

Laszlo SZATHMARY, PhD

Office address:

LATECE Laboratory, Dépt. d'informatique
Université du Québec à Montréal, 201 President Kennedy
Montreal (Québec) H2X 3Y7, Canada

Phone: –

Fax: (+1) (514) 987-8477

E-mail: Szathmary.L@gmail.com

Web: <http://www.loria.fr/~szathmar>

Home address:

XXXXXXXXXXXXXXXXXXXXXXXXXXXX

Montreal (Qc) xxx xxx, Canada

Phone: (+1) (514) xxx-xxxx / xxxxx

Curriculum Vitæ

Laszlo SZATHMARY, PhD

Current Position

since Jan. 2008 *Postdoctoral research fellow* at UQAM (Université du Québec à Montréal) in the LATECE laboratory (Montreal, Canada).

Previous Positions

Sept.–Dec. 2007 *Non-permanent researcher* at the French National Institute for Research in Computer Science and Control (INRIA) in the team Orpailleur (Nancy, France).

School Graduates

2002–2006 *PhD Degree in Computer Science*
University Henri Poincaré, Nancy 1, France
2001–2002 *DEA (5th year degree, equivalent to M.Sc.) in Computer Science*
University Henri Poincaré, Nancy 1, France
1996–2001 *Master Degree in Computer Science*
University of Debrecen, Hungary

Additional Studies

Sept. 2007 Participation at the Reasoning Web 2007 *Summer School* in Dresden, Germany.
July 2003 Participation at the European *Summer School* on Ontological Engineering and the Semantic Web (SSSW-2003) in Cercedilla (near Madrid), Spain.
Sept. 2002 Participation at the European Knowledge Management Forum *Summer School* at Sophia-Antipolis in Nice, France.
1999-2000 A one-year *Erasmus scholarship* in Paderborn, Germany.

Technical Skills

Programming Languages: Java, C/C++, Perl, PHP, Pascal, Prolog, SQL, shell Unix
Operating Systems: Unix/Linux, DOS, Windows
Web Technologies: HTML/Javascript, CSS, XML, RDF(S), SOAP, PHP
Office Applications: Microsoft Office, OpenOffice, L^AT_EX, vi(m), Emacs

Research Interests

- data mining
- formal concept analysis and its applications
- artificial intelligence
- database systems
- knowledge management/representation
- Semantic Web technologies, Web services

Teaching Activities

Legend: Lec. – taught lecture, Tut. – classroom tutorial, Lab. – computer lab work.

2006-2007: teaching assistant and researcher at University Henri Poincaré, Nancy 1, France

- Relational databases with PostgreSQL (Lec. + Tut. + Lab.)
- Programming in Java (Tut. + Lab.)
- Programming environment (Tut. + Lab.)
- Introduction to Unix (Lab.)

2005-2006: teaching assistant and researcher at University Henri Poincaré, Nancy 1, France

- Relational databases with PostgreSQL (Lec. + Tut. + Lab.)
- Databases with Oracle (Lab.)
- Programming in Java (Lab.)
- Programming environment (Lab.)
- Office applications with Microsoft Office (Lab.)

2004-2005: teaching assistant at University Henri Poincaré, Nancy 1, France

- Programming in C (Lab.)
- Office applications with OpenOffice (Lab.)

1998-1999: teaching assistant at University of Debrecen, Hungary

- Programming in C (Lab.)
- Office applications (Lab.)
- Introduction to informatics (Lab.)

Trainee Supervisions

Supervised several students during their training periods, from several educational levels (from 2nd to 4th year of studies).

Competences

Foreign Languages:

English: advanced (English technical translator diploma)
German: advanced (DSH/Oberstufe exam)
French: advanced (DELFI, DELF2, DALF C1 exams)
Hungarian: mother tongue

Miscellaneous:

- Main developer of the CORON data mining platform (<http://coron.loria.fr>).
- Responsible of the team in the technical and network administrators-users interest group of the LORIA laboratory (COMIN) during one year (2005–2006).
- Member of the GALICIA development team (<http://sourceforge.net/projects/galicia>).
- My Erdős number is 4.
- Student volunteer at the ECOOP 2006 conference, Nantes, France.
- Student volunteer at the ECOOP 2001 conference, Budapest, Hungary.

Awards

Best paper award at the conference EGC 2006 in Lille, France. Title of the paper: “Towards Rare Itemset Extraction” (published in French).

Leisure Time Activities

Sports: ice-skating. *Danse:* salsa, rock & roll. *Reading:* science-fiction, adventure.

