# Amélie MARIAN

## Curriculum Vitae

October 2021

#### Research Interests

Explainable Rankings, Accountable and Fair Decision-making Systems, Personal Digital Traces, Data Integration.

## Professional Appointments

2005-present Rutgers University, Computer Science Department.

- o 2012-present: Associate Professor
- o 2014–2020: Undergraduate Director
- o 2005–2012: Assistant Professor

2012–present **Rutgers University**, *Electrical and Computer Engineering Department*, Member of the Graduate Faculty.

- 2000-2005 **Columbia University**, Computer Science Department, Graduate Research Assistant.
  - 2004 AT&T Research Labs. Summer Intern.
- 1999-2000 INRIA Rocquencourt, France, Project VERSO, Research Intern.
- 1998-2000 Pôle Universitaire Léonard de Vinci, Paris La Defense, France, Instructor.

#### Education

- 2005 **Ph.D., Computer Science**, *Columbia University*, NY, USA.
  - Evaluation of Top-k Queries over Structured and Semi-structured Data
  - Advisor: Professor Luis Gravano
- 1999 **D.E.A. (M.Sc. equivalent), Computer Science**, *Université Paris-IX Dauphine*, France. Détection de Changements et Mises-à-jour Incrémentales dans un Système de Vues Actives pour XML Advisor: Serge Abiteboul
- 1998 **Maîtrise (B.Sc. equivalent), Computer Science and Management**, *Université Paris-IX Dauphine*, France.

#### Awards and Grants

2018 Rutgers Research Council, \$1,130.

Connecting, Managing, and Understanding Personal Digital Traces

- 2010-2014 **NSF CDI-Type I**, with Noémie Elhadad, Columbia Biomedical Informatics, BCS-10-27801, \$729,556 (\$302,268 for Rutgers side).
  - Gaining Knowledge from Other Patients: Structuring and Searching the content of Health-Related Web Posts
- 2009-2014 **NSF CAREER Award**, IIS-08-44935, \$499,759.
  - Relaxed Content and Structure Queries over Heterogeneous Data
  - 2012 **Google Research Award**, with Thu D. Nguyen, Rutgers University, \$62,500. Remembrance of Data Past: Using Context in Personal Information Search
  - 2010 **Google Research Award**, with Noémie Elhadad, Columbia Biomedical Informatics, \$77,000 (\$25,000 for Rutgers side).
    - PERSEUS: Structuring and Searching the Content of Health-Related Web Posts
  - 2008 **Google Research Award**, with Noémie Elhadad, Columbia Biomedical Informatics, \$85,000 (\$60,000 for Rutgers side).
    - URSA: Understanding User Reviewing Patterns
  - 2006 Microsoft Research Award, \$48,000.
    - The Truth Is out There: Aggregating Answers From Multiple Web Sources

#### Invited Talks

- 2019 Algorithms in the Wild: A case for Transparency, Accountability and Error Mitigation, Social Responsibility of Algorithms SRA 2019.
- 2019 Explainable Decision-Making, Panel, EARS Workshop at SIGIR 2019.
- 2018 **DigitalSelf: Connecting, Understanding and Managing Personal Digital Traces**, *New York DB Day 2018*.
- 2018 Understanding and Managing Connected Personal Digital Traces, Modas workshop on Web Data Management.
- 2017 Small Data, Panel, ICDE 2017.
- 2016 A la recherche des données perdues: gestion des informations numériques personnelles, Keynote, Colloque TRANSLIT.
- 2016 Managing your Personal Information, Dagstuhl Seminar on "Data, Responsibly".
- 2015 Accommoder les miettes de donnees : Ingredients, Recettes et Astuces, Keynote, BDA 2015, with Arnaud Sahuguet.
- 2015 **Personal Information Systems**, *Tutorial*, *EDBT/ICDT Conference*, with Serge Abiteboul.
- 2015 **Personal Information Search and Recovery**, Keynote, 3rd Singaporean-French Symposium.
- 2013 **Searching Web Forums**, WebDam workshop on Web Data Management.
- 2012 Remembrance of Data Past, WebDam-Modas workshop on Web Data Management.
- 2011 Searching Data with Substance and Style, Technicolor Research Center.
- 2009–2010 **Beyond the Stars: Improving Rating Predictions using Review Text Content**, *University of Pennsylvania; I.N.R.I.A. Gemo; LIP6, Université Pierre et Marie Curie Paris*.
  - 2008 **Multi-Dimensional Search for Personal Information Systems**, Dagstuhl Seminar on Ranked Query Processing.

