

# Pierre Senellart

Professor in computer science

Databases, Web

32 villa de Lourcine  
75014 Paris, France

☎ (+33) 6-48-36-36-20

✉ pierre@senellart.com

<http://pierre.senellart.com/>

French, born 19 June 1981.



## Academic career

- September 2016 – Present **Professor**, *École normale supérieure*, Paris, France.  
*Computer Science Department.*
- August 2014 – July 2016 **Senior research fellow**, *National University of Singapore*, Singapore, Singapore.  
*Computer Science department, IPAL laboratory.*
- January 2013 – August 2016 **Professor**, *Télécom ParisTech*, Paris, France.  
*Network and Computer Science department.*
- November 2012 – April 2013 **Visiting researcher and lecturer**, *The University of Hong Kong*, Hong Kong, China.  
*Computer science department*
- June 2008 – December 2012 **Associate professor**, *Télécom ParisTech*, Paris, France.  
*Network and Computer Science department.*
- September 2008 – December 2008 **Teaching assistant**, *École polytechnique*, Palaiseau, France.  
*Database labs.*
- February 2008 – May 2008 **Teaching assistant**, *École normale supérieure de Cachan*, Cachan, France.  
*Database labs.*
- February 2008 – May 2008 **Post-doctoral researcher**, *Max-Planck-Institut für Informatik*, Saarbrücken, Germany.  
*Databases and Information Systems department.*
- September 2005 – May 2008 **PhD candidate and teaching assistant**, *Université Paris–Sud & INRIA Futurs*, Orsay, France.  
Topic: Understanding the hidden Web. Advisor: Serge Abiteboul. 64hr of teaching a year.
- March 2007 – May 2007 **Visiting research student**, *Oxford University, Department of computer science*, Oxford, United Kingdom.  
Topic: Data exchange Advisor: Georg Gottlob.
- 2003 – 2005 **Part-time teaching assistant**, *École nationale des ponts et chaussées*, Marne-la-Vallée, France.  
C++ course for first-year students (2 semesters, 60 hour each).
- March 2003 – August 2003 **Research intern (master)**, *INRIA Futurs*, Orsay, France.  
Topic: Identifying logical websites. Advisor: Serge Abiteboul.
- February 2002 – June 2002 **Research intern**, *Chalmers Tekhniska Högskola*, Göteborg, Sweden.  
Topic: Automatic verification of multipliers. Advisor: Mary Sheeran.
- June 2001 – July 2001 **Research intern**, *Université catholique de Louvain*, Louvain-la-neuve, Belgium.  
Topic: Automatic extraction of similar words in a dictionary. Advisor: Vincent Blondel.

## Education

- June 2012 **Habilitation to supervise research in Computer science**, *Université Pierre et Marie Curie*, Paris, France.  
Topic: Probabilistic XML: A Data Model for the Web
- December 2007 **PhD in Computer science**, *Université Paris–Sud*, Orsay, France.  
Topic: Understanding the Hidden Web. Mention Très Honorable.

- September 2003 **Magistère (postgraduate diploma) in Fundamental and applied mathematics and Computer science**, *Université Paris–Sud & ENS*, Orsay & Paris, France.
- September 2003 **DEA (M.Sc.) in Computer science**, *Université Paris–Sud*, Orsay, France.  
Mention Très Bien (summa cum laude).
- September 2001 **Maîtrise (undergraduate diploma) in Computer science**, *Université Paris–Sud & ENS*, Orsay & Paris, France.  
Mention Très Bien (summa cum laude).
- June 2001 **Licence (B.Sc.) in Computer science**, *Université Paris–Sud & ENS*, Orsay & Paris, France.  
Mention Très Bien (summa cum laude).
- 2000 – 2005 **École normale supérieure (Paris)**.
- June 1998 **Baccalauréat S (scientific high school diploma)**, *Lycée Mariette*, Boulogne-sur-mer, France.  
Mention Très Bien (summa cum laude).

## Publications

We mention here international publications only; for a complete list (including national conferences and journals, technical reports, unrefereed material, preprints), see <http://pierre.senellart.com/publications/>.

