introduction to recursion.

2011 – 2014 **Science 100**, *3 offerings*, Integrated 1st year Science program, Computing Science component, equivalent to CMPUT174 but with emphasis on science-related coursework.
Note: I completely redesigned the course material, transitioning it from Perl into Python and added 3 new assignments introducing data science.

2011 **CMPUT693**, *Topics in Social Network Analysis*, 1 offering, Introduction to models of social network analysis and applications on real datasets such as twitter.

2009 **CMPUT690**, *Topics in Database Systems – Web Data management*, 1 offering, Semistructured data models, storage and query processing

2008 – 2011 **CMPUT115**, *Practical Programming with Java II*, 4 offerings, Advanced topics and data structures, including recursion, backtracking, cost analysis of algorithms, linked lists, self-balancing tree structures.
Note: I completely redesigned the course material to modernize its coverage, for which I received a teaching award.

Teaching at the University of Calgary

2005 – 2008 **CPSC471**, *Database Systems*, 4 offerings, Introduction to SQL, relational algebra, database modeling and normalization, introduction to indexing and tuning, XML.

2007 **CPSC601.24**, *Social Network Analysis*, 1 offering, Introduction to models of social network analysis and applications on real datasets such as twitter.

2005 – 2006 **CPSC601.15**, *Web Data Management*, 2 offerings, Semistructured data models, storage and query processing.

2005 – 2007 **CPSC457**, *Operating Systems*, 2 offerings, Introduction to OS concepts, devices, memory and process management, the file system, scheduling and concurrency control.
Note: I completely redesigned the course material, an introduced 4 new assignments, based on what was taught at top CS schools in N. America.

Teaching at the University of Toronto

2001 **CSC43**, *Introduction to Databases*, 1 offering, Introduction to SQL, relational algebra, database modeling and normalization, introduction to indexing and tuning.