- 2007 The Truth is Out There: Aggregating Answers from Multiple Web Sources, *Microsoft Search Summit.*
- 2006–2007 **Improving Query Results using Answer Corroboration**, I.N.R.I.A. Gemo; Ask.com; Brooklyn Polytechnic; NJIT.
  - 2006 Ranking and Scoring XML Data, LAMSADE, Université Paris Dauphine; LIP6, Université Pierre et Marie Curie Paris.
  - 2005 Structure and Content Scoring for XML, DIMACS Mixer Series.
  - 2005 **Evaluating Top-k Queries over Structured and Semi-structured Data**, University of Connecticut; IBM Watson; AT&T Labs Research; Rutgers University.

#### **Publications**

#### **Invited Papers**

- [11] Amélie Marian. Top-k selection queries on multimedia datasets. In Ling Liu and M. Tamer Özsu, editors, *Encyclopedia of Database Systems, Second Edition*. Springer, 2018 and 2009.
- [12] Amélie Marian, Ralf Schenkel, and Martin Theobald. Ranked XML processing. In Ling Liu and M. Tamer Özsu, editors, *Encyclopedia of Database Systems, Second Edition*. Springer, 2018 and 2009.
- [13] Amélie Marian and Arnaud Sahuguet. Reconstitution des données personnelles. In Mokrane Bouzeghoub and Rémy Mosseri, editors, *Les big data à découvert*. CNRS, 2017.
- [14] Amélie Marian and Wei Wang. Flexible querying of personal information. *IEEE Data Eng. Bull.*, 32(2):20–27, 2009.

#### Papers in Refereed Journals

- [J1] Gayatree Ganu, Yogesh Kakodkar, and Amélie Marian. Improving the quality of predictions using textual information in online user reviews. *Inf. Syst.*, 38(1):1–15, 2013.
- [J2] Wei Wang, Christopher Peery, Amélie Marian, and Thu D. Nguyen. Efficient multidimensional fuzzy search for personal information management systems. *IEEE Trans. Knowl. Data Eng.*, 24(9):1584–1597, 2012.
- [J3] Minji Wu and Amélie Marian. A framework for corroborating answers from multiple web sources. *Inf. Syst.*, 36(2):431–449, 2011.
- [J4] Surajit Chaudhuri, Luis Gravano, and Amélie Marian. Optimizing top-k selection queries over multimedia repositories. *IEEE Trans. Knowl. Data Eng.*, 16(8):992–1009, 2004.
- [J5] Amélie Marian, Nicolas Bruno, and Luis Gravano. Evaluating top-k queries over web-accessible databases. ACM Trans. Database Syst., 29(2):319–362, 2004.