### Book

Serge Abiteboul, Ioana Manolescu, Marie-Christine Rousset, Philippe Rigaux, and Pierre Senellart. *Web Data Management*. Cambridge University Press, New York, USA, January 2012.

### Book chapters (peer-reviewed)

Thomas Risse, Wim Peters, Pierre Senellart, and Diana Maynard. Documenting contemporary society by preserving relevant information from twitter. In Katrin Weller, Axel Bruns, Jean Burgess, Merja Mahrt, and Cornelius Puschmann, editors, *Twitter and Society*, pages 207–219. Peter Lang International Academic Publishers, November 2013.

Evgeny Kharlamov and Pierre Senellart. Modeling, querying, and mining uncertain XML data. In Andrea Tagarelli, editor, *XML Data Mining: Models, Methods, and Applications*, pages 29–52. IGI Global, November 2011.

### Book chapters (without peer review)

Pierre Senellart. Archivage du web. In Mokrane Bouzhegoub and Rémy Mosseri, editors, *Les Big Data à découvert*, pages 200–201. CNRS Éditions, March 2017.

Benny Kimelfeld and Pierre Senellart. Probabilistic XML: Models and complexity. In Zongmin Ma and Li Yan, editors, *Advances in Probabilistic Databases for Uncertain Information Management*, pages 39–66. Springer-Verlag, May 2013.

Michael Benedikt and Pierre Senellart. Databases. In Edward K. Blum and Alfred V. Aho, editors, *Computer Science. The Hardware, Software and Heart of It*, pages 169–229. Springer-Verlag, January 2012.

Pierre Senellart and Vincent D. Blondel. Automatic discovery of similar words. In Michael W. Berry and Malu Castellanos, editors, *Survey of Text Mining II: Clustering, Classification and Retrieval*, pages 25–44. Springer-Verlag, January 2008.

Pierre Senellart and Vincent D. Blondel. Automatic discovery of similar words. In Michael W. Berry, editor, *Survey of Text Mining: Clustering, Classification and Retrieval*, pages 25–43. Springer-Verlag, August 2003.

### Journal articles (peer-reviewed)

Silviu Maniu, Reynold Cheng, and Pierre Senellart. An indexing framework for queries on probabilistic graphs. *ACM Transactions on Database Systems*, 42(2):13:1–13:34, May 2017.

Debabrota Basu, Qian Lin, Weidong Chen, Hoang Tam Vo, Zihong Yuan, Pierre Senellart, and Stéphane Bressan. Regularized cost-model oblivious database tuning with reinforcement learning. *Transactions on Large-Scale Data and Knowledge-Centered Systems*, 28:96–132, 2016.

Ruiming Tang, Antoine Amarilli, Pierre Senellart, and Stéphane Bressan. A framework for sampling-based XML data pricing. *Transactions on Large-Scale Data and Knowledge-Centered Systems*, 24:116–138, 2016.

Serge Abiteboul, Yael Amsterdamer, Daniel Deutch, Tova Milo, and Pierre Senellart. Optimal probabilistic generation of XML documents. *Theory of Computing Systems*, 57(4):806–842, November 2015.

Thomas Risse, Elena Demidova, Stefan Dietze, Wim Peters, Nikos Papailiou, Katerina Doka, Yannis Stavarakas, Vassilis Plachouras, Pierre Senellart, Florent Carpentier, Amin Mantrach, Bogdan Cautis, Patrick Siehndel, and Dimitris Spiliotopoulos. The ARCOMEM architecture for social and semantic driven web archiving. *Future Internet*, 6(4):688–716, November 2014.

Vassilis Plachouras, Florent Carpentier, Muhammad Faheem, Julien Masanès, Thomas Risse, Pierre Senellart, Patrick Siehndel, and Yannis Stavarakas. ARCOMEM crawling architecture. *Future Internet*, 6(3):518–541, August 2014.

