

Benjamin PIWOWARSKI

Yahoo! Research Latin America

*Oficina 402
Blanco Encalada 2120
Santiago • Chile*
☎ +52 (2) 678 49 81
✉ bpiwowar@yahoo-inc.com
<http://www.bpiwowar.net/>



General information

My research interests are Structured Information Retrieval (SIR), XML Databases, Web Log mining, and Statistical Machine Learning.

Research keywords: (Structured) Information Retrieval, Evaluation, Web Log Mining, XML Databases, Machine Learning.

Teaching keywords: Database, XML Databases, Structured Information Retrieval, Networks, Internet, and Operating Systems.

EDUCATION AND PROFESSIONAL EXPERIENCE

Education

- 10/1999 - 7/2003 **Ph.D. Thesis in Computer Science**, *University Paris 6*, Paris, France.
Machine Learning and Structured data: applications to Information Retrieval, Director: Prof. Patrick Gallinari, Highest Ph.D. grade (Mention Très Honorable)
- 01/10/1998-30/09/1999 **D.E.A. (Master Thesis)**, *University Paris 6*, Paris, France.
Artificial Intelligence and Pattern Recognition. *Main courses*: neural networks, uncertainty reasoning, statistical pattern recognition, distributed artificial intelligence, data mining. University Paris 6, Paris, France Highest grade (Mention Très Bien)
- 4/2000 - 9/2000 **DEA Internship**, *LexiQuest SARL*, Montreuil, France.
Machine Learning and Information Retrieval
- 9/1997 - 6/1998 **3rd and 4th year courses in mathematics**.
University Paris 5, Paris, France Main courses: Applied analysis; Second order processes and signal theory; Advanced Scientific Computation (Numerical solution of partial differential equations); Integration and probabilities; Differential Calculus; Complex analysis.
- 9/1994 - 6/1997 **1st and 2nd year courses in mathematics and psychology**.
University Paris 5, Paris, France

Professional Experience

- 3/2006 - current **Researcher position**, *Yahoo! Research Latin America*.
Main research area: Web Log Mining
- 9/2004 - 2/2006 **Post-doctorate position**, *Departemento de Ciencias de la Computacion (DCC)*, University of Chile, Santiago.
Main research area: XML Databases
- 9/2002 - 8/2004 **Temporary lecturer (ATER, French 1-year full-time position)**, *University Paris 6*, Paris, France.
Main research area: Structured Information Retrieval and Machine Learning

TEACHING AND SUPERVISION

Thesis and Internship Supervision

- 1/2007 - current **Master internship**, *University of Chile*, Santiago, Chile.
XML Database with an XQuery engine based on Proximal Nodes.
- 4/2003 - 9/2003 **D.E.A. (Master) internship**, *University Paris 6*, Paris, France.
XML databases for Structured Information Retrieval. In collaboration with P. Gallinari and A. Doucet
- 10/2002 - 2/2003 **Engineer internship**.
XML editor for letters (Java)
- 4/2001 - 9/2001 **D.E.A. (Master) internship**.
Structure, Machine Learning and Categorisation

Teaching Experience

- 8/2005 - current **XML Databases**, *University of Chile*, Santiago, Chile.
Lecturer, 36 hours per year
- 8/2006 - current **Information Retrieval**, *University of Chile*, Santiago, Chile.
Lecturer (part of the course): 14h hours focusing on XML Information Retrieval
- 2002-04 **Databases**, *University Paris 6*, Paris, France.
Teaching assistant, 44 hours
- 2003-04 **Network**, *University Paris 6*, Paris, France.
Teaching assistant, 44 hours
- 2000-04 **Operating Systems and Internet**, *University Paris 6*, Paris, France.
Teaching assistant, 1st year course, 7×30 hours
- 1999-00, 2003-2004 **Excel and Pascal**, *University Paris 6*, Paris, France.
Teaching assistant, 3×32 hours
- 2001-02, 2003-04 **Computer Organization and Assembler Programming**, *University Paris 6*, Paris, France.
Teaching assistant, 34 hours

PROJECTS AND OTHER CONTRIBUTIONS

Projects

- 2003-2004 **ACIMDD**, *University Paris 6*, Paris, France.
University Paris 6 - INRIA (The French National Institute for Research in Computer Science and Control). XML Information Access.
- 1999-02 **Victor Hugo Letters**, *University Paris 6*, Paris, France.
University Paris 6 - Flohic éditions (French publisher) - Thalès Structured Information
Retrieval in Victor Hugo letters

Software

- XPN XQuery with Proximal Nodes
- EvalJ Evaluation software for Structured Information Retrieval
- EPSIR Experimental Platform for Structured Information Retrieval
- C-shellexec a small cross-platform web server

Internet and networks

- Internet <http://www-connex.lip6.fr>, <http://www.oedipe.org>
- Intranet <http://www-connex.lip6.fr/ressources/>
- Network and System Administration Team network in Paris 6 (Linux)

SKILLS

Computer skills

Programmation	C/C++, Java, Python, Shell, Hadoop	Web	HTML, Javascript, PHP
Databases	SQL, MySQL, Postgresql	XML	XML, XSLT, XQuery