#### Papers in Refereed Conferences and Workshops

- [C1] Varvara Kalokyri, Alexander Borgida, and Amélie Marian. Supporting human memory by reconstructing personal episodic narratives from digital traces. *Accepted for Publication at the International AAAI Conference on Web and Social Media ICWSM'22*, 2022.
- [C2] Daniela Vianna and Amélie Marian. A frequency-based learning-to-rank approach for personal digital traces. Accepted for Publication at the 55th Hawaii International Conference on System Sciences HICSS'22, 2022.
- [C3] Shuchang Liu, Shuyuan Xu, Wenhui Yu, Zuohui Fu, Yongfeng Zhang, and Amélie Marian. Fedct: Federated collaborative transfer for recommendation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval SIGIR'21*, 2021.
- [C4] Abraham Gale and Amélie Marian. Explaining monotonic ranking functions. *Proceedings of the VLDB Endowment VLDB'21*, 14(4):640–652, 2020.
- [C5] Alexander Borgida, Varvara Kalokyri, and Amélie Marian. Description logics and specialization for structured BPMN. In Business Process Management Workshops - BPM 2019 International Workshops, Vienna, Austria, September 1-6, 2019, Revised Selected Papers, volume 362 of Lecture Notes in Business Information Processing, pages 19–31. Springer, 2019.
- [C6] Daniela Vianna, Varvara Kalokyri, Alexander Borgida, Thu D. Nguyen, and Amélie Marian. Searching heterogeneous personal digital traces. Association for Information Science and Technology (ASIS&T), 2019.
- [C7] Abraham Gale and Amélie Marian. Metrics for explainable ranking functions. *EARS International Workshop on ExplainAble Recommendation and Search at SIGIR, 2019.*
- [C8] Varvara Kalokyri, Alexander Borgida, and Amélie Marian. YourDigitalSelf: A personal digital trace integration tool. (Demonstration). In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM'18, Torino, Italy, October 22-26, 2018, pages 1963–1966. ACM, 2018.
- [C9] Varvara Kalokyri, Alexander Borgida, Amélie Marian, and Daniela Vianna. Semantic modeling and inference with episodic organization for managing personal digital traces. In ODBASE OTM Confederated International Conferences" On the Move to Meaningful Internet Systems", pages 273–280. Springer, 2017.
- [C10] Varvara Kalokyri, Alexander Borgida, Amélie Marian, and Daniela Vianna. Integration and exploration of connected personal digital traces. In *Proceedings of the ExploreDB'17, Chicago, IL, USA, May 19, 2017*, pages 3:1–3:6. ACM, 2017.
- [C11] Minji Wu and Amélie Marian. Corroborating facts from affirmative statements. In Proceedings of the 17th International Conference on Extending Database Technology, EDBT'14, Athens, Greece, March 24-28, 2014, pages 157–168.
- [C12] Gayatree Ganu and Amélie Marian. Personalizing forum search using multidimensional random walks. In *Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM'14*, Ann Arbor, Michigan, USA, June 1-4, 2014. The AAAI Press, 2014.

- [C13] Daniela Vianna, Alicia-Michelle Yong, Chaolun Xia, Amélie Marian, and Thu D. Nguyen. A tool for personal data extraction. In Workshops Proceedings of the 30th International Conference on Data Engineering Workshops, ICDE 2014, Chicago, IL, USA, March 31 - April 4, 2014, pages 80–83. IEEE Computer Society, 2014.
- [C14] Gayatree Ganu and Amélie Marian. One size does not fit all: multi-granularity search of web forums. In 22nd ACM International Conference on Information and Knowledge Management, CIKM'13, San Francisco, CA, USA, October 27 November 1, 2013, pages 9–18. ACM, 2013.
- [C15] Wei Wang, Amélie Marian, and Thu D. Nguyen. Unified structure and content search for personal information management systems. In *EDBT'11*, 14th International Conference on Extending Database Technology, Uppsala, Sweden, March 21-24, 2011, Proceedings, pages 201–212. ACM, 2011.
- [C16] Émilien Antoine, Alban Galland, Kristian Lyngbaek, Amélie Marian, and Neoklis Polyzotis. Social networking on top of the webdamexchange system. (Demonstration). In *Proceedings* of the 27th International Conference on Data Engineering, ICDE'11, April 11-16, 2011, Hannover, Germany, pages 1300–1303. IEEE Computer Society, 2011.
- [C17] Serge Abiteboul, Sihem Amer-Yahia, Alban Galland, Amélie Marian, and Pierre Senellart. Birds of a tag flock together. In *Proceedings of the Third Annual Workshop on Search in Social Media (SSM'10)*, 2010.
- [C18] Minji Wu, Laure Berti-Équille, Amélie Marian, Cecilia M. Procopiuc, and Divesh Srivastava. Processing top-k join queries. Proceedings of the VLDB Endowment VLDB'10, 3(1):860–870, 2010.
- [C19] Alban Galland, Serge Abiteboul, Amélie Marian, and Pierre Senellart. Corroborating information from disagreeing views. In *Proceedings of the Third International Conference on Web Search and Web Data Mining*, WSDM'10, New York, NY, USA, February 4-6, 2010, pages 131–140. ACM, 2010.
- [C20] Laure Berti-Équille, Anish Das Sarma, Xin Dong, Amélie Marian, and Divesh Srivastava. Sailing the information ocean with awareness of currents: Discovery and application of source dependence. In CIDR'09, Fourth Biennial Conference on Innovative Data Systems Research, Asilomar, CA, USA, January 4-7, 2009, Online Proceedings. www.cidrdb.org, 2009.
- [C21] Gayatree Ganu, Noemie Elhadad, and Amélie Marian. Beyond the stars: Improving rating predictions using review text content. In 12th International Workshop on the Web and Databases, WebDB 2009, Providence, Rhode Island, USA, June 28, 2009, 2009.
- [C22] Christopher Peery, Wei Wang, Amélie Marian, and Thu D. Nguyen. Multi-dimensional search for personal information management systems. In *EDBT'08*, 11th International Conference on Extending Database Technology, Nantes, France, March 25-29, 2008, Proceedings, volume 261 of ACM International Conference Proceeding Series, pages 464–475. ACM, 2008.
- [C23] Christopher Peery, Wei Wang, Amélie Marian, and Thu D. Nguyen. Fuzzy multi-dimensional search in the wayfinder file system. In *Proceedings of the 24th International Conference* on Data Engineering, ICDE'08, April 7-12, 2008, Cancún, Mexico, pages 1588–1591. IEEE Computer Society, 2008.