Fabian M. Suchanek, Serge Abiteboul, and Pierre Senellart. PARIS: Probabilistic alignment of relations, instances, and schema. *Proceedings of the VLDB Endowment*, 5(3):157–168, December 2011. Presented at the VLDB 2012 conference, Istanbul, Turkey. Also CoRR abs/1111.7164.

Benjamin Nguyen, Antoine Vion, François-Xavier Dudouet, Dario Colazzo, Ioana Manolescu, and Pierre Senellart. XML content warehousing: Improving sociological studies of mailing lists and web data. *Bulletin of Sociological Methodology*, 112(1):5–31, October 2011. Also CoRR abs/1108.4596.

Serge Abiteboul, T-H. Hubert Chan, Evgeny Kharlamov, Werner Nutt, and Pierre Senellart. Capturing continuous data and answering aggregate queries in probabilistic XML. *ACM Transactions on Database Systems*, 36(4), 2011.

Michael Benedikt, Evgeny Kharlamov, Dan Olteanu, and Pierre Senellart. Probabilistic XML via Markov chains. *Proceedings of the VLDB Endowment*, 3(1):770–781, September 2010. Presented at the VLDB 2010 conference, Singapore.

Georg Gottlob and Pierre Senellart. Schema mapping discovery from data instances. *Journal of the ACM*, 57(2), January 2010.

Serge Abiteboul, Benny Kimelfeld, Yehoshua Sagiv, and Pierre Senellart. On the expressiveness of probabilistic XML models. *VLDB Journal*, 18(5):1041–1064, October 2009.

Vincent D. Blondel, Anahí Gajardo, Maureen Heymans, Pierre Senellart, and Paul Van Dooren. A measure of similarity between graph vertices: applications to synonym extraction and Web searching. *SIAM Review*, 46(4):647–666, 2004. Also CoRR cs.IR/0407061.

#### Conference articles (peer-reviewed)

Miyoung Han, Pierre Senellart, and Pierre-Henri Wuillemin. Focused crawling through reinforcement learning. In *Proc. ICWE*, June 2018.

Ben Spencer, Michael Benedikt, and Pierre Senellart. Form filling based on constraint solving. In *Proc. ICWE*, June 2018.

Antoine Amarilli, Mikaël Monet, and Pierre Senellart. Connecting width and structure in knowledge compilation. In *Proc. ICDT*, Vienna, Austria, March 2018.

Antoine Amarilli, M. Lamine Ba, Daniel Deutch, and Pierre Senellart. Possible and certain answers for queries over order-incomplete data. In *Proc. TIME*, pages 4:1–4:19, Mons, Belgium, October 2017.

Antoine Amarilli, Mikaël Monet, and Pierre Senellart. Conjunctive queries on probabilistic graphs: Combined complexity. In *Proc. PODS*, pages 217–232, Chicago, USA, May 2017.

Antoine Amarilli, Pierre Bourhis, Mikaël Monet, and Pierre Senellart. Combined tractability of query evaluation via tree automata and cycluits. In *Proc. ICDT*, pages 6:1–6:19, Venice, Italy, March 2017.

Antoine Amarilli, Yael Amsterdamer, Tova Milo, and Pierre Senellart. Top-k queries on unknown values under order constraints. In *Proc. ICDT*, pages 5:1–5:18, Venice, Italy, March 2017.

Miyoung Han, Pierre Senellart, Stéphane Bressan, and Huayu Wu. Routing an autonomous taxi with reinforcement learning. In *Proc. CIKM*, Indianapolis, USA, October 2016. Industry track, short paper.

Antoine Amarilli, Pierre Bourhis, and Pierre Senellart. Tractable lineages on treelike instances: Limits and extensions. In *Proc. PODS*, pages 355–370, San Francisco, USA, June 2016.

Muhammad Faheem and Pierre Senellart. Adaptive web crawling through structure-based link classification. In *Proc. ICADL*, pages 39–51, Seoul, South Korea, December 2015.

David Montoya, Serge Abiteboul, and Pierre Senellart. Hup-me: Inferring and reconciling a timeline of user activity with smartphone and personal data. In *Proc. SIGSPATIAL*, Seattle, USA, November 2015. Short paper.

