

Daniel Lemire, Ph.D.

Centre de recherche LICEF
TELUQ, Université du Québec
Pavillon Saint-Urbain
5800 Saint-Denis
Office 1105
Montreal (Quebec)
H2S 3L5 Canada

lemire@gmail.com
<http://lemire.me/en/>

Canadian citizen

Experience

- | | |
|-----------|---|
| 2004–... | Computer Science Professor
LICEF Research Center, TELUQ, Université du Québec
Promoted to Full Professor in 2009 |
| 2002–2004 | Research Officer in eBusiness
National Research Council of Canada (NRC)
Team Leader of the e-Health Research Group in 2002–2003 |
| 2001–2002 | Assistant Professor
Acadia University |
| 1999–2001 | Entrepreneur
TechElements Inc. and Ondelette.com |
| 1998–1999 | Post-doctoral fellow
Institut de génie biomédial |

Education

- | | |
|-----------|--|
| 1995–1998 | Ph.D.—“Iterative Schemas”
École Polytechnique de Montréal and Université de Montréal
Supervisors: Gilles Deslauriers and Serge Dubuc |
| 1994–1995 | M.Sc.—“A Priori Approximations for Nonlinear Systems”
University of Toronto
Supervisor: Catherine Sulem |
| 1990–1994 | B.Sc. with “High Distinction”
University of Toronto |

Competitive scholarships and prizes as a student

- FCAR scholarship (Ph.D.) and NSERC scholarship (M.Sc. and Ph.D.);
- C. D. Howe Memorial for 4 years (≈\$50,000);
- NSERC Canada Scholarship for 4 years (\$4,000);
- St. Michael’s College Dean’s medal (class of 1994);
- 3T0 Scholarship (UofT).

Secondary affiliations

- I am an adjunct professor at the University of New Brunswick and at the Université du Québec à Montréal where I supervise graduate students.

Recent Teaching Experience

Undergraduate (in French):

- INF 1250 - Introduction aux bases de données;
- INF 6460 - Recherche et filtrage d'informations;
- INF 6450 - Gestion de l'information avec XML;
- INF 9002 - Évaluation et analyse des systèmes d'information;
- INF 9004 - Informatique des entrepôts de données.

Graduate (in French):

- INF 6104 - Recherche d'informations et web;
- INF 6107 - Le Web social;
- INF 6408 - Informatique de l'analyse multidimensionnelle.

Moreover, I have chaired and cochaired several programs: two certificates, one bachelor degree, and four graduate programs including the master in Information Technology.

Professional Membership

- Association for Computing Machinery (ACM).

Research program

Individual Research Funding (external agencies, competitive grants)

- NSERC Discovery Grant 2017–2022 (\$210,000);
- NSERC Discovery Accelerator Supplement 2017–2020 (\$120,000);
- NSERC Discovery Grant 2012–2017 (\$140,000);
- NSERC Discovery Grant 2007–2012 (\$75,000);
- NSERC Discovery Grant 2003–2007 (\$48,000);
- FQRNT New Researcher Grant 2006–2008 (\$54,249).

Team-Based Research Funding (external agencies, competitive grants)

- CFI Grant 2016–2017 (\$797,481) with N. Bélanger (lead) and E. Filotas;
- CFI Grant 2008–2009 (\$999,618) with G. Paquette (lead) and P. Valtchev;
- New Brunswick Innovation Fund 2007–2008 (\$10,000) with O. Kaser (lead).
- CAA Foundation (2013–2014, \$189 994) with E. Vallières (lead) et al.

Services (external)

Blog

- I run a blog at <http://lemire.me/>. It was featured on Reddit, Hacker News and Slashdot. It has over 25 000 unique visitors a month.

Program committees (incomplete list)

- ACM Conference on Information Retrieval (SIGIR) 2015;
- ACM Conference on Recommender Systems (RecSys) 2009–2014, 2017;

- ACM Conference on Information and Knowledge Management (CIKM) 2012–2017;
- ACM Conference on Web Search and Data Mining (WSDM) 2013–2015;
- World Wide Web Conference (WWW) 2017;
- ACM/IEEE Joint Conference on Digital Libraries (JCDL) 2011–2017.

