

PARTIE RÉDACTIONNELLE DU RAPPORT D'ACTIVITÉ

NOM : BENZAKEN

PRÉNOM : VERONIQUE

NUMEN: 01S9346459ORA

Note à l'attention des rapporteurs.

J'ai, en 2006, été brusquement frappée d'une tumeur à l'œil. J'ai été traitée par proton-thérapie et suis toujours suivie par l'Institut Curie. Ceci a entraîné une sévère baisse de mon acuité visuelle, accompagnée de pertes d'équilibre et de divers autres troubles. C'est pourquoi, depuis 2006, je ne peux plus m'investir autant dans les charges collectives. Etant désormais considérée en «vision basse», je bénéficie, depuis la rentrée 2009, suite aux recommandations de la médecine du travail, d'un aménagement de service à hauteur de 96h eq/TD annuelles.

Situation Professionnelle.

1998 Depuis septembre 1998, **Professeur** à l'Université de Paris Sud 11. **Promue à la première classe** depuis 2004, en CRCT au titre local pour 2003-2004. **Titulaire de la prime d'encadrement doctoral** d'octobre 1993 à novembre 2001 puis depuis octobre 2006. Depuis octobre 2010 titulaire de la **prime d'excellence scientifique**. Membre de l'équipe Bases de Données du LRI (UMR 8623 du CNRS) de 1998 à 2010 puis, depuis septembre 2010, de l'équipe Proval- Preuve de programmes, équipe commune du LRI et de l'INRIA - Saclay.

1990-1998 D'octobre 1990 à août 1998, **Maître de Conférences** à l'Université de Paris 1 (Panthéon-Sorbonne). Membre du CRI (Centre de Recherche en Informatique) de Paris 1.

1990-1991 De 1990 à 1991, **Conseiller extérieur** à l'INRIA.

Sociétés savantes Membre de l'ACM ; Membre de SPÉCIF

Études et diplômes.

1996 Habilitation à diriger des recherches - Université Paris 11.

JURY : C. Delobel Président, M. Adiba Rapporteur, A. Albano Rapporteur, M.P. Atkinson Rapporteur, M-C. Gaudel, F. Gire, C. Rolland, N. Spyrtos.

TITRE : "Langages de programmation pour bases de données et contraintes d'intégrité".

1990 Thèse de Doctorat, effectuée sous la direction de Mr le Professeur C. Delobel.

JURY : C. Delobel Président, D. J. DeWitt Rapporteur, P. Valduriez Rapporteur, F. Bancilhon, J. Beauquier.

TITRE : "Regroupement d'objets sur disque dans un système de bases de données orienté-objet". Spécialité Informatique, Université de Paris 11.

Synthèse

Professeur (1998)
Paris 11, Maître de conférences (1990) Paris 1.

HDR (12/12/96) et Doctorat (16/1/90) Paris 11,
DEA (15/9/86) Paris 7 ;
Premier Prix de chant et d'art-lyrique, (15/6/83) CNR Grenoble.

1986 D.E.A. d'Informatique Fondamentale - Université Paris 7. Mention : Très bien.

1983 Premier prix de Chant et Premier prix d'Art-Lyrique du Conservatoire National de Région (CNR) de Grenoble.

Expérience industrielle.

1984 Ingénieur d'application, Société ORDINA.

1986-1989 Dans le cadre de ma thèse sous convention CIFRE, Ingénieur, Société IN2.

I. Activités scientifiques.

Mes activités de recherche sont à la confluence de deux domaines : **gestion de données** et **langages de programmation**. Mon objectif est l'obtention de langages et/ou de systèmes adaptés au traitement de grandes masses de données. La méthodologie qui me guide dans mon travail consiste à développer de solides bases théoriques pour bâtir sur de telles fondations des outils robustes et efficaces.

Récemment, je me suis principalement intéressée au **format XML** ce qui m'a conduit à définir, en collaboration avec G. Castagna et A. Frisch, un **langage de programmation**, CDuce, adapté au traitement de telles données ainsi qu'à l'étude de leur stockage natif et de l'optimisation des langages de requêtes permettant de les manipuler. CDuce fait désormais partie des **principales distributions de Linux** (Debian, Ubuntu, Mandriva et Fedora), une version est opérationnelle sur les plateformes Windows XP et Mac OS X.

Depuis septembre 2008, je travaille, en collaboration avec J-D Fekete et I. Manolescu (INRIA), à la définition de **systèmes de workflow interactifs** et d'une plateforme pour le traitement et la visualisation de données scientifiques massives et distribuées.

Très récemment, j'ai entamé un travail de recherche sur des aspects liés à la **parallélisation sur GPU de langages fonctionnels** en collaboration avec G. Castagna (PPS) et J. Falcou (LRI). Le détail de mes thématiques est présenté en Annexe.

Cinq contributions majeures

Articles dans des revues internationales avec comité de lecture

- [1] Alain Frisch, Giuseppe Castagna and Véronique Benzaken. Semantic subtyping: Dealing set-theoretically with function, union, intersection, and negation types. *Journal of the ACM*, 55(4):1–64, 2008.
- [2] V. Benzaken and A. Doucet. Thémis: A Database Programming Language Handling Integrity Constraints. *Very Large Databases (VLDB) Journal*, 4(3):493–518, 1995.

Communications dans des conférences internationales avec comité de lecture

- [3] V. Benzaken, G. Castagna, and A. Frisch. CDuce: an XML-centric general-purpose language. In *ACM International Conference on Functional Programming (ICFP'03)*, 2003.

Ingénieur d'application
SSII (1984) ; Convention
CIFRE (86-90)

Thématiques :
Aspects système des
bases de données
(1986-1990, doctorat),
Contraintes d'intégrité
dans les bases de don-
nées (1991-1999, maître
de conférences), typage
et sécurité pour données
semi-structurées,
langages centrés
données (depuis 2001),
workflows pour données
scientifiques (depuis
2008), programmation
fonctionnelle parallèle et
GPUs.

Encadrement :
Direction de 10 thèses,
de 20 stages de
DEA, d'un stage pour
normaliens.

- [4] V. Benzaken, G.Castagna, D. Colazzo, and K. Nguyen. Type-based XML projection. In *32nd International Conference on Very Large Databases (VLDB 2006)*, September 2006.
- [5] A. Arion, V. Benzaken, I. Manolescu, and Y. Papakonstantinu. Structured materialized views for XML queries. In *33rd International Conference on Very Large Databases (VLDB 2007)*, September 2007.

La publication [1] présente une contribution en théorie des types. Dans cet article nous proposons une approche sémantique au sous-typage en présence de (tous les) connecteurs booléens et de types fonctionnels. Cette approche se distingue des approches traditionnelles qui soit définissent le sous-typage de manière syntaxique (au moyen d'un système formel) soit le définissent sémantiquement au moyen d'une interprétation des types dans un modèle dénotationnel non typé. La revue *Journal of the ACM* est une des revues les plus prestigieuses en informatique théorique. La publication [2] présente un langage de programmation pour bases de données permettant la prise en compte de contraintes d'intégrité. En outre, au moyen de techniques basées sur l'interprétation abstraite de programmes combinées à de la déduction automatique de théorèmes au premier ordre, le compilateur de transactions permet d'obtenir du code certifié. La revue *The VLDB Journal* est une revue dont le facteur d'impact est un des plus important dans la communauté des bases de données. Se fondant sur les travaux théoriques établis dans [1], la publication [3] présente le langage de programmation CDuce. *ACM -ICFP* est une des conférences de tout premier plan dans le domaine des langages de programmation et en particulier des langages fonctionnels. Enfin, les travaux présentés dans [4, 5] proposent différentes approches à l'optimisation de requêtes pour données au format XML. La conférence *VLDB* est une des tribunes les plus sélectives dans le domaine des bases de données.