Debabrota Basu, Qian Lin, Weidong Chen, Hoang Tam Vo, Zihong Yuan, Pierre Senellart, and Stéphane Bressan. Cost-model oblivious database tuning with reinforcement learning. In *Proc. DEXA*, pages 253–268, Valencia, Spain, September 2015.

Siyu Lei, Silviu Maniu, Luyi Mo, Reynold Cheng, and Pierre Senellart. Online influence maximization. In *Proc. KDD*, pages 645–654, Sydney, Australia, August 2015.

Antoine Amarilli, Pierre Bourhis, and Pierre Senellart. Provenance circuits for trees and treelike instances. In *Proc. ICALP*, pages 56–68, Kyoto, Japan, July 2015.

Changping Meng, Reynold Cheng, Silviu Maniu, Pierre Senellart, and Wangda Zhang. Discovering meta-paths in large heterogeneous information networks. In *Proc. WWW*, pages 754–764, Florence, Italy, May 2015.

Georges Gouriten, Silviu Maniu, and Pierre Senellart. Scalable, generic, and adaptive systems for focused crawling. In *Proc. Hypertext*, pages 35–45, Santiago, Chile, September 2014. Douglas Engelbart Best Paper Award.

Ruiming Tang, Antoine Amarilli, Pierre Senellart, and Stéphane Bressan. Get a sample for a discount: Sampling-based XML data pricing. In *Proc. DEXA*, pages 20–34, Munich, Germany, September 2014.

M. Lamine Ba, Talel Abdesslem, and Pierre Senellart. Uncertain version control in open collaborative editing of tree-structured documents. In *Proc. DocEng*, pages 27–36, Florence, Italy, September 2013.

Muhammad Faheem and Pierre Senellart. Intelligent and adaptive crawling of Web applications for Web archiving. In *Proc. ICWE*, pages 306–322, Aalborg, Denmark, July 2013.

Yael Amsterdamer, Yael Grossman, Tova Milo, and Pierre Senellart. Crowd mining. In *Proc. SIGMOD*, pages 241–252, New York, USA, June 2013.

Asma Souihli and Pierre Senellart. Optimizing approximations of DNF query lineage in probabilistic XML. In *Proc. ICDE*, pages 721–732, Brisbane, Australia, April 2013.

Michael Benedikt, Pierre Bourhis, and Pierre Senellart. Monadic datalog containment. In *Proc. ICALP*, pages 79–91, Warwick, United Kingdom, July 2012.

Jiaheng Lu, Pierre Senellart, Chunbin Lin, Xiaoyong Du, Shan Wang, and Xinxing Chen. Optimal top-k generation of attribute combinations based on ranked lists. In *Proc. SIGMOD*, pages 409–420, Scottsdale, USA, May 2012.

Georges Gouriten and Pierre Senellart. API Blender: A uniform interface to social platform APIs. In *Proc. WWW*, Lyon, France, April 2012. Developer track. Also CoRR abs/1301.2086.

Serge Abiteboul, Pierre Senellart, and Victor Vianu. The ERC Webdam on foundations of Web data management. In *Proc. WWW*, pages 211–214, Lyon, France, April 2012. European project track.

Serge Abiteboul, Yael Amsterdamer, Daniel Deutch, Tova Milo, and Pierre Senellart. Finding optimal probabilistic generators for XML collections. In *Proc. ICDT*, pages 127–139, Berlin, Germany, March 2012.

Michael Benedikt, Georg Gottlob, and Pierre Senellart. Determining relevance of accesses at runtime. In *Proc. PODS*, pages 211–222, Athens, Greece, June 2011.

Serge Abiteboul, T-H. Hubert Chan, Evgeny Kharlamov, Werner Nutt, and Pierre Senellart. Aggregate queries for discrete and continuous probabilistic XML. In *Proc. ICDT*, pages 50–61, Lausanne, Switzerland, March 2010.

Alban Galland, Serge Abiteboul, Amélie Marian, and Pierre Senellart. Corroborating information from disagreeing views. In *Proc. WSDM*, pages 131–140, New York, USA, February 2010.

