

Benoît GROZ
Université Paris-Saclay
Bâtiment 650 (PCRI/S)
1, rue Raimond Castaing
91190 GIF-SUR-YVETTE
France

Curriculum Vitae



Tél. : (0)1 69 15 64 02
E-mail : benoit.groz@lri.fr

Né le 18/02/1986
Nationalité Française

Situation actuelle

Sep. 2014– **Maître de conférences** à l'Université Paris-Saclay.

Thèse et post-doctorat

Nov. 2012–2014 **Post-doctorat** sur le Crowdsourcing avec Tova Milo à Tel-Aviv.

Sept 2008–2012 **Thèse** sous la direction de Sophie Tison et Sławek Staworko à l'Université de Lille, au sein de l'équipe MOSTRARE : XML SECURITY VIEWS : QUERIES, UPDATES, AND SCHEMAS
Sujets étudiés : XPath, réécriture de requêtes mise à jour à travers des vues, approximation et validation de schémas, automates visibly pushdown...

Jury : Michael Rusinowitch, Thomas Schwentick, Hélène Touzet et Jean-François Raskin.
Thèse financée par une Allocation Moniteur Normalien, et soutenue le 5/10/ 2012.

Activités connexes :

- Collaboration avec Sebastian Maneth sur les expressions régulières déterministes, dans le cadre d'une visite de 15 jours au NICTA à Sydney.
- Encadrement d'un stage étudiant de licence.

Études

2006–2008 **Master au MPRI** *Master Parisien de Recherche en Informatique co-habilité entre Paris 7, les ENS et l'École Polytechnique* mention Très Bien. Rang : 3^e/58
2005–2006 **Licence d'Informatique** à l'ENS de Cachan, mention Bien
2005–2008 **Normalien à l'École Normale Supérieure de Cachan**
2003-2005 **Classes Préparatoires aux Grandes Écoles : MPSI/MP***, Grenoble (Champollion)
2003 **Baccalauréat Scientifique option Mathématiques**, mention Très Bien

Stages de Recherche

Mars–Juillet 2008 **Stage de Master 2** au LSV (ENS Cachan) encadré par Luc SEGOUFIN
Mars–Juillet 2007 **Stage de Master 1** à l'ULB (Bruxelles) encadré par Jean-François RASKIN
Juin–Aout 2006 **Stage de Licence** au RWTH (Aachen), encadré par Wolfgang THOMAS

Activités d'animation de la recherche

Relectures d'articles pour : ASIACCS'11, LID'11, ICDT'12, CIAA'13, ICDE'13, STACS'13, JCSS'14, CIKM'15, JACM'15, VLDB'15, TKDE'15, PODS'16, TKDE'16, KAIS'17, SIGMOD'17, ICCS'18, FSTTCS'18, JCOMP'18, RANDOM'19, VLDB'20, SAC'21, STACS'21, ICALP'22, BDA'22 (membre du PC BDA'22).

Membre d'un jury de recrutement MCF en 2023.

Prise en charge de l'organisation du séminaire interne de l'équipe MOSTRARE de 2010 à 2012.

(co)Encadrement doctoral

2020– **Doctorat de Taha Halal** (encadrant principal : Bogdan Cautis) *en cours de rédaction*
Apprentissage et optimisation sur les graphes
2023– **Doctorat de Lilia Izri** (encadrant principal : Silviu Maniu) *en cours*
Optimisation en présence de données incertaines et restreintes

Enseignement

2023-	Social graphs	Paris Saclay	M2
2020-	Systèmes pour le big data	Paris Saclay	M1, M2
2019-	Introduction aux Bases de Données	Univ. Paris-Saclay	2e année prépa
2018-2019	Bases de Données	École Polytechnique	M1
2014-2015	Bases de Données	Paris Sud	L3,M1
2014-2015	Bases de Données	UFE(Egypte)	L3
2014-	Entrepôts de Données	Paris Saclay	M1, M2
2011-2012	Algorithmes de graphes et Programmation Linéaire	Lille 1	L3
2009-2012	Algorithmique Avancée et Complexité	Lille 1	M1
2009-2010	Codage	Lille 1	L2
2009-2010	Technologies du web	Lille 1	L2
2009-2010	Assembleur	Lille 1	L2
2008-2009	Bases de Données	Lille 1	M1

Responsabilité et activités d'enseignement

Responsable d'année du M2 Miage Informatique Décisionnelle en apprentissage depuis 2021.