Les lettres jointes en annexe complètent et attestent de la qualité du travail mené tout au long de ma carrière.

Encadrement.

Le pourcentage de participation à l'encadrement est indiqué entre crochets lorsque ce n'est pas 100%.

Encadrement post-doctoral

1. Dario Colazzo 2004-2005
Invité par mes soins dans le cadre du contrat RNTL GraphDuce dont j'étais responsable.
Maître de conférences à l'université de Paris-Sud depuis septembre 2005.

Thèses

1. Mathias Kende, *Programmation fonctionnelle et GPU*, depuis Septembre 2010, [33%]
2. Wael Khemiri, "*Workflow Interactifs pour données scientifiques*", depuis Septembre 2008 [33%]

1 post-doc, 20 dea - master recherche, 10 thèses ;

3. Kim Nguyen, “*Langages de combineurs pour XML: conception, typage, implantation*”, 2004-2008 (thèse soutenue en Mai 2008) [50%]
Maître de conférences à l’Université de Paris-Sud depuis septembre 2010 (avait le choix avec un poste de CR2 à l’INRIA Rhône-Alpes)
4. Andrei Arion, “*XML access modules: toward physical independence in XML databases XML*”, 2004-2007 [50%].
"Software developer" chez Scort.
5. Cédric Miachon. *Requêtes dans CDuce : implantation, optimisation et mesures de performances*, 2003-2006.
"EQ Manager" chez Courtanet.
6. Marwan Burelle. *Sécurité de transformations XML* (depuis septembre 2002) Co-encadrement avec G. Castagna, LIENS, ENS [50%]. (Abandon pour raisons familiales).
7. S. Praud 1997-2000. Co-encadrement avec S. Cerrito (LRI)
Sujet : “Contraintes temporelles, analyse statique et déduction automatique”. Thèse soutenue en Février 2002, Université de Paris XI. [50%]
"Software R&D Bid and Project Manager" chez Thalès.
8. F. Watez 1997-2000 co-encadrement avec S. Cluet (Projet Verso-INRIA).
Sujet : “Optimisation de requêtes OQL”.
Thèse effectuée sous convention CIFRE (Société O₂Technology), Thèse soutenue en Novembre 2000. [50%]
Enseignant chercheur à l’INAPG.
9. X. Schaefer, 1994-1997
Sujet : “Bases de données orientées objet, Contraintes d’intégrité et Analyse statique”.
Thèse soutenue le 15 Octobre 1997, Université de Paris 1 - Panthéon - Sorbonne.
Ingénieur chez Devoteam.
10. P. Y. Policella de 1992 à 1994 co-encadrement en collaboration avec A. Doucet. Thèse soutenue le 11 Juin 1996, Université d’Orsay.
Sujet : “Cohérence dans les bases de données orientées objet”. [50%]

DEA - Master Recherche

1. Mathias Kende, “Programmation fonctionnelle et GPU”, Année 2010, [33%].
2. Elise Hénaux, “Validation sémantique de données au format XML pour la norme CGNS”, Année 2004-2005
3. Kim Nguyen, “Algèbres de filtrage pour l’interrogation de documents XML”, Année 2003-2004 [50%]
4. Andrei Arion, “Optimisation à base d’index et mises à jour dans un système de gestion de données XML”, Année 2003-2004 [50%].
5. Cédric Miachon. “Requêtes dans CDuce : implantation, optimisation et mesures de performances”, Année 2002-2003.
6. Marwan Burelle. “Sécurité de transformations de documents XML” DEA I₃, Année 2001-2002 [50%]

7. Alain Frisch. "Types récursifs, combinaisons booléennes et fonctions surchargées: application au typage de XML" DEA Programmation. Année 2000-2001. [50%]
8. M. Pasquier (98-99)
Sujet : "Vérification statique d'assertions en java".
9. A. Diallo (98-99)
Sujet : "étude de quelques calculs pour la mobilité".
10. H. Benajam (97-98)
Sujet : "Intégrité et test de logiciel".
11. S. Praud (96-97)
Sujet : "Contraintes d'intégrités temporelles et analyse statique de programmes".
12. R. Zhang (96-97)
Sujet : "Analyse statique du langage DEL : une application à la gestion de l'intégrité pour le système Validity".
13. N. Hugon (95-96) en collaboration avec E. Pelz.
Sujet : "Vers une sémantique pour les règles actives via le Petri-box calculus".
14. L. Tonnelier (94-95)
Sujet : "Contraintes d'intégrité et OQL: vers une intégration de contraintes d'intégrité dans le système O₂".
15. M. Sarkis (93-94)
Sujet : "Contraintes d'intégrité dans les bases de données orientées objet : approche procédurale versus déclarative."
16. R. Souah (93-94)
Sujet : "Migration d'objets en présence de contraintes d'intégrité".
17. S. Montiès (92-93)
Sujet : "Algorithmes de vérification de contraintes d'intégrité fondés sur des méthodes de simplification : application aux quantificateurs existentiels".
18. X. Schaefer (92-93)
Sujet : "Interprétation abstraite de langages persistants : une application à la satisfaction de contraintes d'intégrité".
19. S. Attali (91-92)
Sujet : "Contrôle d'intégrité et méthodes de vérification"
20. P. Y. Policella (91-92)
Sujet : "Etude et définition de contraintes d'intégrité dans un environnement orienté-objet".

Direction de stages pour élèves normaliens

- E.-O. Lebigot, G. Lecerf. "Génération de Bases Abstraites à Partir de Définitions de Type et d'une Contrainte Logique". Année 1994-1995. [50%]

Responsabilités, animation et rayonnement scientifique.

Instances d'évaluation

1. Janvier 2011, Membre, en qualité d'**expert scientifique** de la commission d'évaluation **AERES** des laboratoires marseillais (LIF et LSIS).
2. Novembre 2009, évaluation d'un projet de **collaboration COFECUB (partenariats avec le Brésil)**.
3. Septembre 2009, Membre de la commission pour la **prime d'excellence scientifique (PES)** 27ème section.
4. Octobre 2006 **rapporteur d'une proposition pour le Netherlands Organisation for Scientific Research (NWO)**.

Jurys de concours

1. Mai 2010, Membre du **jury du concours CR1 pour l'Inria Rocquencourt**.
2. Mai 2010, Membre du **jury du concours CR2 pour l'Inria Rocquencourt**.
3. Avril 2010, Membre du **jury du concours CR1 pour l'Inria Lille**.
4. Avril 2010, Membre du **jury du concours CR2 pour l'Inria Lille**.
5. Mai 2009, Membre du **comité de sélection poste de Maître de Conférences**, IUT, Paris 11.
6. Mai 2009, Membre du **comité de sélection poste de Professeur**, INPG
7. Mai 2008, Membre du **jury du concours CR2 pour l'Inria Saclay**.
8. Mai 2008, Membre du **jury du concours CR2 pour l'Inria de Rennes**.
9. Mai 2007 Membre du **jurys du concours CR2 pour l'Inria de Rennes**.
10. Mai 2007 Membre du **jurys du concours CR1 pour l'Inria de Rennes**.
11. Mai 2005, Membre du **jury du concours CR2 pour l'Inria de Rennes**.
12. Mai 2004, Membre du **jury du concours CR2 pour l'Inria de Rennes**.
13. Mai 2004, Membre du **jury du concours CR2 pour l'Inria de Futurs**.
14. Mai 2004, Membre du **jury Post-doc de l'Inria**.
15. Mai 2003, Membre du **jury du concours CR2 pour l'Inria de Rennes**.