PhD Thesis Overview

- *Title:* **Symbolic Data Mining Methods with the Coron Platform**
- *Date of thesis defense:* November 24, 2006
- *Location:* University Henri Poincaré, Nancy 1, laboratory of LORIA – INRIA Lorraine, France
- *Note:* the thesis is written in English and it includes an extended abstract in French

Thesis Abstract

The main topic of this thesis is *knowledge discovery in databases* (KDD). More precisely, we have investigated two of the most important tasks of KDD today, namely itemset extraction and association rule generation. Throughout our work we have borne in mind that our goal is to find *interesting* association rules from various points of view: for efficient mining purposes, for minimizing the set of extracted rules and for finding intelligible (and easily interpretable) knowledge units. We have developed and adapted specific algorithms in order to achieve this goal.

The main contributions of this thesis are: **(1)** We have developed and adapted algorithms for finding minimal non-redundant association rules; **(2)** We have defined a new basis for association rules called Closed Rules; **(3)** We have investigated an important but relatively unexplored field of KDD namely the extraction of rare itemsets and rare association rules; **(4)** We have packaged our algorithms and a collection of other algorithms along with other auxiliary operations for KDD into a unified software toolkit called CORON.

Keywords: knowledge discovery in databases (KDD), data mining, itemset extraction, association rule generation, rare item problem.

Software – The CORON Toolkit

I have implemented all the algorithms presented in my thesis in a unified software platform called CORON. CORON is a domain and platform independent, multi-purposed data mining toolkit, which incorporates not only a rich collection of data mining algorithms, but also allows a number of auxiliary operations. To the best of our knowledge, a data mining toolkit designed specifically for itemset extraction and association rule generation like CORON does not exist elsewhere. CORON also provides support for preparing and filtering data, and for interpreting the extracted units of knowledge.

Board of Examiners

- Claude Godart, Professeur, UHP Nancy 1, France (president)
- Bruno Crémilleux, Professeur, Université de Caen, France (rapporteur)
- Sergei O. Kuznetsov, Professor, Higher School of Economics, Moscow, Russia (rapporteur)
- Katalin Bognár, Associate professor, University of Debrecen, Hungary (evaluator)
- Marzena Kryszkiewicz, Associate professor, Warsaw Univ. of Technology, Poland (evaluator)
- Amedeo Napoli, Directeur de recherche CNRS, UHP Nancy 1, France (supervisor)

References

- Dr. Amedeo NAPOLI, (+33) 3.83.59.20.68, napoli@loria.fr
- Dr. Katalin BOGNÁR, (+36) (52) 31.66.66 / 22822, bognar@inf.unideb.hu
- Dr. Sergei O. KUZNETSOV, (+7) (495) 771-3238, skuznetsov@hse.ru
- Dr. Petko VALTCHEV, (+1) (514) 343-7599, valtchev.petko@uqam.ca

Publications

PhD Thesis / Thèse

1. L. Szathmary. *Symbolic Data Mining Methods with the Coron Platform*. PhD Thesis in Computer Science, University Henri Poincaré – Nancy 1, France, Nov 2006

International Journals /

Revue internationale avec comités de sélection et actes

1. L. Szathmary, P. Valtchev, and A. Napoli. Generating Rare Association Rules Using the Minimal Rare Itemsets Family. *International Journal of Software and Informatics*, 4(3):219–238, September 2010