Centres d'intérêts scientifiques

Logique et Automates, Fouille de données, Théorie des Graphes, Requêtes dans les graphes, Schémas dans les graphes, Données incertaines (et plus généralement : Algorithmes, Technologies du Web et Vérification)...

Langues

Anglais (TOEIC : 970/990), Allemand lu, écrit

Divers

Musique : piano et orgue à tuyaux, histoire, lecture...

Accueil et visite de monuments.

(par le passé) Activités d'enseignement auprès d'enfants hospitalisés et pour la réintégration de détenus dans la société,

Participation à des projets de recherches

PEPS JCJC INS2I Erable (2016-2017) avec Silviu Maniu :

Evaluation de Requêtes d'Agrégation sur Bases de données probabiListEs.

Internal Project WebStore (2016-2018) avec F.Bugiotti, N.Bennacer, S.Maniu :

Multi-target news crawling through API

Projet AIDA (2020-2024) financé pour un tiers par Bpifrance dans le cadre du Programme d'investissements d'avenir.

Publications

Les rangs indiqués entre parenthèses après les noms de journaux ou conférences sont ceux attribués par CORE.

Articles journal :

Guillaume Bagan, Angela Bonifati, Benoît Groz :

A trichotomy for regular simple path queries on graphs.

Journal of Computer and System Sciences, 2020. (A*)

<https://inria.hal.science/hal-02435355/document>

| Cet article complète la publication de la conférence PODS 2013, essentiellement en précisant les preuves.

Benoît Groz, S.Maniu :

Hypervolume Subset Selection with Small Subsets.

Evolutionary Computation, 2019. (A)

https://www.mitpressjournals.org/doi/abs/10.1162/evco_a_00235

Le problème Hypervolume Subset Selection cherche à approximer un ensemble de n solutions à un problème d'optimisation multicritère. Il s'agit de sélectionner k points parmi les n points de $\mathbb{R}^d$, de sorte que les points sélectionnés maximisent l'hypervolume dominé (et donc couvrent au mieux l'espace des solutions). Cet article analyse et compare les algorithmes existants en dimension 2 et introduit le premier algorithme asymptotiquement optimal pour une heuristique gloutonne de la littérature. En dimension 3, l'article propose des algorithmes lorsque k ou $n - k$ sont petits, et présente aussi une structure de donnée permettant de représenter efficacement les contributions conjointes de sous-ensembles de points. Enfin, nous étendons nos techniques à des dimensions arbitraires pour le cas où $k \leq 3$.

Benoît Groz, Sebastian Maneth :

Efficient testing and matching of deterministic regular expressions.

Journal of Computer and System Sciences, 2017. (A*)

<https://www.sciencedirect.com/science/article/pii/S0022000017300843>

| Cet article complète la publication de la conférence PODS'12 avec notamment des résultats sur les expressions régulières mettant en jeu des indicateurs numériques permettant de répéter certaines sous-expressions.

Benoît Groz, Slawomir Staworko, Anne-Cécile Caron, Y. Roos, and S. Tison :

Static analysis of XML security views and query rewriting

<https://hal.archives-ouvertes.fr/hal-00796297/document>

numéro spécial "d'Information and Computation" sur la sécurité et les techniques de réécriture, 2014. (B)

Cet article s'intéresse à la réécriture de requêtes sur les vues lorsque les vues sont définies par Regular XPath ou MSO sélectionnant les parties d'un documents XML accessible à l'utilisateur. L'article propose aussi différentes manières pour comparer des vues, et analyse la complexité de ces comparaisons selon les restrictions que l'on impose sur les vues et documents.

Workshops avec comité de sélection :

Benoît Groz, Frederik Mallmann-Trenn, Claire Mathieu, Victor Verdugo :

Skyline Computation with Noisy Comparisons

International Workshop (IWOCA), Bordeaux, June 8-10, 2020 (B)

<https://arxiv.org/abs/1710.02058>

| Ce travail améliore en général les bornes du papier PODS'2015 ci-dessous.