Funding Agencies

- At the FQRNT, I have been a member of the review committee 03F (Theoretical Computer Science) since 2007.
- Still at the FQRNT, I was member of the review committee 309 (Computer Science team projects) in 2006–2007, in 2013–2014, 2014–2015 and 2016–2017.
- At NSERC, I was a member of the Computer Science Research Tools and Instruments (RTI) Review Committee in 2012–2013, in 2013–2014 and in 2014–2015.

External examiner/referee

- External examiner for Ph.D. thesis:
 - Mohammed Shaaban from Université Pierre et Marie Curie, France (2017) — supervised by Patrick Garda;
 - Mehdi Boukhechba from UQAC, Canada (2016) — supervised by Abdenour Bouzouane and Charles Gouin-Vallerand;
 - Hicham Assoudi from UQAM, Canada (2016) — supervised by Hakim Lounis;
 - Khaled Dehdouh from Lyon 2, France (2015) — supervised by Omar Boussaid;
 - Martin Leginus from Aalborg University, Denmark (2015) — supervised by Peter Dolog;
 - Ahmad Taleb from Concordia University, Canada (2011) — supervised by Todd Eavis.
- External referee for habilitation:
 - Sabine Loudcher Rabaseda from Lyon 2 University, France (2011).
- External referee for promotion:
 - Jason Sawin from the University of St. Thomas (2015);
 - Amer Nizar AbuAli from Philadelphia University (2013);
 - Ken Pu from the Ontario Institute of Technology (2011);
 - Jinan Fiaidhi from Lakehead University (2005).

Peer-Reviewed Journal Articles

Impact factors of some journals: *Information Sciences* (2.8), *Pattern Recognition* (2.6), *Journal of the Association for Information Science and Technology* (2.0), *Data & Knowledge Engineering* (1.7), *ACM Transactions on Database Systems* (1.4), *Computer Speech and Language* (1.4), *Information Retrieval* (1.3), *ACM Transactions on Information Systems* (1.3) and *Information Systems* (1.2).

- [1] Wojciech Muła, Nathan Kurz, Daniel Lemire, Faster Population Counts Using AVX2 Instructions, *Computer Journal* (to appear)
<https://doi.org/10.1093/comjnl/bxx046>

- [2] Antonio Badia and Daniel Lemire, On Desirable Semantics of Functional Dependencies over Databases with Incomplete Information, *Fundamenta Informaticae* (to appear) <https://arxiv.org/abs/1703.08198>
- [3] Dmytro Ivanchykhin, Sergey Ignatchenko, Daniel Lemire, Regular and almost universal hashing: an efficient implementation, *Software: Practice and Experience* (to appear) <http://dx.doi.org/10.1002/spe.2461>
- [4] Daniel Lemire, Christoph Rupp, Upscaledb: Efficient Integer-Key Compression in a Key-Value Store using SIMD Instructions, *Information Systems* **66**, 2017. <http://dx.doi.org/10.1016/j.is.2017.01.002>
- [5] Jing Li, Yuhong Yan, Daniel Lemire, Full Solution Indexing for top-K Web Service Composition, *IEEE Transactions on Services Computing* **99**, 2016. <http://dx.doi.org/10.1109/TSC.2016.2578924>
- [6] Samy Chambi, Daniel Lemire, Robert Godin, Vers de meilleures performances avec des Roaring bitmaps, *Technique et Science Informatiques* **35** 3, 2016.
- [7] Daniel Lemire, Gregory Ssi-Yan-Kai, Owen Kaser, Consistently faster and smaller compressed bitmaps with Roaring, *Software: Practice & Experience* **46** 11, 2016. <http://dx.doi.org/10.1002/spe.2402>
- [8] Daniel Lemire, Owen Kaser, Faster 64-bit universal hashing using carry-less multiplications, *Journal of Cryptographic Engineering* **6** 3, 2016. <http://dx.doi.org/10.1007/s13389-015-0110-5>
- [9] Daniel Lemire, Nathan Kurz, Leonid Boytsov, SIMD Compression and the Intersection of Sorted Integers, *Software: Practice & Experience* **46** 6, 2016. <http://dx.doi.org/10.1002/spe.2326>
- [10] Samy Chambi, Daniel Lemire, Owen Kaser, Robert Godin, Better bitmap performance with Roaring bitmaps, *Software: Practice & Experience* **46** 5, 2016. <http://dx.doi.org/10.1002/spe.2325>
- [11] Owen Kaser and Daniel Lemire, Compressed bitmap indexes: beyond unions and intersections, *Software: Practice & Experience* **46** 2, 2016. <http://dx.doi.org/10.1002/spe.2289>
- [12] Adina Crainiceanu and Daniel Lemire, Multidimensional Bloom Filters, *Information Systems* **54**, 2015. <http://dx.doi.org/10.1016/j.is.2015.01.002>