International Conferences /

Conférences internationales avec comités de sélection et actes

1. H. Mili, P. Valtchev, Y. Charif, L. Szathmary, N. Daghrir, M. Béland, A. Boubaker, L. Martin, F. Bédard, S. Caid-Essebsi, and A. Leshob. E-Tourism Portal: A Case Study in Ontology-Driven Development. In *E-Technologies: Transformation in a Connected World*, volume 78 of *LNBI*, pages 76–99. Springer, 2011
2. L. Szathmary, P. Valtchev, and A. Napoli. Finding Minimal Rare Itemsets and Rare Association Rules. In Y. Bi and M.-A. Williams, editors, *Proc. of the 4th Intl. Conf. on Knowledge Science, Engineering and Management (KSEM '10)*, volume 6291 of *LNAI*, pages 16–27, Belfast, Northern Ireland, UK, 2010. Springer, Berlin
3. M. Kaytoue, F. Marcuola, A. Napoli, L. Szathmary, and J. Villerd. The Coron System. In L. Boumedjout, P. Valtchev, L. Kwuida, and B. Sertkaya, editors, *Suppl. Proc. of the 8th Intl. Conf. on Formal Concept Analysis (ICFCA '10)*, pages 55–58, 2010. (demo paper)
4. L. Szathmary, P. Valtchev, A. Napoli, and R. Godin. Efficient Vertical Mining of Frequent Closures and Generators. In *Proc. of the 8th Intl. Symposium on Intelligent Data Analysis (IDA '09)*, volume 5772 of *LNCS*, pages 393–404, Lyon, France, 2009. Springer
5. J. Baixeries, L. Szathmary, P. Valtchev, and R. Godin. Yet a Faster Algorithm for Building the Hasse Diagram of a Galois Lattice. In *Proc. of the 7th Intl. Conf. on Formal Concept Analysis (ICFCA '09)*, volume 5548 of *LNAI*, pages 162–177, Darmstadt, Germany, May 2009. Springer
6. L. Szathmary, P. Valtchev, A. Napoli, and R. Godin. An Efficient Hybrid Algorithm for Mining Frequent Closures and Generators. In *Proc. of the 6th Intl. Conf. on Concept Lattices and Their Applications (CLA '08)*, pages 47–58, Olomouc, Czech Republic, Oct 2008
7. L. Szathmary, P. Valtchev, A. Napoli, and R. Godin. Constructing Iceberg Lattices from Frequent Closures Using Generators. In *Proc. of the 11th Intl. Conf. on Discovery Science (DS '08)*, volume 5255 of *LNAI*, pages 136–147, Budapest, Hungary, Oct 2008. Springer
8. J. Lieber, A. Napoli, L. Szathmary, and Y. Toussaint. First Elements on Knowledge Discovery Guided by Domain Knowledge (KDDK). In S. B. Yahia, E. M. Nguifo, and R. Belohlavek, editors, *Concept Lattices and Their Applications (CLA 06)*, Lecture Notes in Artificial Intelligence 4923, pages 22–41. Springer, Berlin, 2008
9. L. Szathmary, A. Napoli, and P. Valtchev. Towards Rare Itemset Mining. In *Proc. of the 19th IEEE Intl. Conf. on Tools with Artificial Intelligence (ICTAI '07)*, volume 1, pages 305–312, Patras, Greece, Oct 2007
10. L. Szathmary, A. Napoli, and S. O. Kuznetsov. ZART: A Multifunctional Itemset Mining Algorithm. In *Proc. of the 5th Intl. Conf. on Concept Lattices and Their Applications (CLA '07)*, pages 26–37, Montpellier, France, Oct 2007

11. M. d'Aquin, F. Badra, S. Lafrogne, J. Lieber, A. Napoli, and L. Szathmary. Case Base Mining for Adaptation Knowledge Acquisition. In *Proc. of the 20th Intl. Joint Conf. on Artificial Intelligence (IJCAI '07)*, pages 750–755, Hyderabad, India, Jan 2007. Morgan Kaufmann, Inc
12. M. d'Aquin, F. Badra, S. Lafrogne, J. Lieber, A. Napoli, and L. Szathmary. Adaptation Knowledge Discovery from a Case Base. In G. Brewka, editor, *Proc. of the 17th European Conf. on Artificial Intelligence (ECAI '06)*, Trento, Italy, pages 795–796. IOS Press, 2006. (poster paper)
13. L. Szathmary and A. Napoli. CORON: A Framework for Levelwise Itemset Mining Algorithms. In B. Ganter, R. Godin, and E. Mephu Nguifo, editors, *Suppl. Proc. of the Third Intl. Conf. on Formal Concept Analysis (ICFCA '05)*, Lens, France, pages 110–113, Feb 2005. (demo paper)
14. L. Szathmary and A. Napoli. Knowledge organisation and information retrieval based on Galois lattices. In Pham Dinh Tao Le Thi Hoai An, editor, *Modelling, Computation and Optimization in Information Systems and Management Sciences (MCO '04)*, Metz, France, pages 611–618. Hermes Science Publishing, Jul 2004
15. L. Szathmary and A. Napoli. Knowledge organisation and information retrieval with Galois lattices. In E. Motta, N. Shadbolt, A. Stutt, and N. Gibbins, editors, *14th Intl. Conf. on Engineering Knowledge in the Age of the Semantic Web (EKAW '04)*, Whittlebury Hall, UK, volume 3257 of *Lecture Notes in Computer Science*, pages 511–512. Springer, Oct 2004. (poster paper)
16. L. Szathmary and A. Napoli. Towards knowledge management using Galois lattices. In *Proc. of the 6th Intl. Conf. on Applied Informatics (ICAI '04)*, Eger, Hungary, pages 403–417, 2004