Slawomir Staworko, Iovka Boneva, and Benoît Groz :

The View Update Problem for XML,

Workshop on Updates in XML (WUXML), 2010 (EDBT/ICDT Workshops Proceedings)

Benoît Groz, Slawomir Staworko, Anne-Cécile Caron, Y. Roos, and S. Tison :

Query Rewriting for XML Access Control,

Workshop on Security and Rewriting Techniques (Secret), 2009

Conférences :

Iovka Boneva, Benoît Groz, Jan Hidders, Filip Murlak, Slawek Staworko :

Static Analysis of Graph Database Transformations

Symposium on Principles of Database Systems (PODS), Seattle, June 18-23, 2023 (A*)

<https://hal.science/hal-03937274v2/document>

Ce travail étudie des problèmes d'analyse statique (équivalence de transformations, type-checking, construction du schéma cible) pour les transformations de graphes en présence de schémas. Les transformations sont définies par des règles de type datalog dont le corps est une requête de chemin régulière (C2RPQ acyclique). Les schémas sont définis par des contraintes de participation similaires à celles définies par le modèle Entité Relation : des contraintes cardinalité régissant le nombre d'arêtes entre deux nœuds d'un type donné (chaque nœud a une unique étiquette qui détermine son type). L'article réduit les problèmes étudiés à la satisfaisabilité (sans restriction à un domaine fini) de C2RPQ p en présence d'un schéma défini par une TBox $\mathcal{T}$ appartenant à la classe Horn-ALC_{IF}. Et il montre que ce problème de satisfaisabilité peut se résoudre en temps $O(\text{poly}(|\mathcal{T}|) \cdot 2^{\text{poly}(|p|, k, \ell)})$ où k est le nombre de concepts de la TBox et ℓ le nombre de contraintes de cardinalité de type "at most" dans la TBox.

Benoît Groz, Aurélien Lemay, Slawek Staworko, Piotr Wiecezorek :

Inference of Shape Graphs for Graph Databases.

International Conference on Database Theory (ICDT), Edinburgh, UK (virtual), March 29-April 1, 2022 (A)

<https://hal.science/hal-03937274v2/document>

Ce travail s'attaque au problème d'inférer un schéma de type ShEx pour les graphes RDF. Le problème est défini et résolu par une approche d'inférence grammaticale.

Benoît Groz, Ezra Levin, Isaac Meilijson, and Tova Milo :

Filtering With the Crowd : CrowdScreen Revisited

International Conference on Database Theory (ICDT), Bordeaux, March 15-18, 2016 (A)

<https://hal.archives-ouvertes.fr/hal-01239458/document>

Ce travail présente des algorithmes pour filtrer les objets satisfaisant une requête booléenne dans le cadre du CrowdSourcing. On se place dans le cadre du modèle CrowdScreen proposé par Parameswaran et al. (Sigmod'12). Ce modèle suppose que chaque travailleur se comporte comme un oracle booléen probabiliste renvoyant la réponse correcte avec une probabilité constante. L'objectif est de définir une stratégie qui minimise le nombre moyen d'appels à l'oracle pour chaque objet, tout en maintenant la probabilité de mal classifier un objet sous un certain seuil. Nous analysons les algorithmes proposés par CrowdScreen pour expliquer leur propriétés et proposons de nouvelles heuristiques passant mieux à l'échelle. Enfin, nous évaluons les différentes heuristiques sur des données synthétiques ainsi que dans le cadre d'une réelle expérience de CrowdSourcing.

Benoît Groz, Tova Milo :

Skyline Queries with Noisy Comparisons,

Symposium on Principles of Database Systems (PODS), Melbourne, May 31-June 4, 2015 (A*)

<https://hal.inria.fr/hal-01146568/document>

Ce travail présente des algorithmes pour évaluer les requêtes de skyline dans le cadre du CrowdSourcing. Plus précisément, le problème suppose donné un oracle pouvant comparer les objets passés en entrée avec une probabilité d'erreur fixe connue. Pour augmenter la fiabilité d'une comparaison, on peut répéter une comparaison, les réponses de l'oracle étant indépendantes, car chaque appel à l'oracle modélise l'utilisation d'un travailleur différent dans le cadre du CrowdSourcing. On suppose que les points peuvent être comparés selon d critères et l'objectif est de calculer exactement l'ensemble des k points formant la skyline (les optima de Pareto), avec une probabilité d'erreur au plus δ , en minimisant les appels à l'oracle. Nous analysons et renforçons différents algorithmes pour trier ou évaluer des formules booléennes à partir de données bruitées, et montrons entre autres que la skyline peut être calculée ces algorithmes peuvent s'en assemblant et renforçant différents algorithmes de tris, recherches binaire et entre autres que l'on peut résoudre le problème avec $O(dnk \log(dn/\delta))$ comparaisons.