Pierre Senellart and Georg Gottlob. On the complexity of deriving schema mappings from database instances. In *Proc. PODS*, pages 23–32, Vancouver, Canada, June 2008.

Yann Ollivier and Pierre Senellart. Finding related pages using Green measures: An illustration with Wikipedia. In *Proc. AAAI*, pages 1427–1433, Vancouver, Canada, July 2007.

Pierre Senellart and Serge Abiteboul. On the complexity of managing probabilistic XML data. In *Proc. PODS*, pages 283–292, Beijing, China, June 2007.

Serge Abiteboul and Pierre Senellart. Querying and updating probabilistic information in XML. In *Proc. EDBT*, pages 1059–1068, Munich, Germany, March 2006. Short paper.

François-Xavier Dudouet, Ioana Manolescu, Benjamin Nguyen, and Pierre Senellart. XML warehousing meets sociology. In *Proc. IADIS ICWI*, pages 170–174, Lisbon, Portugal, October 2005.

Mats Attnäs, Pierre Senellart, and Jean Senellart. Integration of SYSTRAN MT systems in an open workflow. In *Proc. MT Summit*, Phuket, Thailand, September 2005.

Pierre Senellart. Identifying Websites with flow simulation. In *Proc. ICWE*, pages 124–129, Sydney, Australia, July 2005.

### Workshop articles (peer-reviewed)

- David Montoya, Thomas Pellissier Tanon, Serge Abiteboul, Pierre Senellart, and Fabian M. Suchanek. A knowledge base for personal information management. In *Proc. LDOW*, Lyon, France, April 2018.
- Ognjen Savkovic, Evgeny Kharlamov, Werner Nutt, and Pierre Senellart. Towards approximating incomplete queries over partially complete databases. In *Proc. AMW*, Montevideo, Uruguay, June 2017.
- M. Lamine Ba, Roxana Horincar, Pierre Senellart, and Huayu Wu. Truth finding with attribute partitioning. In *Proc. WebDB*, pages 27–33, Melbourne, Australia, May 2015.
- Marilena Oita and Pierre Senellart. FOREST: Focused Object Retrieval by Exploiting Significant Tag paths. In *Proc. WebDB*, pages 55–61, Melbourne, Australia, May 2015.
- Silviu Maniu, Reynold Cheng, and Pierre Senellart. ProbTree: A query-efficient representation of probabilistic graphs. In *Proc. BUDA*, Snowbird, USA, June 2014. Workshop without formal proceedings.
- M. Lamine Ba, Talel Abdessalem, and Pierre Senellart. Merging uncertain multi-version XML documents. In *Proc. DChanges*, Florence, Italy, September 2013.
- Thomas Risse, Wim Peters, and Pierre Senellart. Social and semantic driven web harvesting. In *Proc. Building Web Observatories*, Paris, France, May 2013.
- Marilena Oita, Antoine Amarilli, and Pierre Senellart. Cross-fertilizing deep Web analysis and ontology enrichment. In *Proc. VLDS*, pages 5–8, Istanbul, Turkey, August 2012. Vision article.
- Meghyn Bienvenu, Daniel Deutch, Davide Martinenghi, Pierre Senellart, and Fabian M. Suchanek. Dealing with the deep Web and all its quirks. In *Proc. VLDS*, pages 21–24, Istanbul, Turkey, August 2012. Vision article.
- M. Lamine Ba, Talel Abdessalem, and Pierre Senellart. Towards a version control model with uncertain data. In *Proc. PIKM*, Glasgow, United Kingdom, October 2011.
- Marilena Oita and Pierre Senellart. Deriving dynamics of web pages: A survey. In *Proc. TWAU*, pages 25–32, Hyderabad, India, March 2011.
- Evgeny Kharlamov, Werner Nutt, and Pierre Senellart. Value joins are expensive over (probabilistic) XML. In *Proc. LID*, pages 41–48, Uppsala, Sweden, March 2011.
- Marilena Oita and Pierre Senellart. Archiving data objects using Web feeds. In *Proc. IWAU*, pages 31–41, Vienna, Austria, September 2010.
- Evgeny Kharlamov, Werner Nutt, and Pierre Senellart. Updating probabilistic XML. In *Proc. Updates in XML*, Lausanne, Switzerland, March 2010.
- Serge Abiteboul, Sihem Amer-Yahia, Amélie Marian, Alban Galland, and Pierre Senellart. Birds of a tag flock together. In *Proc. SSM*, New York, USA, February 2010. Workshop without formal proceedings.
- Marc Spaniol, Dimitar Denev, Arturas Mazeika, Pierre Senellart, and Gerhard Weikum. Data quality in Web archiving. In *Proc. WICOW*, pages 19–26, Madrid, Spain, April 2009.
- Pierre Senellart, Avin Mittal, Daniel Muschick, Rémi Gilleron, and Marc Tommasi. Automatic wrapper induction from hidden-Web sources with domain knowledge. In *Proc. WIDM*, pages 9–16, Napa, USA, October 2008.
- Vincent D. Blondel and Pierre Senellart. Automatic extraction of synonyms in a dictionary. In *Proc. TMW*, pages 7–13, Arlington, USA, April 2002. Also registered as Technical Report 89 (2001), Université catholique de Louvain, Louvain-la-neuve.