Jurys et rapports de thèses

1. **Rapporteur** de la thèse de Muath Alrammal, Université de Paris 12, Avril 2011.
2. **Rapporteur** de la thèse d'Everardo Bárcenas Patiño, Université Joseph Fourier, janvier 2011.
3. **Rapporteur** de la thèse de H. Idabal, Université de Paris 1 - Panthéon - Sorbonne, octobre 2010.
4. Membre du jury de thèse de Radu Tofan, Université de Bordeaux- I, septembre 2010.

Instances d'évaluation AERES, PES ; Participation à 31 jurys de thèse ou d'HDR dont 10 en tant que rapporteur ; Membre de 12 jurys de concours INRIA, Présidente du Comité de programme de BDA'2005, Membre de plus de 19 comités de programme (ICDE, VLDB, XSYM, IS-MIS). Coéditeur d'un numéro spécial de Techniques et Sciences Informatiques; coéditeur des actes du 4ème International Workshop on Persistent Object Systems, membre du comité de programme du workshop international Xime-P

Organisatrice de journées et ateliers dont 2 internationaux. Lecteur pour nombreuses revues et conférences internationales.

5. Membre du jury de thèse de A. Benameur, Université Claude Bernard Lyon 1, janvier 2010.
6. **Rapporteur** de la thèse de P. ElKhoury, Université Claude Bernard Lyon 1, décembre 2009.
7. Membre du jury de thèse de Y. Gripay, Université Claude Bernard Lyon 1, décembre 2009.
8. Membre du jury de l'Habilitation à diriger des recherches de Ioana Manolescu, Université de Paris 11, octobre 2009.
9. Membre du jury de la thèse de Bogdan Marinoiu, Université de Paris 11, Juin 2009.
10. Membre du jury de la thèse de V. Sans, Université de Cergy, Décembre 2008.
11. Membre du jury de la thèse de M. Sozeau, Université de Paris 11, Décembre 2008.
12. Membre du jury de la thèse de N. Preda, Université de Paris 11, Mai 2008.
13. Membre du jury de l'Habilitation à diriger des recherches de Mirian Halfeld, Université de Blois, décembre 2007.
14. Membre du jury de la thèse de N. Travers, soutenue le 9 décembre 2006 à l'Université de Versailles Saint Quentin.
15. Membre du jury de la thèse de L. Pilard, soutenue le 15 décembre 2005 à l'Université de Paris 11.
16. **Rapporteur** de la thèse de D. Debarbieux, soutenue le 9 décembre 2005 à l'Université de Lille.
17. **Rapporteur** de la thèse de S. Medjdoub, soutenue le 8 décembre 2005 à l'Université de Versailles Saint Quentin.
18. Membre du jury de la thèse de N. Gesbert soutenue le 2 décembre 2005 à l'Université de Marne la Vallée.
19. **Rapporteur** de la thèse de D. Duarte, soutenue en juillet 2005 à l'Université de Francois Rabelais Blois.
20. Membre du Jury de la thèse d'O. Benjelloun, soutenue en septembre 2004 à l'Université de Paris 11.
21. Membre du Jury de la thèse de P. Letouzey, soutenue le 10 Juillet 2004 à l'Université de Paris 11.
22. Membre du Jury de la thèse de J. Forest, soutenue le 29 Septembre 2003 à l'Université de Paris 11.
23. Membre du Jury de la thèse de P-A. Veltri, soutenue le 1 Octobre 2002 à l'Université de Paris 11.
24. Membre du Jury de la thèse de J. Barbay, soutenue le 24 Septembre 2002 à l'Université de Paris 11.
25. **Rapporteur** de la thèse de L. Nemirovski, soutenue à l'Université de Paris 1 Janvier 2002.
26. Membre du jury de la thèse de F. Watez, Soutenue à l'Université de Paris 1 le 31 Octobre 2000.

27. Membre du jury de la thèse de D. Dvekautsas, Soutenue à l'Université de Paris XI le 15 Juin 1999.
28. **Rapporteur** de la thèse de A. Ait-Bouziad, Soutenue à l'Université de Paris VII, le 1 Juin 1999.
29. **Rapporteur** de la thèse de J. Darmont, Soutenue à l'Université de Clermont-Ferrand, le 18 Janvier 1999.
30. **Rapporteur** de la thèse de F. Bentayeb, Soutenue à Université d'Orléans en Juin 1998,
31. Membre du jury de la thèse de F. Semmak soutenue à l'Université de Paris 1 le 24 Fevrier 1998.
32. Membre du jury de la thèse de P.Y. Policella soutenue à l'Université d'Orsay le 11 Juin 1996.

Comités de rédaction

1. Membre du comité de rédaction de la revue I_3 , depuis 1999.
2. Membre du comité de rédaction de la revue TSI, de Janvier 1996 à Janvier 2000.
3. Membre du comité de lecture du numéro spécial de la revue L'Objet, "Méthodes formelles pour les systèmes à objets", 1999.

Comités de pilotage

1. Membre du comité de pilotage de BDA depuis 2004

Comités de programme

1. Membre du comité de programme de la conférence Bases de Données Avancées (BDA) 2011.
2. Membre du comité de programme de international conference on data engineering **ICDE, 2011**.
3. Membre du comité de programme de l'International XML Symposium (**XSYM 2010**)
4. Membre du comité de programme de la conférence internationale **VLDB, 2009**.
5. Membre du comité de programme de l'**ACM** International workshop on Programming Languages for XML (**PLAN-X 2008**).
6. Membre du comité de programme de la conférence Bases de données avancées (BDA 2007).
7. Membre du comité de programme de l'**ACM** - International workshop on Xquery Implementation and Experience **XIME-P 2007**.
8. Membre du comité de programme de l'International XML Symposium (**XSYM 2007**)
9. Membre du comité de programme de l'International XML Symposium (**XSYM 2006**)
10. Membre du comité de programme de l'International XML Symposium (**XSYM 2005**)

11. Membre du comité de programme de l' International Workhop on Programming Language technologies for XML (PLAN-X 2005) in conjunction with ACM POPL'2005.
12. Membre du comité de programme de l' International Symposium on Methodologies for Intelligent Systems (ISMIS) 2005
13. **Présidente du comité de programme** de la conférence Bases de Données Avancées (BDA) 2005.
14. Membre du comité de programme du workshop on XQuery, **XIME-P** joint with ACM SIGMOD-PODS, Paris Juin 2004.
15. Membre du comité de programme de la conférence Bases de Données Avancées (BDA) 2003.
16. Membre du comité de programme de la conférence Bases de Données Avancées (BDA) 2002.
17. Membre du comité de programme de la conférence Bases de Données Avancées (BDA) 1999.
18. Membre du comité de programme de la conférence Bases de Données Avancées (BDA) 1996.
19. **Co-chair** du comité de programme de International Workshop on Persistent Object Systems (POS6), 1994.
20. Membre du comité de programme de la conférence Bases de Données Avancées (BDA) 1993.
21. Membre du comité de programme de la conférence Inforsid, 1992.

Comités d'organisation

1. **Co-responsable** de l'organisation de "Persistent Object Systems 6" (POS6), 5-9 Septembre 1994, Tarascon, France.
2. Participation à l'organisation du "2nd Workshop on Foundation of Object-Oriented Languages", Paris juillet 1994.
3. Organisation du meeting des partenaires francophones du réseau d'excellence RENOIR, Décembre 1997, Paris.

Autres

- Lecteur pour les revues Journal of Functional Programming, Information Systems, Theoretical Computer Science, Theory and Practice of Object-Oriented Systems et pour les conférences DBPL, PODS, VLDB, EDBT, ICDT, ICALP, ESOP, ECOOP.