- [13] Antonio Badia and Daniel Lemire, Functional dependencies with null markers, *Computer Journal* **58** 5, 2015. <http://dx.doi.org/10.1093/comjnl/bxu039>
- [14] Wayne Xin Zhao, Xudong Zhang, Daniel Lemire, Dongdong Shan, Jian-Yun Nie, Hongfei Yan, Ji-Rong Wen, A General SIMD-based Approach to Accelerating Compression Algorithms, *ACM Transactions on Information Systems* **45** 1, 2015. <http://dx.doi.org/10.1145/2735629>
- [15] Xiaodan Zhu, Peter Turney, Daniel Lemire, Andre Vellino, Measuring academic influence: Not all citations are equal, *Journal of the Association for Information Science and Technology* **66** 2, 2015. <http://dx.doi.org/10.1002/asi.23179>
- [16] Daniel Lemire and Leonid Boytsov, Decoding billions of integers per second through vectorization, *Software: Practice & Experience* **45** 1, 2015. <http://dx.doi.org/10.1002/spe.2203>
- [17] Owen Kaser and Daniel Lemire, Strongly universal string hashing is fast, *Computer Journal* **57** 11, 2014. <http://dx.doi.org/10.1093/comjnl/bxt070>
- [18] Hazel Webb, Daniel Lemire, Owen Kaser, Diamond Dicing, *Data & Knowledge Engineering* **86**, 2013. <http://arxiv.org/abs/1006.3726>
- [19] Zoltán Prekopcsák and Daniel Lemire, Time Series Classification by Class-Specific Mahalanobis Distances, *Advances in Data Analysis and Classification* **6** 3, 2012. <http://arxiv.org/abs/1010.1526>
- [20] Daniel Lemire, Owen Kaser, Eduardo Gutarra, Reordering Rows for Better Compression: Beyond the Lexicographic Order, *ACM Transactions on Database Systems* **37** 3, 2012. <http://arxiv.org/abs/1207.2189>
- [21] Daniel Lemire, The universality of iterated hashing over variable-length strings, *Discrete Applied Mathematics* **160** (4-5), 2012. <http://arxiv.org/abs/1008.1715>
- [22] Antonio Badia and Daniel Lemire, A Call to Arms: Revisiting Database Design, *SIGMOD Record* **40** 3, 2011. <http://arxiv.org/abs/1105.6001>
- [23] Daniel Lemire and Owen Kaser, Reordering columns for smaller indexes, *Information Sciences* **181** 12, 2011. <http://arxiv.org/abs/0909.1346>