- [C24] Minji Wu and Amélie Marian. Corroborating answers from multiple web sources. In *Tenth International Workshop on the Web and Databases, WebDB 2007, Beijing, China, June 15, 2007, 2007.*
- [C25] Kenneth A. Ross, Peter J. Stuckey, and Amélie Marian. Practical preference relations for large data sets. In Proceedings of the 23rd International Conference on Data Engineering Workshops, (DBRank '07) ICDE 2007, 15-20 April 2007, Istanbul, Turkey, pages 229–236. IEEE Computer Society, 2007.
- [C26] Yannis Kotidis, Amélie Marian, and Divesh Srivastava. Circumventing data quality problems using multiple join paths. In *Proceedings of the First Int'l VLDB Workshop on Clean Databases, CleanDB 2006, September 11, 2006, Seoul, Korea (Co-located with VLDB 2006)*, 2006.
- [C27] Sihem Amer-Yahia, Nick Koudas, Amélie Marian, Divesh Srivastava, and David Toman. Structure and content scoring for XML. In *Proceedings of the 31st International Conference on Very Large Data Bases VLDB'05*, Trondheim, Norway, August 30 - September 2, 2005, pages 361–372. ACM, 2005.
- [C28] Amélie Marian, Sihem Amer-Yahia, Nick Koudas, and Divesh Srivastava. Adaptive processing of top-k queries in XML. In *Proceedings of the 21st International Conference on Data Engineering*, *ICDE'05*, 5-8 April 2005, Tokyo, Japan, pages 162–173. IEEE Computer Society, 2005.
- [C29] Amélie Marian and Jérôme Siméon. Projecting XML documents. In Proceedings of 29th International Conference on Very Large Data Bases, VLDB'03, Berlin, Germany, September 9-12, 2003, pages 213–224. Morgan Kaufmann, 2003.
- [C30] Mary F. Fernández, Jérôme Siméon, Byron Choi, Amélie Marian, and Gargi Sur. Implementing xquery 1.0: The galax experience. (Demonstration. In *Proceedings of 29th International Conference on Very Large Data Bases*, *VLDB'03*, Berlin, Germany, September 9-12, 2003, pages 1077–1080. Morgan Kaufmann, 2003.
- [C31] Gregory Cobena, Serge Abiteboul, and Amélie Marian. Detecting changes in XML documents. In *Proceedings of the 18th International Conference on Data Engineering*, *ICDE'02* San Jose, CA, USA, February 26 March 1, 2002, pages 41–52. IEEE Computer Society, 2002.
- [C32] Nicolas Bruno, Luis Gravano, and Amélie Marian. Evaluating top-k queries over web-accessible databases. In Proceedings of the 18th International Conference on Data Engineering, ICDE'02 San Jose, CA, USA, February 26 - March 1, 2002, pages 369–380. IEEE Computer Society, 2002.
- [C33] Amélie Marian, Serge Abiteboul, Gregory Cobena, and Laurent Mignet. Change-centric management of versions in an XML warehouse. In VLDB'01, Proceedings of 27th International Conference on Very Large Data Bases, September 11-14, 2001, Roma, Italy, pages 581–590. Morgan Kaufmann, 2001.
- [C34] Gregory Cobena, Serge Abiteboul, and Amélie Marian. Detecting changes in XML documents. In 17èmes Journées Bases de Données Avancées, BDA 2001, 29 octobre - 2 novembre, Agadir, Maroc, Actes (Informal Proceedings), 2001.