Conférences et séminaires invités

La personne qui m'a invité est indiquée entre parenthèses carrées.

Invitée, pour un court séjour, à IBM T-J Watson Research, [Jérôme Siméon], Avril 2011.

Participation au workshop RADICAL 2010, sur **invitation uniquement**, Microsoft Research Cambridge. [Andy Gordon] Mai 2010.

“Forward and backward analysis of database programming languages”, Séminaire invité, [A. Carbone] Université de Paris 12, Mai 1998.

“Forward and backward analysis of database programming languages”, Séminaire invité, [J. Vitek] Université de Genève, Suisse, Septembre 1997.

“Static verification of integrity constraints in object-oriented database programming languages via predicate transformers”, Séminaire invité, Sémantique et Interprétation Abstraite [P. Cousot], École Normale Supérieure, Avril 1997.

“Systèmes de Bases de données orientées-objet”, Tutorial invité. Conférence Inforsid 1991 [Colette Rolland].

GDR Bases de Données

1. De 1995 à 1998 responsable du serveur Web du GDR BD3.
2. De 1995 à 1998 modérateur de la “mailing list” du GDR BD3.

Animation de la recherche

1. De Septembre 2002 à la fin des RTP, membre du **comité de pilotage du RTP Bases de données et d’informations hétérogènes et distribuées (RTP 9)**.
2. Janvier 2003, **rapporteur extérieur** pour l’évaluation d’une **proposition RNTL**.
3. Octobre 2006 **rapporteur d’une proposition pour le Netherlands Organisation for Scientific Research (NWO)**.
4. Novembre 2009, évaluation d’un projet de **collaboration COFECUB (partenariats avec le Brésil)**.

Contrats et collaborations.

1989-1992: membre du **projet ESPRIT BRA (Basic Research Action) n° 3070 FIDE**.

1997-1999 membre du **réseau d’excellence RENOIR (Requirement Engineering Network On International cooperating Research groups)**.

1999-2000 Membre du groupe “Spécification, vérification, sémantique”, pôle *Objets et composants logiciels*, GDR ALP (1999).

2003-2004 Responsable d’un projet d’action spécifique “Langages, types, sécurité et intégrité pour données semi-structurées”, dans le cadre du RTP 9 acceptée en mai 2003

2003-2006 Responsable pour l’équipe Bases de données d’une proposition soumise à l’appel d’offre **RNTL** : “GraphDuce”, labellisée le 13 mars 2003. Le projet impliquait l’équipe langage du LIENS et la société Brixlogic.

2003-2006 Responsable pour le partenaire LRI, du projet CASC, **ACI-sécurité du CNRS-MENRT-INRIA**. Le projet impliquait : l'équipe SMIS de l'INRIA Rocquencourt, l'équipe Lanages du LIENS, ENST bretagne et l'Université de Pau.

2004-2008 Responsable pour le partenaire LRI, du projet TRALALA (TRansformation LAnguages for XML : Logics and Applications), **ACI-masse de données** du CNRS-MENRT-INRIA. Le projet impliquait l'équipe langage du LIENS, le projet Mostrare de l'INRIA Lille, le LIFL, le LIF de Marseille, le projet Gemo INRIA Saclay et l'équipe vérification du LIAFA.

2008-2011 Responsable pour l'équipe Bases de données du LRI du **projet Digitéo** Ediflow (Efficient Interactive Workflow for Scientific Data Management).

2009-2012 membre du projet **ANR Domaines Emergents - CODEX** (Performance, évolution et composition pour XML : modèles, algorithmes et systèmes)..

Réalisations, Prototypes, Valorisation, Transfert.

- Réalisation du module de regroupement d'objets sur disque pour le système de gestion de bases de données commercial O₂.
- CDuce est **intégré aux distributions Linux** suivantes : Ubuntu/Debian, Fedora et Mandriva.
- Membre de la commission Logiciel du LRI depuis 2009.

II. Activités d'enseignement.

Je ne donne pas le détail de mes activités d'enseignement assurées en qualité de maître de conférences. J'ai enseigné dans tous les cycles de l'Université tant des cours théoriques (théorie des langages et compilation) que des cours plus pratiques (bases de données, programmation). Je décris mes activités effectuées en qualité de professeur dont le tableau suivant dresse le récapitulatif.

Enseignement dans plusieurs écoles doctorales en Italie; enseignement en DEA (Paris 1, Paris 11), DESS (Paris 11), cours et travaux dirigés en DEUG Mass, Licence et Maîtrise d'informatique, Miage, Mass et DU, cycle ingénieur, tant en formation classique que par la voie de l'apprentissage ; encadrement de stagiaires et d'apprentis

Niveau	Intitulé	C/TD	Eq/TD	années
L3	Miage ; Lic. info Introduction aux bases de données	C/TD	60	1998 - 2003
L3	Miage apprentissage Bases de données	C	40	1998 - 2001
L3	Bio-info Bases de données	C	20	2002 - 2003
L3-M1	Encadrement 10 apprentis	autre	bénévolat	1998 - 2000
M1	VAE - 3 professionnels	autre	bénévolat	2002 - 2003
M1	C4-DU Bases de données	C	30	2001 - 2003
M1	Miage apprentissage Objet méthodes et langages	C	22,5	1998 - 2001
M1	cycle ingénieur apprentissage BD	C	30	2004 - 2005
M1	cycle ingénieur BD système	C	30	2004 - 2009
M1	VAA - 300 dossiers	autre	bénévolat	2009-2010
M1	cycle ingénieur BD système	C	27	2009 - 2011
M2	cycle ingénieur prog internet	C	22,5	2004 - 2008
M2	cycle ingénieur BD (tuning)	C	22,5	2008 - 2009
M2	Master Pro Gestion de l'information sur internet	C	18	2004 - 2010
M2	Master Pro XML et Programation Internet	C	18	2010 - 2011
M2	M2R Fondement des BD	C	12	1998 - 2003
M2	M2R Langages et données	C	18	1998 - 2010
M2	M2R Données et connaissances sur le Web	C	6	2010 - 2011
M2	M2R Modèles et Langages pour systèmes	C	18	2010 - 2011
M2	M2R Coordination stages	autre	5	2010 - 2011
M2	M2pro Encadrement stage	autre	15	2004 - 2009
M2	M2pro Encadrement stage	autre	6	2009 - 2010
M2	cycle ingénieur Encadrement stage	autre	35	2004 - 2011

1998 à ce jour : Professeur Université de Paris 11

Niveau L

Licence d'Informatique et IUP Miage. cours "Introduction aux bases de données". Université de Paris 11 (1998-2003). J'ai assuré la coordination de six groupes de TD. (160 étudiants)

Licence Miage par apprentissage cours "Bases de données". Université de Paris 11, (1998-2001). J'ai assuré la coordination de 2 groupes de TD. (40 étudiants)

Licence Miage classique. TD du cours "Bases de données". Université Paris 11, (1998-1999) et (2001-2002).

C4-DU Diplôme d'Université en Informatique, cours de "Bases de données", Université de Paris 11, (2001-2003). Ce cours était dispensé a' des étudiants issus de filières non informatique. J'ai assuré la coordination de deux groupes de TD. (60 étudiants).

Licence de Bio Informatique cours de "Bases de Données". Université de Paris 11, (2002-2003). (35 étudiants)

IUP-Miage Tutrice de dix apprentis et encadrement de stages annuellement.

Miage : VAE Responsable de trois professionnels dans le cadre de la VAE.

Niveau M : Professionnel

VAA - M₁ Etude des dossiers (600 dossiers) de Validation des Acquis Académiques pour l'entrée en M₁. J'ai assuré, bénévolement, la sélection de 300 dossiers.