- [24] Daniel Lemire and Owen Kaser, Recursive n-gram hashing is pairwise independent, at best, *Computer Speech and Language* **24** 4, 2010. <http://arxiv.org/abs/0705.4676>
- [25] Daniel Lemire, Owen Kaser, Kamel Aouiche, Sorting improves word-aligned bitmap indexes, *Data & Knowledge Engineering* **69** 1, 2010. <http://arxiv.org/abs/0901.3751>
Cited at least **39 times**.
- [26] Daniel Lemire, Faster retrieval with a two-pass Dynamic-Time-Warping lower bound, *Pattern Recognition* **42** 9, 2009. <http://arxiv.org/abs/0811.3301>
Cited at least **31 times**.
- [27] Daniel Lemire, Martin Brooks, Yuhong Yan, An optimal linear time algorithm for quasi-monotonic segmentation, *International Journal of Computer Mathematics* **86** 7, 2009. <http://arxiv.org/abs/0709.1166>
- [28] Kamel Aouiche, Daniel Lemire and Robert Godin, Web 2.0 OLAP: From data cubes to tag clouds, *Lecture Notes in Business Information Processing* **18**, 2009. <http://arxiv.org/abs/0905.2657>
- [29] Daniel Lemire and Owen Kaser, Hierarchical Bin Buffering: Online local moments for dynamic external memory arrays, *ACM Transactions on Algorithms* **4** 1, 2008.
- [30] Daniel Lemire, Streaming maximum-minimum filter using no more than three comparisons per element, *Nordic Journal of Computing* **13** 4, 2006.
Cited at least **24 times**.
- [31] Owen Kaser and Daniel Lemire, Attribute value reordering for efficient hybrid OLAP, *Information Sciences* **176** 16, 2006.
Cited at least **33 times**.
- [32] Daniel Lemire, Harold Boley, Sean McGrath, Marcel Ball, Collaborative filtering and inference rules for context-aware learning object recommendation, *International Journal of Interactive Technology & Smart Education* **2** 3, 2005.
Cited at least **48 times**.
- [33] Daniel Lemire, Scale and Translation Invariant Collaborative Filtering Systems, *Information Retrieval* **8** 1, 2005.
Cited at least **64 times**.

- [34] Serge Dubuc, Daniel Lemire, Jean-Louis Merrien, Fourier analysis of 2-point Hermite interpolatory subdivision schemes, *Journal of Fourier Analysis and Applications* **7** 5, 2001.
- [35] Daniel Lemire, Chantal Pharand, Jean-Claude Rajaonah, Bruno Dubé, A.-Robert LeBlanc, Wavelet time entropy, T wave morphology and myocardial ischemia, *IEEE Transactions in Biomedical Engineering* **47** 7, 2000.
Cited at least **56 times**.
- [36] Gilles Deslauriers, Serge Dubuc, and Daniel Lemire, Une famille d'ondelettes biorthogonales sur l'intervalle obtenue par un schéma d'interpolation itérative, *Annales des Sciences Mathématiques du Québec* **23** 1, 1999.

In Peer-Reviewed Proceedings

- [37] Dara Aghamirkarimi, Daniel Lemire, Discovering the Smart Forests with Virtual Reality, WIMS 2017, 2017.
- [38] Samy Chambi, Daniel Lemire, Robert Godin, Kamel Boukhalfa, Charles Allen, Fangjin Yang, Optimizing Druid with Roaring bitmaps, IDEAS 2016, 2016.
- [39] Jing Li, Yuhong Yan, Daniel Lemire, Scaling up Web Service Composition with the Skyline Operator, IEEE ICWS 2016, 2016.
- [40] Samy Chambi, Daniel Lemire, Robert Godin, Nouveaux modèles d'index bitmap compressés à 64 bits, EDA 2016, 2016.
- [41] Perrine Ruer, Charles Guin-Vallerand, Le Zhang, Daniel Lemire, and Evelyne F. Vallières, An analysis tool for the contextual information from field experiments on driving fatigue, CONTEXT 2015, 2015.
- [42] Jing Li, Yuhong Yan, Daniel Lemire, A web service composition method based on compact K2-trees, IEEE SCC 2015, 2015.
- [43] Jeff Plaisance, Nathan Kurz, Daniel Lemire, Vectorized VByte Decoding, International Symposium on Web Algorithms 2015, 2015.
- [44] Samy Chambi, Daniel Lemire, Robert Godin, Owen Kaser, Roaring bitmap : nouveau modèle de compression bitmap, EDA 2014, 2014.
- [45] Jing Li, Yuhong Yan, Daniel Lemire, Full Solution Indexing Using Database for QoS-aware Web Service Composition, IEEE SCC 2014, 2014. **(Best Student Paper Award)**
- [46] Andre Vellino and Daniel Lemire, Extracting, Transforming and Archiving Scientific Data, VLDL 2011, 2011.