Languages

French	Native
English	Very good
Spanish	Very good
Polish	Notions

PUBLICATIONS AND RELATED ACTIVITIES

Conference related activities

PC International conferences	WWW 2008, SIGIR 2007, CIKM 2007, CIKM 2005
PC Workshops	SIGIR 2006 Workshop on XML Element Retrieval Methodology, INEX 2005 Workshop on Element Retrieval Methodology
Organizing Committee	SIGIR 2007 IRGM Workshop (Information retrieval and applications of graphical models), INEX 2007 (assessment interface and metrics), INEX 2006 (assessment interface), INEX 2005 (assessment interface and metrics)
Workshop Chair	Chairman of the "Assessment Interface" session INEX (Initiative for the Evaluation of XML Retrieval), 12/2003, 12/2005)

Publications

Journals

Benjamin Piwowarski, Andrew Trotman, and Mounia Lalmas, *Sound and complete relevance assessments for xml retrieval*, *Accepted with minor revisions in ACM TOIS*.

Benjamin Piwowarski, Patrick Gallinari, and Georges Dupret, *An extension of precision-recall with user modelling (PRUM): Application to XML retrieval*, *ACM Transactions on Information Systems* **25** (2007), no. 1, 1.

Benjamin Piwowarski and Patrick Gallinari, *A bayesian network for XML information retrieval: Searching and learning with the INEX collection*, *Information Retrieval* **8** (2005), no. 4, 655–681.

International conferences

Gabriella Kazai, Benjamin Piwowarski, and Stephen Robertson, *Effort precision and gain-recall based on a probabilistic navigation model*, 1st International Conference on the Theory of Information Retrieval, October 2007.

Olivier Motelet, Nelson Baloian, Benjamin Piwowarski, and Jose A. Pino, *Taking advantage of the semantics of a lesson graph based on learning objects*, The 13th International Conference on Artificial Intelligence in Education (AIED 2007), IOS Press, July 2007.

Olivier Motelet, Benjamin Piwowarski, Georges Dupret, Jose A. Pino, and Nelson Baloian, *Enhancing educational-material retrieval using authored-lesson metadata*, Fourteenth String Processing and Information Retrieval Symposium (SPIRE 2007) (Santiago, Chile), October 2007.

Benjamin Piwowarski and Hugo Zaragoza, *Predictive user click models based on click-through history*, Proceedings of the Sixteenth Conference on Information and Knowledge Management (CIKM 2007), November 2007.

G. Dupret and Piwowarski, *Principal components for automatic term hierarchy building*, Proceedings of the 13th International Symposium on String Processing and Information Retrieval (SPIRE 2006), LNCS 4209, Springer, 2006, pp. 37–48.

G. Dupret, B. Piwowarski, C. Hurtado, and M. Mendoza, *A statistical model of query log generation*, Proceedings of the 13th International Symposium on String Processing and Information Retrieval (SPIRE 2006), LNCS 4209, Springer, 2006, pp. 217–228.

B. Piwowarski and G. Dupret, *Evaluation in (XML) information retrieval: Expected precision-recall with user modelling (EPRUM)*, SIGIR 2006: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2006, pp. 260–267.

Benjamin Piwowarski and Mounia Lalmas, *Providing consistent and exhaustive relevance assessments for XML retrieval evaluation*, Proceedings of the Thirteenth Conference on Information and Knowledge Management (CIKM 2004) (Washington D.C., U.S.A.), November 2004.

Benjamin Piwowarski and Patrick Gallinari, *A machine learning model for information retrieval with structured documents*, Machine Learning and Data Mining in Pattern Recognition (Leipzig, Germany) (Petra Perner, ed.), Springer Verlag, July 2003, pp. 425–438.

Benjamin Piwowarski, *Learning in information retrieval: a probabilistic differential approach*, Proceedings of the BCS-IRSG, 22nd Annual Colloquium on Information Retrieval Research (Sidney Sussex College, Cambridge, England), April 2000.

National conferences

Benjamin Piwowarski and Mounia Lalmas, *Interface pour l'évaluation de systèmes de recherche sur des documents XML*, Première Conférence en Recherche d'Information et Applications (CORIA'04) (Toulouse, France), Hermès, March 2004.

Benjamin Piwowarski, Ludovic Denoyer, and Patrick Gallinari, *Un modèle pour la recherche d'information sur des documents structurés*, 6èmes Journées internationales d'Analyse statistique des Données Textuelles (JADT 2002) (Saint-Malo, France), March 2002.

Benjamin Piwowarski, *Apprentissage et recherche documentaire : une approche probabiliste différentielle*, Colloque Francophone sur l'Apprentissage Automatique (CAP'2000) (Saint-Etienne, France), Hermès, June 2000.

Other

Jovan Pehcevski and Benjamin Piwowarski, *Evaluation metrics for semi-structured text retrieval*, Entry for the *Encyclopedia of Database Systems* (accepted for publication).

Benjamin Piwowarski, *Techniques d'apprentissage pour le traitement d'informations structurées : application à la recherche d'information*, Ph.D. thesis, Université Pierre et Marie Curie, Paris, France, July 2003.