- [C35] Amélie Marian, Serge Abiteboul, and Laurent Mignet. Chance-centric management of versions in an XML warehouse. In 16èmes Journées Bases de Données Avancées, BDA 2000, 24-27 octobre 2000, Blois, France, Actes (Informal Proceedings), 2000.
- [C36] Laurent Mignet, Serge Abiteboul, Sébastien Ailleret, Bernd Amann, Amélie Marian, and Mihai Preda. Acquiring XML pages for a webhouse. In Anne Doucet, editor, 16èmes Journées Bases de Données Avancées, BDA 2000, 24-27 octobre 2000, Blois, France, Actes (Informal Proceedings), 2000.
- [C37] Serge Abiteboul, Vincent Aguilera, Sébastien Ailleret, Bernd Amann, Sophie Cluet, Brendan Hills, Frédéric Hubert, Jean-Claude Mamou, Amélie Marian, Laurent Mignet, Tova Milo, Cassio Souza dos Santos, Bruno Tessier, and Anne-Marie Vercoustre. XML repository and active views demonstration. In VLDB'99, Proceedings of 25th International Conference on Very Large Data Bases, September 7-10, 1999, Edinburgh, Scotland, UK, pages 742–745. Morgan Kaufmann, 1999.

## Teaching Experience

198:142 Data 101: Data Literacy, Instructor, Rutgers University, Fall 2021, Spring 2020.

**198:437** Systems Aspects of Information Management (Undergraduate level), *Instructor*, Rutgers University, Fall 2018, 2016, 2015, 2014.

**198:539** Systems Aspects of Information Management (Graduate level), *Instructor*, Rutgers University, Fall 2016, 2015, 2014.

**198:541 Database Systems (Graduate level)**, *Instructor*, Rutgers University, Spring 2017, 2016, 2014, 2010, 2008, 2007, 2006.

**198:336** Principles of Information and Data Management, *Instructor*, Rutgers University, Fall 2012, 2011, 2006.

**198:673 Web Data Management (Graduate Level)**, *Instructor*, Rutgers University, Fall 2008.

**198:500** Readings in Data Management, Light Seminar (Graduate level), *Instructor*, Rutgers University, Spring 2008, Fall 2006, Spring 2006.

## Student Advising

#### Current Ph.D. Advisees

- since 2014 Varvara (Valia) Kalokyri, co-advised with Alex Borgida.
- since 2017 **Schuchang Liu**.
- since 2018 Abraham (Yehuda) Gale.

Current Undergraduate Advisees

2021 Alborz Jelvani.

Graduated Ph.D. Advisees

- October 2019 Daniela Vianna.
- October 2015 Minji Wu, First Employment: Microsoft.
- January 2014 Gayatree Ganu, First Employment: Facebook.
- October 2010 Wei Wang, co-advised with Thu Nguyen, First Employment: EA.

#### Ph.D. Defense Committee

2019	Daniela Vianna (Chair)	2018	Matt Mitsui
2017	Long Le	2015	Minji Wu (Chair)
2014	Vukosi Marivate	2013	Gayatree Ganu (Chair)
2013	Imen Ben Dhia, Telecom ParisTech	2012	Nora Derouiche, Telecom ParisTech
2010	Smriti Bhagat	2010	Wei Wang (Chair)
2010	Geetha Jagannathan	2009	Christopher Peery
2007	Irina Rozenbaum	2007	Vladislav Shkapenyuk
	M.S. Defense Committee		
2011	Yogesh Kakodkar (Chair)	2009	Vikas Menon
2006	Preetham Mysore		

#### Professional Activities

**Agorithms in the Wild**, *Blog posts*, Wrote on the behaviors on public decision-making algorithms, https://medium.com/algorithms-in-the-wild.