- [47] Owen Kaser, Daniel Lemire, Kamel Aouiche, Histogram-aware sorting for Enhanced Word-Aligned Compression in bitmap indexes, DOLAP 2008, 2008.
- [48] Kamel Aouiche, Daniel Lemire, Owen Kaser, Tri de la table de faits et compression des index bitmaps avec alignement sur les mots, BDA 2008, 2008.
- [49] Hazel Webb, Owen Kaser, Daniel Lemire, Pruning attributes from data cubes with Diamond Dicing, IDEAS 2008, 2008.
- [50] Kamel Aouiche, Daniel Lemire, Robert Godin, Collaborative OLAP with tag clouds: Web 2.0 OLAP formalism and experimental evaluation, WEBIST 2008, 2008.
- [51] Kamel Aouiche and Daniel Lemire, A comparison of five probabilistic view-size estimation techniques in OLAP, DOLAP 2007, 2007.
- [52] Owen Kaser and Daniel Lemire, Removing manually-generated boilerplate from electronic texts: Experiments with project Gutenberg e-books, CASCON 2007, 2007.
- [53] Owen Kaser and Daniel Lemire, Tag-Cloud Drawing: Algorithms for cloud visualization, Tagging and Metadata for Social Information Organization (WWW 2007), 2007.
Cited at least **167 times**.
- [54] Kamel Aouiche and Daniel Lemire, Unassuming view-size estimation techniques in OLAP: An experimental comparison, ICEIS 2007, 2007.
- [55] Daniel Lemire, A better alternative to piecewise linear time series segmentation, SIAM Data Mining 2007, 2007.
Cited at least **38 times**.
- [56] Dan Kucerovsky and Daniel Lemire, Monotonicity analysis over chains and curves, Curves and Surfaces 2006, 2007.
- [57] Owen Kaser, Daniel Lemire, Steven Keith, The LitOLAP Project: data warehousing with literature, CaSTA 2006, 2006.
- [58] Daniel Lemire, Martin Brooks, Yuhong Yan, An optimal linear time algorithm for quasi-monotonic segmentation, ICDM 2005, 2005.
- [59] Will Fitzgerald, Daniel Lemire, Martin Brooks, Quasi-monotonic segmentation of state variable behavior for reactive control, AAAI 2005, 2005.

- [60] Yuhong Yan, Martin Brooks, Daniel Lemire, Scale-based monotonicity analysis in qualitative modelling with flat segments, IJCAI 2005, 2005.
- [61] Daniel Lemire and Anna Maclachlan, Slope One predictors for online rating-based collaborative filtering. SIAM Data Mining 2005, 2005.
Cited at least **282 times**.
- [62] Yuhong Yan, Daniel Lemire, Martin Brooks, Monotone pieces analysis for qualitative modeling, MONET 2004, 2004.
- [63] Michelle Anderson, Marcel Ball, Harold Boley, Stephen Greene, Nancy Howse, Daniel Lemire, Sean McGrath, RACOFI: A rule-applying collaborative filtering system, IEEE/WIC COLA 2003, 2003.
Cited at least **102 times**.
- [64] Owen Kaser and Daniel Lemire, Attribute value reordering for efficient hybrid OLAP. DOLAP 2003, 2003.
- [65] Daniel Lemire, A family of 4-point dyadic high resolution subdivision schemes, Curves and Surfaces 2002, 2003.
- [66] Daniel Lemire, Wavelet-based relative prefix sum methods for range sum queries in data cubes, CASCON 2002, 2002. (**Best Paper Award**)

Books

- [67] Mamadou Tadiou Koné and Daniel Lemire (Eds.), Canadian Semantic Web, Springer, 2006.

Special Issues and editorials

- [68] Cameron Neylon, Jan Aerts, C. Titus Brown, Daniel Lemire, Jarrod Millman, Peter Murray-Rust, Fernando Perez, Neil Saunders, Arfon Smith, Gaël Varoquaux and Egon Willighagen, Changing computational research: The challenges ahead, *Source Code for Biology and Medicine* **7** (2), 2012.
- [69] Daniel Lemire and Richard Hotte (Eds.), Special issue on learning and the social web, *Journal of Emerging Technologies in Web Intelligence* **2** 1, 2010.
- [70] Mamadou Tadiou Koné and Daniel Lemire (Eds.), Special issue on Canadian Semantic Web, *Computational Intelligence* **23** 3, 2007.