Ecole d'ingénieurs IFIPS (niveau M2) Cours aspects systèmes des bases de données : concurrence, reprise, distribution et tuning. 2008-2009. Un groupe de TD.(30 étudiants).

Master Pro M2 Cours de Gestion de l'information sur Internet (12h) de 2004 à 2010. Ce cours est commun aux master pro, miage et reseaux. Je coordonne 4 groupes de TD en moyenne (80 étudiants).

Master Pro M2 Cours de XML et Programmation Internet (12h) depuis 2010. Ce cours est commun aux master pro, miage et reseaux. Je coordonne 4 groupes de TD en moyenne (80 étudiants). La page décrivant le contenu du cours est consultable à :

<http://www.lri.fr/benzaken/giinal.en.html>

Ecole d'ingénieurs IFIPS (niveau M1) Cours de Bases de données avancées. Depuis 2004. La page décrivant le contenu du cours est à

<http://www.lri.fr/benzaken/ifips.en.html> (30+30 étudiants)

Ecole d'ingénieurs IFIPS (niveau M2) Cours de Nouvelles architectures logicielles (15h), de 2004 à 2008. (25 étudiants)

Maîtrise Miage par apprentissage cours "Objets, méthodes et langages" Université Paris 11, (1998-2000) (40 étudiants).

Encadrements de stages Depuis 1998 à raison de cinq à six stagiaires par an en moyenne.

DESS Ingénierie Informatique Université de Paris 11, cours "Bases de données avancées", (2000-2001).

Dess Encadrement de dix stages (année 2000-2001).

Niveau M : Recherche

Master Recherche NSI Cours "Systèmes enfouis, distribués, parallèles, internet : modèles, langages, programmation", **Responsable** des 24h de cours. Depuis 2010.

Master Recherche IAC Cours du tronc commun "Données et connaissances sur le Web" (4h) depuis 2010.

Master Recherche NSI Responsable de la coordination des stages. Depuis 2010.

Master Recherche Cours "Langages et systèmes pour bases de données" (depuis 2004). L'intitulé de ce cours a changé il s'agissait auparavant du cours "Sûreté, Intégrité et Sécurité des Données". **Responsable**.

DEA d'Informatique I3: information, interaction, intelligence Universités Paris 1, Paris 11. Cours "Sûreté, Intégrité et Sécurité des Données" en collaboration avec G. Castagna (1998-2003).

DEA d'Informatique I3: information, interaction, intelligence Universités Paris 1, Paris 11. Cours "Fondements des bases de données", en collaboration avec N. Bidoit et N. Spyrtos(1999-2003)

DEA Théorie et Ingénierie des Bases de Données. Universités Paris 11 - Paris 1. Module "Langages de Programmation pour Bases de Données". 1998-1999.

Écoles doctorales

Scuola Nazionale dei Dottorati di Informatica delle Facoltà di Scienze. Cours “Linguaggi per basi di dati ad oggetti”, Bertinoro (Italie), 1999 (8 heures) [en Italien].

III. Responsabilités collectives.

Période 1990-1998 : Maître de conférences Université de Paris 1

1. Membre du **Conseil d'Administration** de l'Université de Paris 1 de mars 1993 à mars 1997.
2. Membre du **Conseil de l'U.F.R. de Mathématiques et Informatique** de Paris 1 de Juin 1992 à Septembre 1998.
3. Membre du **Bureau du Conseil de l'U.F.R. de Mathématiques et Informatique** de Paris 1 d' Octobre 1992 à Septembre 1998.
4. Responsable, pour l'U.F.R., de 1992 à 1994, de **l'organisation des enseignements d'Informatique** dans les diverses filières de Paris 1.
5. **Assesseur** de la **Commission de Spécialistes** Mathématiques et Informatique de l'Université de Paris 1, d' Avril 1995 à Septembre 1998.
6. Membre de la **Commission de Spécialistes** Mathématiques et Informatique de l'Université de Paris 1 de Septembre 1993 à Septembre 1998.
7. Membre de la **Commission de Spécialistes** Informatique de l'Université d'Orléans de Mai 1994 à Mai 1995.
8. Membre de la **Commission de Spécialistes** Informatique de l'Université de Paris 12 (Créteil), de Janvier 1996 à Septembre 1998.
9. Membre de la **Commission de Spécialistes** Informatique de l'Université de Versailles de Décembre 1996 à Décembre 1997.
10. Membre de la **Commission de documentation** de l'Université de Paris 1 d' Avril 1994 à mars 1997.

Période 1998 à ce jour : Professeur Université de Paris 11

Responsabilités locales à l'établissement

1. **Présidente du comité de sélection** du poste de Maître de conférences n° 600, UFR des sciences Université de Paris-sud 11.
2. Depuis septembre 2010, **Coordination des stages** M2 Recherche NSI
3. Mai 2010, Membre du Bureau (VP collège A) du **comité de sélection poste de Maître de Conférences - UFR**, Université de Paris Sud XI.
4. Membre de la **commission Logiciel du LRI** depuis 2009.
5. Mai 2009, Membre du **comité de sélection poste de Maître de Conférences**, IUT, Paris 11.
6. Membre titulaire de la **Commission de Spécialistes** Informatique puis **CCSU** de l'Université de Paris 11 depuis décembre 2003.

Maître de conférences :
Membre du conseil d'administration, membre de 4 commissions de spécialistes, membre du conseil d'UFR et de diverses commissions de l'Université.
Professeur :
Membre du conseil de département, bureau du département enseignement, membre du conseil de laboratoire du LRI, membre de la commission de spécialistes, directrice de l'IUP-Miage, membre du conseil scientifique, membre de la commission “Logiciel” du LRI, membre élue du CTPU.

7. **Directrice de l'IUP-Miage** de 1998 à 2003. J'ai bénéficié, pour cela, d'une décharge de 64heq/TD. L'IUP est une filière du Département d' Informatique, il regroupait un effectif annuel moyen de 230 étudiants répartis en deux sous-filières : la miage classique (150 étudiants) et la miage par apprentissage (80 étudiants).

Le service Miage était composé d'une responsable administrative, d'une secrétaire pédagogique (rétribuée sur nos fonds propres), d'un ingénieur et d'un technicien. J'étais, en outre, secondée dans ma tâche par un directeur des études, un coordinateur de l'apprentissage et un coordinateur des enseignements d'informatique. Je décris, en Annexe, de façon plus détaillée la charge que représentait, alors, la direction de l'IUP.

De 1998 à 2003, j'ai consacré quatre jours par semaine en moyenne à la gestion pédagogique et administrative de l'IUP-Miage. Les autres responsabilités ne m'ont occupée qu'une demi journée par mois.

8. Membre du **Conseil du département enseignement** du département d'Informatique de Paris 11 de 1998 à 2003.
9. Membre du **Bureau du Conseil du département enseignement** du département d'Informatique de Paris 11 de 1998 à 2003.
10. Membre du **Conseil du laboratoire** du LRI, UMR 8623 CNRS de 1999 à 2008..
11. Membre suppléant de la **Commission de Spécialistes** Informatique de l'Université de Paris 11 de mars 2001 à décembre 2003.
12. Membre du **Conseil Scientifique** de l'Université de Paris 11, de mars 2002 à mars 2005.
13. Membre du **Conseil de la recherche de l'UFR des Sciences** de mars 2004 à mars 2005.