**NYSIP-PLC Team**, *Volunteer member of the design team for the NYC Department of Education District 2 New York State Integration Plan*, The purpose of the plan is to increase student achievement in New York State Title I Schools, by encouraging greater integration by race and ethnicity, as well as socioeconomic, special education, and English Language Learner (ELL) status, Grant awarded in October 2020 (\$2,688,562).

Worked on the school composition component part of the grant to improve diversity in school admissions.

**Area Editor**, Information Systems 2012–2016.

Tutorial Chair, EDBT 2022, EDBT 2016.

Workshop co-Chair, WebDB 2011, DBRank 2010, VLDB Ph.D. Workshop 2010.

Workshops co-Chair, WWW 2010.

Publicity Chair, SIGMOD 2013.

Demonstrations Co-chair, CIKM 2012.

Co-organizer, Third New York Area DB/IR day, April 2006.

Senior Program Committees, SIGMOD 2016, WSDM 2013.

Program Committees, EEAMO 2022, AiDM 2021, EAAI 2021 (Student Competition), VLDB 2020 (Demonstrations), EDBT 2019, CIKM 2018, WebDB 2018, VLDB'18 PhD Workshop, ExploreDB'17, ICDE '16, SIGMOD 2015, VLDB 2014/PVLDB, EDBT 2014, BigData 2014, VLDB 2013/PVLDB, WSDM 2013, DBRank 2012, BDA 2012, ER 2012, SSDBM 2012, WSDM'2012, WWW 2012, EDBT 2012, ESAIR 2011, EDBT 2011, CIKM 2010) (Database Track), WebDB 2010, EDBT 2010, KEYS 2010, WebDB 2009, EC 2009, ECIR 2009 Workshop on Information Retrieval over Social Networks WSCD 2009, MPREF 2008, DBRank 2008, VLDB 2008, PIKM 2007, VLDB 2007 (Infrastructure for Information Systems Track), SIGMOD 2007 (Demonstrations), DBRank 2007, WIDM 2006, CIKM 2006 (Database Track), XSym 2006.

**Journal Referee**, DAPD 2009; VLDB Journal, 2009, 2008, 2007,2005; IS 2009, 2008, 2007, 2006; DKE 2007; ACM TODS, 2010, 2009, 2006; ACM TOMCCAP, 2006; IEEE TKDE, 2009, 2006, 2005, 2004; ACM TOIT, 2004; Journal of Computer and System Sciences, 2001.

**Conference Referee**, PODS 2006, SIGMOD 2005, SIGMOD 2004, SIGMETRICS 2004, ICDE 2004, SIGMOD 2002, and VLDB 2002 conference.

Grant Referee, Agence Nationale de la Recherche (ANR), 2006.

Grant Panel, ANR (France) 2018; NSF IIS, 2016, 2014, 2013, 2010, 2009, 2007.

## University Service

- 2021 MS Admissions Committee, Computer Science Dept. Rutgers University.
- 2014–2020 **Undergraduate Director**, *Computer Science Dept. Rutgers University*.
- 2014–2020 **Executive Committee**, Computer Science Dept. Rutgers University.
- 2014–2020 Undergraduate Advising Committee (Chair), Computer Science Dept. Rutgers University.
- 2012–2020 **Undergraduate Committee (Chair 2014–2020)**, Computer Science Dept. Rutgers University.
- 2013–2014 Academic Planning Committee, Computer Science Dept. Rutgers University.
- 2008–2012 **Graduate Committee**, Computer Science Dept. Rutgers University, (also 2005-2006).
  - 2010 Rutgers Day Coordinator, Computer Science Dept. Rutgers University.
  - 2009 **Hiring Committee**, Computer Science Dept. Rutgers University.
- 2006–2008 **Graduate Admissions Committee**, Computer Science Dept. Rutgers University.
- 2000–2004 **Ph.D. admissions Committee (Student member)**, Computer Science Dept. Columbia University.
- 1997–1998 Elected Student Representative, MIAGE Council, Université Paris Dauphine.