### Demonstrations (peer-reviewed)

- M. Lamine Ba, Sébastien Montenez, Talel Abdessalem, and Pierre Senellart. Monitoring moving objects using uncertain web data. In *Proc. SIGSPATIAL*, pages 565–568, Dallas, USA, November 2014. Demonstration.
- Muhammad Faheem and Pierre Senellart. Demonstrating intelligent crawling and archiving of web applications. In *Proc. CIKM*, pages 2481–2484, San Francisco, USA, October 2013. Demonstration.
- Yael Amsterdamer, Yael Grossman, Tova Milo, and Pierre Senellart. Crowd miner: Mining association rules from the crowd. In *Proc. VLDB*, pages 1250–1253, Riva del Garda, Italy, August 2013. Demonstration.
- Asma Souhli and Pierre Senellart. Demonstrating ProApprox 2.0: A predictive query engine for probabilistic XML. In *Proc. CIKM*, pages 2734–2736, Maui, Hawaii, USA, October 2012. Demonstration.



Serge Abiteboul, Yael Amsterdamer, Tova Milo, and Pierre Senellart. Auto-completion learning for XML. In *Proc. SIGMOD*, pages 669–672, Scottsdale, USA, May 2012. Demonstration.

Michael Benedikt, Tim Furche, Andreas Savvides, and Pierre Senellart. ProFoUnd: Program-analysis-based form understanding. In *Proc. WWW*, pages 313–316, Lyon, France, April 2012. Demonstration.

Pierre Senellart and Asma Souihli. ProApproX: A lightweight approximation query processor over probabilistic trees. In *Proc. SIGMOD*, pages 1295–1298, Athens, Greece, June 2011. Demonstration.

Talel Abdesslem, M. Lamine Ba, and Pierre Senellart. A probabilistic XML merging tool. In *Proc. EDBT*, pages 538–541, Uppsala, Sweden, March 2011. Demonstration.

#### Posters (peer-reviewed)

Thomas Risse, Stefan Dietze, Wim Peters, Katerina Doka, Yannis Stavarakas, and Pierre Senellart. Exploiting the social and semantic web for guided web archiving. In *Proc. TPD*, pages 426–432, Paphos, Cyprus, September 2012. Poster.

Benjamin Nguyen, Antoine Vion, François-Xavier Dudouet, Ioana Manolescu, Dario Colazzo, and Pierre Senellart. The WebStand project. In *Proc. WebSci*, Athens, Greece, March 2009. Also CoRR abs/1002.0971 (Poster).

Michalis Vazirgiannis, Dimitris Drosos, Pierre Senellart, and Akrivi Vlachou. Web page rank prediction with Markov models. In *Proc. WWW*, pages 1075–1076, Beijing, China, April 2008. Poster.