Responsabilités nationales

1. Membre du **comité de sélection** du poste de Maître de conférences n°0670 INPG, Mai 2011.
2. Membre du **comité de sélection** d'un poste de Professeur Paris 1, Mai 2011
3. Janvier 2011, Membre, en qualité **d'expert scientifique** de la commission d'évaluation **AERES** des laboratoires marseillais (LIF et LSIS).
4. de 2003 à 2010, j'ai été **membre de 12 jurys de concours CR1 et CR2** pour pour les unités de recherche de l'INRIA de Rennes, Saclay, Lille et Rocquencourt.
5. Mai 2010, Membre du **comité de sélection poste de Maître de Conférences**, Université de Versailles Saint Quentin - IUT de Velizy.
6. Mai 2010, Membre du **comité de sélection poste de Maître de Conférences**, Université de Paris IX - Dauphine.
7. Mai 2009, Membre du **comité de sélection poste de Professeur**, INPG - Grenoble.
8. Novembre 2009, évaluation d'un projet de **collaboration COFECUB (partenariats avec le Brésil)**.
9. Septembre 2009, Membre de la commission pour la **prime d'excellence scientifique (PES)** 27ème section.
10. Mai 2004, Membre du **jury Post-doc de l'Inria**.
11. Membre élue du Comité Technique Paritaire Universitaire **CTPU, Instance nationale** en charge de l'étude de nos **statuts**, de décembre 2003 à 2006.
12. Je participe régulièrement à l'évaluation pour le MESR de maquettes de réhabilitations ou de projets PPF.

Depuis 2004, en moyenne, les tâches d'intérêt collectif m'occupent deux jours par semaine.

ANNEXES.

Dans cette partie, je décris de façon plus détaillée la charge que représentait la direction de l'IUP-Miage, l'ensemble de mes thématiques de recherche, je donne ma liste de publications depuis mon passage à la première classe ainsi que ma liste complète de publications. Enfin, j'ai joint des lettres de soutien à ma candidature tant pour la partie recherche que pour la partie administration de mon dossier.

Direction de l'IUP-Miage (1998-2003)

La direction de l'IUP-Miage a représenté 85% de ma charge administrative à l'époque. Je précise que cette responsabilité a donné lieu à une décharge de service de 64h eq/TD. L'IUP-Miage, est depuis la rentrée 2004 rentré dans le schéma Master de l'Université de Paris XI.

L'IUP-Miage d'Orsay en trois mots et quelques chiffres...

L'Institut Universitaire Professionnalisé (IUP) Miage (Maîtrise de méthodes informatiques appliquées à la gestion) de l'Université de Paris 11 était une filière du Département d'Informatique qui en comptait neuf (Deug Mias, Licence-Maîtrise d'informatique, NFI FIIFO, le DU d'informatique, deux DESS et deux DEA).

L'IUP regroupait un effectif annuel moyen de 230 étudiants répartis en deux sous-filières : la miage classique (150 étudiants) et la miage par apprentissage (80 étudiants). De part le volume de ses effectifs et ses trente ans d'existence, la Miage d'Orsay est une des plus importantes Miage de France¹.

Le service Miage était composé d'une responsable administrative, d'une secrétaire pédagogique (rétribuée sur nos fonds propres), d'un ingénieur et d'un technicien. J'étais, en outre, secondée dans ma tâche par un directeur des études, un coordinateur de l'apprentissage et un coordinateur des enseignements d'informatique. En dernier lieu, je tiens à préciser que le département d'informatique était cruellement sous-encadré en particulier en enseignants-chercheurs.

La dimension professionnalisante

Le processus de recrutement

Le processus de recrutement de l'IUP-Miage débute au mois de janvier et se termine en juin. Nous recrutons des étudiants à Bac+1 et à Bac+2. Il convient donc d'opérer un travail de communication fort, travail qui consiste à se déplacer en premier cycle (Deug Mias, Mass) tant à l'Université de Paris 11 que dans d'autres Universités parisiennes. Pour le recrutement au niveau Bac+2, nous nous déplaçons, en outre, dans les IUT de la région parisienne afin de présenter notre filière.

¹Voir enquête annuelle réalisée pour la conférence des directeurs des 20 miages de France

Ensuite le processus de recrutement stricto-sensu se déroule en trois phases. La première phase consiste en l'organisation de tests à l'entrée puis en un jury purement académique qui sélectionne les dossiers qui sont acceptés à l'IUP (en moyenne nous avons à étudier de l'ordre de huit cent dossiers par an dont une trentaine seront acceptés en IUP première année, une trentaine en IUP deuxième année et une cinquantaine en IUP deuxième année par la voie de l'apprentissage.

Les deuxième et troisième phases concernent uniquement le recrutement pour la filière en apprentissage. La deuxième phase consiste en l'organisation d'entretiens devant un jury composé d'un universitaire et d'un membre du centre de formation par apprentissage (CFA). Ces entretiens se déroulent sur deux à trois jours et permettent de constituer une liste d'admis ainsi qu'une liste d'attente. La troisième phase consiste à affecter les apprentis aux postes qui seront les leurs au sein des entreprises partenaires du CFA de l'Affia avec qui nous collaborons depuis bientôt dix ans.

j'ai consacré à cette activité, de janvier à juin, environ une journée et demi par semaine.

Prise en compte du parcours professionnel

Le statut d'IUP impose un certain nombre de contraintes qui ont un impact direct sur le déroulement de l'année scolaire. En particulier, pour la filière Miage classique, la dimension professionnelle de ce type de filière requiert l'organisation de périodes de stages (deux mois en première année, trois mois en deuxième année (licence) et trois mois en troisième année (maîtrise). La prise en compte du parcours professionnel de l'étudiant donne lieu à l'organisation de soutenances de stages.

Pour la filière en apprentissage, la notion de stage n'existe pas mais, afin de prendre en compte l'aspect professionnel de ce type de formation, nous avons créé une unité d'enseignement "synthèse de l'apprentissage" qui donne lieu à une soutenance dont le jury est composé du maître d'apprentissage de l'étudiant et d'un universitaire.

Enfin, des soutenances particulières doivent être mises en place pour l'obtention du titre d'ingénieur-maître.

j'ai consacré à cette activité, de mai à mi-septembre une demi journée par semaine.

Mise en place de nouveaux enseignements et campus numérique

Nous avons depuis la rentrée 2001 mis en place des enseignements en communication pour un volume global de 500h Eq/TD. J'ai assuré la recherche ainsi que le recrutement des vacataires pour ces enseignements. De janvier 2000 à novembre 2000, *j'ai consacré à cette activité deux heures par semaine.*

Enfin, dans le cadre des projets "campus numérique", l'IUP-Miage a intégré le projet "International E-Miage".

Gestion administrative et financière

Dans le cadre du contrat d'établissement nous avons été amenés à rédiger le dossier de réhabilitation de la filière en 2000-2001. Une telle tâche nous a occupé de janvier à juillet environ un jour par semaine.

Chaque année, afin d'obtenir des crédits, nous devons répondre à une série d'appels d'offre internes à l'Université. Il faut noter que l'IUP-Miage ne touche que très partiellement et par ce biais une partie de sa dotation SanRemo (les bonnes années nous avons obtenus 150KF.

De novembre à février, nous sollicitons, par l'envoi de courriers ciblés ou de visites, de nombreuses entreprises afin de collecter de la taxe d'apprentissage.

La préparation des budgets primitifs et des DBM (Décisions Budgétaires Modificatives) nous occupe tout au long de l'année.

Enfin, le recrutement de vacataires et de professionnels est un aspect important de la dimension professionnalisante de ce type de filière.

j'ai consacré globalement à tous ces aspects un jour par semaine.

L'apprentissage : la convention

À l'IUP-Miage, le dispositif de l'apprentissage repose sur deux partenaires : l'Université représentée par l'IUP et le Centre de Formation par Apprentissage (CFA) de l'Affia (Association pour la Formation d'Informaticiens par la voie de l'Apprentissage). Ces deux partenaires passent une convention pour chaque promotion recrutée. Cette convention a pour objet d'établir quelles sont les compétences et responsabilités de chacune des parties et de préciser les conditions financières du partenariat.