National (French) Journal / Revue nationale (francophone)

1. L. Szathmary and A. Napoli. Les treillis de Galois pour l'organisation et la gestion des connaissances. In M. Chavent, M. et Langlais, editor, *11èmes Rencontres de la Société Francophone de Classification (SFC '04)*, Bordeaux, France, volume C of *Revue des Nouvelles Technologies de l'Information (RNTI)*, pages 153–164. CEPAD, Sep 2004. (long version)

National (French) Conferences / Conférences francophones avec comités de sélection et actes

1. B. Ducatel, M. Kaytoue, F. Marcuola, A. Napoli, and L. Szathmary. Coron : Plate-forme d'Extraction de Connaissances dans les Bases de Données. In *17ème conférence en Reconnaissance des Formes et Intelligence Artificielle (RFIA '10)*, pages 883–884, 2010. (demo paper, in French)
2. S. Maumus, A. Napoli, L. Szathmary, and Y. Toussaint. Réflexions sur l'extraction de motifs rares. In M. Nadif and F.-X. Jollois, editors, *Comptes-rendus des 13èmes rencontres de la Société Francophone de Classification (SFC '06)*, Metz, France, pages 157–162. Presses Universitaires de Montréal, 2006
3. L. Szathmary, S. Maumus, P. Petronin, Y. Toussaint, and A. Napoli. Vers l'extraction de motifs rares. In G. Ritschard and C. Djeraba, editors, *Extraction et Gestion des Connaissances (EGC '06)*, Lille, France, pages 499–510. RNTI-E-6, Cépaduès-Éditions Toulouse, 2006. **(best paper award)**
4. S. Maumus, A. Napoli, L. Szathmary, and S. Visvikis-Siest. Fouille de données biomédicales complexes : extraction de règles et de profils génétiques dans le cadre de l'étude du syndrome métabolique. In G. Perrière, A. Guénoche, and C. Gourgeon, editors, *Journées Ouvertes Biologie Informatique Mathématiques (JOBIM '05)*, Lyon, France, pages 169–173, Jul 2005
5. L. Szathmary and A. Napoli. Les treillis de Galois pour l'organisation et la gestion des connaissances. In M. Chavent, O. Dordan, C. Lacomblez, M. Langlais, and B. Patouille, editors, *11èmes Rencontres de la Société Francophone de Classification (SFC '04)*, Bordeaux, France, pages 298–301, Sep 2004. (short version)

International Workshop / Atelier international avec comités de sélection et actes

1. L. Szathmary and A. Napoli. Knowledge organisation and information retrieval using Galois lattices. In R. Dieng-Kuntz and N. Matta, editors, *Workshop on Knowledge Management and Organizational Memories – ECAI '04 (16th European Conf. on Artificial Intelligence)*, Valencia, Spain, pages 73–78, Aug 2004

National (French) Workshop / Atelier national avec comités de sélection et actes

1. S. Maumus, A. Napoli, L. Szathmary, and S. Visvikis-Siest. Exploitation des données de la cohorte STANISLAS par des techniques de fouille de données numériques et symboliques utilisées seules ou en combinaison. In *Workshop on Fouille de Données Complexes dans un Processus d'Extraction des Connaissances – EGC '05 (5èmes Journées d'Extraction et Gestion des Connaissances)*, Paris, France, pages 73–76, 2005

Technical Reports / Rapports de recherche

1. L. Szathmary, P. Valtchev, and A. Napoli. Efficient Mining of Frequent Closures with Precedence Links and Associated Generators. Research Report RR-6657, INRIA, 2008
2. L. Szathmary, S. Maumus, and A. Napoli. Mining Rare Association Rules. LORIA Research Report 00102909 (<http://hal.inria.fr/inria-00102909>), Oct 2006
3. S. Maumus, L. Szathmary, A. Napoli, E. Albuissou, and S. Visvikis-Siest. Towards a global methodology for mining cohorts with biological and genetic data. Rapport de Contrat, Nov 2005. (Contrat de Plan Etat-Région, PRST Intelligence Logicielle)

Miscellaneous / Autres

1. L. Szathmary and A. Napoli. CORON: A Platform for Itemset Mining Algorithms. In *Fourth Intl. Conf. on Formal Concept Analysis (ICFCA '06)*, Dresden, Germany, Feb 2006. (oral communication and demo, without paper)

Software / Logiciel

1. L. Szathmary and A. Napoli. The CORON Data Mining Platform. (<http://coron.loria.fr>)