Guillaume Bagan, Angela Bonifati, Benoît Groz :

A Trichotomy for Regular Simple Path Queries on Graphs,

Symposium on Principles of Database Systems (PODS), New York, USA, June 22-27, 2013 (A*)

<https://hal.archives-ouvertes.fr/hal-00806448>

Ce travail présente une caractérisation des langages réguliers L pour lesquels l'évaluation des requêtes de chemins simples régulières ($RSPQ(L)$) est décidable en temps polynomial. Plus précisément nous établissons une trichotomie : en fonction du langage L le problème est NP-complet, NL-complet ou dans AC_0 . Nous proposons un algorithme permettant d'évaluer efficacement de telles requêtes dans le cas abordable, et proposons une caractérisation décidable en fonction de propriétés de pompage sur le langage L , se traduisant en propriétés sur l'automate minimal de L . Une caractérisation équivalente du fragment "abordable" est aussi exposée en terme d'expression régulières.

Benoît Groz, Sebastian Maneth and Slawomir Staworko :

Deterministic Regular Expressions in Linear time,

Symposium on Principles of Database Systems (PODS), Scottsdale, AZ, USA, May 20-24, 2012 (A*)

<http://hal.inria.fr/docs/00/61/84/51/PDF/1.pdf>

Ce travail présente le premier algorithme en temps linéaire pour tester si une expression régulière est déterministe, ainsi que de nouveaux algorithmes pour évaluer les expressions régulières déterministes. Ces algorithmes construisent des pointeurs sur l'arbre d'analyse de l'expression en combinaison avec des structures de données pour accélérer des requêtes retournant des ancêtres de noeuds.

Iovka Boneva, Benoît Groz, S.Tison, Anne-Cécile Caron, Y. Roos, and Slawomir Staworko :

View update translation for XML,

International Conference on Database Theory (ICDT), Uppsala, Sweden, March 21-24, 2011 (A)

<http://hal.inria.fr/docs/00/53/48/57/PDF/paper.pdf>

Ce travail étudie le problème de mise à jour à travers une vues, pour des fonctions de mise à jour qui transforment n'importe quel document d'entrée en un nouveau document. Le problème considéré est de calculer les modifications qui doivent être apportées au document réel pour obtenir l'effet désiré sur la vue. Dans ce travail, nous représentons vues et fonctions de mise à jour par des ensembles réguliers d'arbres d'alignement.

Benoît Groz, Slawomir Staworko, Anne-Cécile Caron, Y. Roos, and S. Tison :

XML Security views revisited,

Database Programming Languages International Symposium (DBPL), Lyon, France, August 24, 2009. (B)

<https://hal.archives-ouvertes.fr/hal-00396796/document>

Dans ces travaux nous définissons un modèle général pour les vues de sécurité XML, modèle dans lequel les vues sont définies par des requêtes RegularXpath. Pour ce modèle nous étudions plusieurs approches permettant de comparer des politiques de sécurité. Ces comparaisons se basent essentiellement sur l'inclusion de requêtes ou la réécriture de requêtes. Nous proposons aussi différentes techniques pour obtenir une représentation approchée du schéma de vues lorsque le schéma de vue dérivé du schéma des documents n'est pas un langage régulier d'arbres.

Ouvrages :

Grégory Quenet (coord.), Boris Beade, Michel Beaudouin-Lafon, Youness Bousraf, Benoît Groz, Jean-Noël Lafargue, Bertrand Maigné, Olivier Marcé, Emmanuel Waller

Sciences Numériques et Technologie (SNT), 2nde,

Editions Didier, 192 pages, 2019.

ISBN 978-2-278-09446-2

Michel Beaudouin-Lafon (ed.), Benoît Groz, Mathieu Nancel, Olivier Marcé, Emmanuel Waller.

Numérique et Sciences Informatiques (NSI), 1ère spécialité,

Hachette Education, Paris, France. 288 pages, 2021.

ISBN 978-2-01-786630-5.

Michel Beaudouin-Lafon (ed.), Céline Chevalier, Gilles Grimaud, Benoît Groz, Philippe Marquet, Mathieu Nancel, Cristel Pelsser, Xavier Redon, Thomas Vantroys, Emmanuel Waller.

Numérique et Sciences Informatiques (NSI), Terminale spécialité

Hachette Education, Paris, France. 352 pages, 2022

ISBN 978-2-01-786634-3