Lors de ma prise de fonction, en 1998, j'ai naïvement pensé que le texte de la convention était reconduit d'année en année et qu'il s'agissait d'un processus bien rodé. Je me trompais. Chaque année le texte de la convention fait l'objet d'une longue et âpre négociation. Ce processus m'a conduit à re-travailler en profondeur ce texte en étroite collaboration avec mon collègue B. Goldfarb Directeur de l'UFR informatique et gestion de Paris 9 et directeur de la Miage de Dauphine. Parallèlement, le CFA de l'Affia ayant élargi, en trois ans, ses collaborations à six miages parisiennes (Paris 1 et Paris 9 en 1999, Paris 5 et Evry en 2000 et Paris 10 en 2001), j'ai souhaité travailler en réseau avec mes collègues comme c'est l'usage dans la communauté des directeurs de Miage. Ce travail s'est effectué sur deux plans : le plan juridique (étude des textes régissant l'apprentissage dans le supérieur) et le plan financier (gestion analytique des coûts du dispositif apprentissage). Ceci nous a conduit à élaborer un même texte de convention pour les six Miages partenaires et surtout des conditions financières identiques pour tous. À ce jour, seul le volet financier a atteint son objectif : les six miages parisiennes sont sur un même pied d'égalité.

j'ai consacré à cette tâche une demi journée par semaine.

L'apprentissage : le rythme d'alternance et l'implication pédagogique

Bien que le maître mot en matière de transmission des savoirs à l'Université, et tout particulièrement en second cycle, soit la continuité, le rythme d'alternance souhaité par nos

partenaires du CFA est de trois semaines/trois semaines. L'étudiant-apprenti alterne toutes les trois semaines une session à l'Université et une session en entreprise. Ceci induit un certain nombre de conséquences. D'un point de vue purement logistique, il n'est pas possible de prévoir un emploi du temps par semestre. Seules les grandes lignes sont planifiables (quels cours, quels TD). L'emploi du temps est préparé session par session. Je supervise, en collaboration avec mon directeur des études, cette tâche.

Sur le plan purement pédagogique, les étudiants souffrent très rapidement d'un tel rythme. Ceci se traduit par une forme de dissociation entre les deux mondes et conduit à de réelles difficultés pour l'acquisition de savoirs fondamentaux pouvant conduire à une détérioration de la qualité du diplôme national Miage. Afin d'amortir l'impact néfaste de ce rythme, nous avons mis en place une réunion de bilan de session, (à laquelle participent les délégués étudiants, la coordinatrice de l'apprentissage, un chargé de missions du CFA et moi-même) ainsi que du tutorat enseignant. Chaque apprenti bénéficie des conseils et de l'aide personnalisée d'un enseignant-chercheur du Département d'Informatique. Un tel dispositif est lourd à mettre en place et à faire vivre. *j'ai consacré à cet aspect une demi-journée par semaine.*

De 1998 à 2003, j'ai consacré quatre jours par semaine en moyenne à la gestion pédagogique et administrative de l'IUP-Miage. Les autres responsabilités ne m'ont occupée qu'une demi journée par mois. Depuis 2004, en moyenne, les tâches d'intérêt collectif m'occupent deux jours par semaine

Thématiques de Recherche

Doctorat

Aspects système des systèmes de gestion de bases de données

En 1987, j'ai rejoint, dans le cadre de ma thèse, le GIP Altaïr (INRIA). J'ai participé à la conception puis à la réalisation du système de gestion de bases de données orientées objets O₂. Ma contribution a particulièrement porté sur des problèmes de gestion du disque [33, 48, 47, 31, 46, 32, 3] et a consisté à définir et implanter des politiques de placement de données afin de minimiser les surcoûts d'entrée-sortie.

À partir de 1989, le GIP s'est scindé en deux équipes : l'une chargée de l'industrialisation du prototype O₂, l'autre poursuivant une activité de recherche. Au sein de cette dernière équipe, j'ai commencé une activité de recherche plus fondamentale. Dans le cadre du projet ESPRIT II "Basic Research Action" n° 3070 FIDE², démarré en 1989, j'ai étudié plus en détail les problèmes de la définition des systèmes de gestion d'objets persistants et des langages de programmation pour bases de données. Ce projet m'a permis de collaborer avec des équipes internationales de pointe sur le domaine des bases de données et des langages de programmation persistants (Equipe de Malcolm Atkinson, Université de Glasgow ; Equipe de Ron Morrison, Université de St-Andrews ; Equipe d'Antonio Albano, Université de Pise)

Période 1990 à 1998 : Maître de Conférences Université de Paris 1

²Formally Integrated Data Environment

Langages pour bases de données et intégrité

À partir de Novembre 1991, j'ai poursuivi mes activités de recherche dans le domaine de la définition de langages de programmation pour bases de données. Dans ce cadre, j'ai plus particulièrement étudié le problème de la définition et de la vérification de contraintes d'intégrité pour les bases de données orientées objet.

Dans le domaine des bases de données il est crucial de pouvoir spécifier des propriétés (contraintes d'intégrité) que les données stockées dans la base doivent satisfaire (par exemple que toute personne est soit célibataire, soit le conjoint de son conjoint).

La plupart des langages de programmation pour bases de données tels que *Galileo*, *Napier88*, *Fibonacci* ou O_2 ne permettaient pas d'exprimer ce type de propriétés de façon globale et déclarative.

Ma contribution a consisté à étudier la possibilité d'intégrer ce type de fonctionnalités au sein d'un langage de programmation pour bases de données. Trois aspects importants ont été étudiés tant sur le plan formel que pratique : (i) définir un langage de spécification de contraintes d'intégrité, (ii) réduire statiquement (à la compilation) le nombre de contraintes d'intégrité à vérifier lors de l'exécution d'une transaction (d'un programme), et (iii) générer automatiquement des algorithmes de vérification efficaces.

Thémis

Le premier point nous a conduit à développer un langage (*Thémis*) offrant ce type de fonctionnalités. Une première version de ce langage, proposée dans [30, 29] puis reprise et étendue dans [9], permet d'exprimer de façon globale et déclarative des contraintes d'intégrité qui ne portent que sur des attributs statiques (des données stockées).

Analyse statique et déduction automatique :

Un premier mécanisme, assez rudimentaire [29, 9] (basé sur des informations relatives au typage des données) permettant de détecter *statiquement* si une transaction ne violera pas une contrainte donnée a été défini pour ce langage.

Ce type de mécanisme de détection statique de la préservation de l'intégrité proposé s'avérant assez grossier, j'ai, en collaboration avec Xavier Schaefer (dans le cadre de son stage de D.E.A., que j'ai dirigé en 1993), utilisé la théorie de l'interprétation abstraite (définie par P et R Cousot en 1976) afin d'obtenir des algorithmes de détection de non violation de l'intégrité plus précis, fiables et implantables. Nous adaptons cette théorie à la problématique des bases de données en considérant des langages de programmation persistants. Les premiers résultats de ces travaux ont été présentés dans [43, 10].

Nous avons, ensuite, approfondi ces résultats. Notre approche consiste à définir un contexte d'analyse de programme fondé sur la notion de transformeurs de prédicats (définis par Dijkstra en 1976) et de ramener le problème de la préservation de l'intégrité à celui de la preuve d'un théorème de logique du premier ordre. Plus précisément, un *transformeur de prédicat*

avant est une fonction qui prend comme arguments un programme et une formule satisfaite par les données en entrée du programme (une pré-condition) et donne comme résultat une nouvelle formule (une post-condition) qui est satisfaite par tout état qui résulterait de l'exécution du programme. La contrepartie naturelle d'un transformeur de prédicats avant est la notion de *transformeur arrière* qui génère, à partir d'une post-condition, une pré-condition suffisante sur les données en entrée du programme pour que celui ci préserve la post-condition. Ainsi, si la post-condition générée implique la contrainte d'intégrité ou si la contrainte implique la pré-condition, nous sommes sûrs que le programme préserve la cohérence des données.