Book chapters

- [71] Sylvie Noël and Daniel Lemire, On the challenges of collaborative data processing, in Collaborative Information Behaviour: User Engagement and Communication Sharing (edited by Jonathan Foster), IGI Global, 2010.

Magazines

- [72] Daniel Lemire, Marketing your ideas: Don't sell yourself short, ACM XRDS: Crossroads **16** (4), 2010.

Technical Reports

- [73] Owen Kaser and Daniel Lemire, Threshold and Symmetric Functions over Bitmaps, UNBSJ CSAS Technical Report TR-14-001, 2014.
- [74] Hazel Webb, Owen Kaser, Daniel Lemire, Pruning attributes from data cubes with Diamond Dicing, UNBSJ CSAS Technical Report TR-08-011, 2008.
- [75] Owen Kaser and Daniel Lemire, Removing manually-generated boilerplate from electronic texts: Experiments with project Gutenberg e-books. UNBSJ CSAS Technical Report TR-07-001, 2007.
- [76] Daniel Lemire and Owen Kaser, One-pass, one-hash n-gram statistics estimation, UNBSJ CSAS Technical Report TR-06-001, 2006.
- [77] Steven Keith, Owen Kaser, Daniel Lemire, Analyzing large collections of electronic text using OLAP, UNBSJ CSAS Technical Report TR-05-001, 2005.
- [78] Jean-Michel Nonglaton, Franco Lenardon, and Daniel Lemire, Wavelet shrinkage of LINAC III and protons synchrotron booster transformers by the Haar Transform, CERN Technical Report Number AB-Note-2003-033, avril 2003, VNRC 45816.
- [79] Daniel Lemire, Local interpolation by high resolution subdivision schemes. Technical Report 200205-01, Acadia University, 2002.

Lecture Notes

- [INF9004] Daniel Lemire, Informatique des entrepôts de données (Data Warehousing and Computer Science), 2012. Online course: <http://benhur.telug.ca/SPIP/inf9004/>.
- [INF6104] Daniel Lemire, Recherche d'informations et web (Information Retrieval and the Web), 2008. Online course: <http://benhur.telug.ca/SPIP/inf6104/>.

- [INF6460] Daniel Lemire, Recherche et filtrage d'informations (Information Retrieval and Filtering), 2007. Online course: <http://benhur.telug.ca/SPIP/inf6460/>.
- [INF6450] Daniel Lemire, Gestion de l'information avec XML (Information Management with XML), ISBN 2762418747, 2005. (Revised in 2006, 2007 and 2008.) Online course: <http://www.telug.ca/inf6450/>.

Open Source Software (sample)

I am a beneficiary of the Google Open Source Peer Bonus Program.

- Daniel Lemire et al., Roaring Bitmaps, 2014. (<https://github.com/RoaringBitmap/RoaringBitmap>)
This library is used by Apache Spark, Apache Kylin (eBay's OLAP engine) and Druid. A reimplementaion of this library is part of Apache Lucene, and used by systems such as Solr and Elastic.
- Daniel Lemire et al., JavaFastPFOR: A simple integer compression library in Java, 2012. (<https://github.com/lemire/JavaFastPFOR>)
A derivative of this library is part of Apache Lucene.
- Daniel Lemire et al., The FastPFOR C++ library: Fast integer compression, 2012. (<https://github.com/lemire/FastPFor>)
- Daniel Lemire et al., JavaEWAH, A compressed alternative to the Java BitSet class, 2009. (<http://code.google.com/p/javaewah/>)
This project was recognized by Google's Open Source Programs in 2012.
JavaEWAH is used by Apache Hive for bitmap indexing, **part of major Linux distributions** like Redhat and Ubuntu. It is used by JGit.
- Daniel Lemire, EWAHBoolArray: A compressed bitmap class in C++, 2008. (<http://code.google.com/p/lemurbitmapindex/>)
Used by the OLAP engine Hustle.
Ported to C by GitHub and integrated in the Git version control system.

Hobbies

- My favorite science-fiction authors are Charles Stross, Peter F. Hamilton and Alastair Reynolds.
- My family loves geocaching.
- I build radio-controlled sailboat models.
- Since I was twelve years old, I have been programming computers for fun.