---

### Student supervision

Official supervision of PhD students

Su Yang (with Serge Abiteboul, since 2016). *Semantic extraction of personal information, integration, and querying.*

Debabrota Basu (with Stéphane Bressan, since 2015). *Reinforcement learning applications to data management problems.*

Mikaël Monet (since 2015). *Efficient querying of large uncertain graphs by exploiting their structure.*

Miyong Han (with Stéphane Bressan and Huayu Wu, since 2015). *Learning approaches to dynamic data management.*

Quentin Lobbé (with Dana Diminescu, since 2015). *Diachronic analysis of diaspora communities through web archives enrichment.*

David Montoya (with Serge Abiteboul, graduated in 2017). *Improving the Usability of Public Transportation.*

Antoine Amarilli (graduated in 2016). *Uncertainty over Intensional Data.*

M. Lamine Ba (with Talel Abdesslem, graduated in 2015). *Probabilistic approaches in integrating social Web data.*

Muhammad Faheem (graduated in 2014). *Intelligent content acquisition in Web archiving.*

Ioana Ileana (with Bogdan Cautis, graduated in 2014). *Query rewriting using views: a theoretical and practical perspective.*

Georges Gouriten (with Bogdan Cautis, 2012–2013). *Knowledge-based content suggestions on the social Web.*

Asma Souihli (graduated in 2012). *Querying probabilistic XML.*

Marilena Oita (graduated in 2012). *Deriving semantic objects from the structured Web.*

Evgeny Kharlamov (with Werner Nutt, graduated in 2010). *A probabilistic approach to XML data management.*

Informal supervision of PhD students

Yael Amsterdamer (with Tova Milo and Serge Abiteboul, 2011). *Learning of probabilistic XML generators.*

Other

Supervision of 2 Master's students and 6 engineering students on long-term projects

---

### Participation in contract research projects

2013–2015

**XData**, *Collaboration grant with industrials on the acquisition of social network data and integration with company data.*

Responsible for Télécom, with funding 175 kEUR.

- 2013 – 2015 **CCIPX**, *STIC-Asia collaborative project between Institut Mines–Télécom, Hanoi University of Science and Technology, and National University of Singapore.*  
Principal co-investigator. Funded 26 kEUR.
- 2012 – 2017 **MoDaS**, *Tova Milo’s ERC advanced grant on Mob Data Sourcing.*  
Collaboration on research papers, exchange of students.
- 2012 – 2013 **Private Recommendation System in Social Networks**, *Campus France PROCORE France–Hong Kong partnership.*  
Principal co-investigator. Funded 8 kEUR.
- 2010 – 2014 **Diadem**, *Georg Gottlob’s ERC advanced grant on Domain-centric Intelligent Automated Data Extraction Methodology.*  
Collaboration on research papers, exchange of students.
- 2011 – 2013 **ARCOMEM**, *European Union FP7 project on the Archiving of the social Web.*  
Leader of the work package on content acquisition ; responsible for Télécom, with funding 600 kEUR.
- 2009 – 2013 **Webdam**, *Serge Abiteboul’s ERC advanced grant on the Foundations of Web Data Management.*  
Leader of the research axis on uncertain data; management board (with S. Abiteboul and V. Vianu).
- 2010 – 2012 **PANIC**, *ANR CONTINT project on the Proactivity of audience and the digitization of cultural industries.*  
Collaborator.
- 2009 – 2011 **DataRing**, *ANR VERSO project on P2P data sharing for online communities.*  
Leader of the work package on data modeling; responsible for Télécom, with funding 120 kEUR.

## Teaching activities

- Created courses** Data Scientist (professional training, Télécom ParisTech), ACM-ICPC training (problem-solving course leading to a programming contest, Télécom ParisTech), Web search (intensive course over one or several weeks, Télécom ParisTech & Renmin University of China), Web technologies (high-level course for business and engineering students, Télécom ParisTech & HEC), L<sup>A</sup>T<sub>E</sub>X (two-day intensive course for graduate students, EDITE de Paris), Web programming (introductory course, Université Paris-Sud)
- Courses with strong involvement** Formal languages (Télécom ParisTech), Web engineering (Télécom ParisTech), Web data management (Université Paris-Sud, Télécom-ParisTech, MPRI, & The University of Hong Kong), Database application design (National University of Singapore), Information theory (National University of Singapore)
- Participation in other courses** C++ programming (École des Ponts ParisTech), Java programming (Télécom ParisTech), Databases (Université Paris-Sud, Télécom ParisTech, & ENS Cachan), Collective intelligence (Télécom ParisTech), Operating systems (Université Paris-Sud)

## Service to the scientific community

- Organization** PC chair of BDA 2017, PC co-chair of WebDB 2017, PC co-chair of SUM 2016, Tutorial co-chair of VLDB 2015, DASFAA 2015, DEXA 2015, ICDE 2011, PC chair of UnCrowd 2014, PC co-chair of DBCrowd 2013, Industrial chair of EDBT 2011, Organizer of the SIGMOD 2010 programming contest, Submission chair of VLDB 2009
- Editorial responsibilities** Information director of the *Journal of the ACM* (2009–2015), Guest co-editor for a special issue of *Distributed and Parallel Databases* (2014)
- Program committees** SIGMOD 2008&2009 (repeatability committee), 2013&2017 (regular PC member), 2015 (group leader), PODS 2012&2016, VLDB 2013, WWW 2009&2010, ICDE 2010, ICDT 2011&2017, CIKM 2010&2014, EDBT 2013 (demonstrations), 2014, ECML/PKDD 2008, BDA 2010&2012&2014, various workshops
- Journal refereeing activities** *Journal of the ACM*, *Journal of Computer Science and Systems*, *VLDB Journal*, *Transactions on Database Systems*, *Information Systems*, *Transactions on Knowledge and Data Engineering*, *Data and Knowledge Engineering*, *Communications of the ACM*, ...

PhD thesis committees Mohammadreza Khelgati (reviewer, Universiteit Twente, 2016), Ahmad Assaf (Télécom ParisTech, 2015), François Rousseau (École polytechnique, 2015), Disheng Qiu (reviewer, Università degli Studi Roma Tre, 2015), Andrés Sanoja (Université Pierre-et-Marie-Curie, 2015), Zeynep Pehlivan (Université Pierre-et-Marie-Curie, 2013), Andrew Sellers (reviewer, The University of Oxford, 2011), Federico Ulliana (reviewer, Université Paris-Sud, 2012), Tom Ameloot (Hasselt University, 2013), reviewer for numerous MSc theses

## Awards

- Distinguished SIGMOD 2017 PC Member
- ACM HyperText 2014 Douglas Engelbart best paper award
- Adviser of Clément Genzmer, winner of the SIGMOD Programming Contest in 2009
- PhD fellowship from École normale supérieure (2005–2008)
- Admitted to the competitive exams of ENS (21st), ENS Cachan (11th), ENS Lyon (2nd), École polytechnique (67th); these competitions draw several thousands of the best science students in France (2000).

## Industrial experience

November 2004 – February 2005 **Software author**, SYSTRAN S.A., La Défense, France.  
SYSTRAN Mobile for Pocket PC.

September 2003 – September 2004 **R&D engineer**, SYSTRAN S.A., La Défense, France.  
SYSTRAN version 5.

September 2002 – October 2002 **Software engineer**, *Le Robert*, Paris, France.  
Analysis of the dictionary graph and automatic extraction of similar words.

## Information technology

Programming languages	C, C++, Java, Perl, Python	Operating systems	Unix/Linux, Windows, Windows CE, Palm OS
XML technologies	DOM, SAX, DTD, XML Schema, XPath, XSLT, XQuery, XSL-FO	Web	(X)HTML, CSS, JavaScript, PHP, JSP, SVG
DBMS	Oracle, MySQL, PostgreSQL	Miscellaneous	L <sup>A</sup> T <sub>E</sub> X, Maple, Unicode, Hadoop

## Languages

French	native language	English	fluent
German	proficient	Classical Arabic	basics
Ancient Greek	high-school level		