Nous avons défini et implanté ces deux transformeurs distincts. Pour le transformeur avant, la méthode de déduction utilisée et également implantée pour prouver l'implication est la méthode des tableaux (car le théorème à prouver est en forme normale disjonctive, la résolution opérant sur des formules en forme normale conjonctives). Pour le transformeur arrière, la méthode de déduction utilisée est la résolution (pour des raisons symétriques).

Ceci permet de déduire, de façon plus précise, qu'un programme donné ne violera pas une contrainte [28, 42, 27, 26]. Les résultats de mesures qui ont été menées au moyen de ce prototype sont tout à fait satisfaisants et démontrent la "faisabilité" d'une telle approche. La thèse de X. Schaefer que j'ai dirigée est le fruit de ce travail.

Une évolution naturelle de notre travail a consisté à l'adapter aux contraintes d'intégrité temporelles. De telles contraintes, contrairement aux précédentes, portent sur une succession d'états de la base de données. Un exemple classique de telles contraintes est donné par : "l'âge d'une personne ne peut décroître". Nous avons défini des mécanismes de détection statique inspirés des techniques précédemment décrites combinés à des techniques de déduction automatique en logique temporelle du premier ordre. Le problème majeur qui s'est posé, dans ce contexte, a été de définir un démonstrateur de théorèmes pour la logique temporelle du premier ordre. Ceci a fait l'objet du stage de DEA de S. Praud et s'est poursuivi, dans le cadre de sa thèse que j'ai dirigée en collaboration avec S. Cerrito (LRI) [8, 41].

Génération automatique de tests :

Concernant la génération automatique d'algorithmes de vérification le problème se ramène au problème théorique suivant : étant donné une formule de la logique du premier ordre (la contrainte) et un modèle fini de cette formule (la base de données) si l'on modifie (le programme exécuté) ce modèle fini existe-t-il un moyen constructif efficace pour savoir si la nouvelle interprétation constitue à nouveau un modèle fini de la formule ? Dans le cas général, si l'on ne suppose pas que l'on dispose d'un modèle fini initialement, le problème est PSPACE-complet (réduction de QBF). Une classe de formules a été isolée (formules possédant au plus une alternance de quantificateurs existentiels/universels ou universel/existential) pour lesquelles les algorithmes proposés sont efficaces (polynômiaux) [29, 9]. S. Monties, dans son stage de D.E.A., étend ces algorithmes au cas des quantificateurs existentiels. Initialement écrits dans un style impératif, ces algorithmes ont été réécrits dans le langage de requête standard pour les bases de données orientées-objet : OQL. Ceci a fait l'objet du stage de D.E.A., de L. Tonnelier. Ces algorithmes de vérification sont définis et automatiquement générés lors de la compilation des transactions.

Période 1998 à ce jour : Professeur Université de Paris 11

Langages de programmation et systèmes pour données au format XML

L'explosion récente d'Internet comme support de calculs globaux conduit à l'échange croissant d'informations. La nature hautement distribuée du réseau induit que les informations sont généralement collectées, traitées puis restructurées à partir de sources multiples et hétérogènes. Afin d'en permettre le traitement ces informations sont publiées sous un même format. XML devient le standard de-facto pour l'échange et la manipulation de documents sur le Web. De nombreux facteurs plaident en faveur de son acceptation : les documents XML sont lisibles par un humain, ils sont par nature auto-descriptifs ; il fait l'objet d'un processus actif de standardisation au sein du W3C et de nombreuses entreprises, parmi les plus représentatives dans le domaine des technologies de l'information, l'ont dores et déjà adopté. Dans un tel contexte nous nous intéressons aux points suivants :

1. Définition de langages de programmation pour XML typés et expressifs
2. Systèmes de stockage natifs et optimisation de requêtes

Langages de programmation

Comme le reflète le processus actuel de standardisation, les documents XML sont intrinsèquement typés (notions de validité et de bonne formation relativement à une DTD, XML-*Schema*, Relax-NG). Or, les langages de transformation associés

(XPath/XSLT) sont essentiellement non-typés.

La définition de langages de transformations pour XML typés permettra d'obtenir une plus grande fiabilité du code et surtout d'offrir un contrôle (partiel) statique de correction de la transformation. Ainsi, dans un tel contexte, il sera possible de garantir que le résultat d'une transformation donnée est bien conforme au type (ou DTD, ou Schemas) attendu.

Enfin, dans le contexte de l'interrogation stricto-sensu de documents XML, il est indispensable d'offrir une interface plus déclarative qu'un langage turing-complet. Ce type d'interface est communément appelé langage de requête. Notre contribution des dernières années pour cet axe est brièvement développée.

Le Langage CDuce (www.cduce.org)

Les langages de transformation de documents XML actuels sont loin d'être complètement satisfaisants. Dans le cas de DOM (Document Object Model) —un API pour XML très répandu— par exemple, il n'y a aucune manière de garantir statiquement que le document produit par un programme est du type attendu. Ceci est aussi vrai pour XSLT par exemple, à l'exception près que dans ce dernier cas des programmes non corrects n'engendrent pas du XML mal formé mais des documents vides. D'une certaine manière on peut considérer ceux-ci comme des langages non typés.

Proposer un langage typé pour manipuler des documents XML est assez naturel : XML n'est qu'une syntaxe, et différents systèmes de types sont proposés : DTD, XML-Schema, ... ; un document ne prend de sens que si on le voit avec un certain type XML.

Dans le cadre du stage de DEA de A. Frisch que j'ai co-encadré avec G. Castagna nous avons étudié l'introduction du sous-typage sémantique ceci nous a permis d'établir le socle théorique sur lequel bâtir les développements futurs. Les résultats de ce travail sont présentés dans [23, 7] et donnent lieu au développement du langage CDuce [22, 21] qui est disponible sur le site www.cduce.org et est intégré aux principales distributions Linux (Ubuntu/Debian, Mandriva, Fedora).

Nous poursuivons ces recherches dans deux directions parallèles : (i) définition de langages de requêtes et d'interfaces graphiques pour les données au format XML et (ii) stockage et optimisation pour l'accès aux documents XML.

ⒸQL un langage de requête pour XML:

Le système de types de XDuce sert de base à celui de XQuery, une algèbre de requêtes pour XML proposée par le W3C. L'adoption d'un tel système montre l'intérêt que porte le monde des bases de données à celui des langages typés adaptés à XML. Cet intérêt ne se limite pas uniquement au monde académique. En effet, les types peuvent apporter plusieurs caractéristiques au monde des bases de données (comme ils l'ont toujours fait de par le passé).

Dans le cadre de la manipulation de documents XML, on peut considérer le mécanisme de filtrage des langages fonctionnels de la famille d'OCaml ("pattern-matching") comme une primitive pour exprimer et exécuter une *requête*. Dans le contexte du stage de Cédric Michon [40], puis de sa thèse [19], nous avons suivi cette approche en intégrant dans CDuce un langage de requête, ⒸQL, exploitant pleinement la puissance expressive des motifs ("patterns") ainsi que leur compilation efficace au moyen d'automates d'arbres non-uniformes. Il s'agit d'un langage déclaratif qui laisse toute latitude au compilateur pour exploiter au mieux des réécritures algébriques (en particulier de motifs) afin de générer des plans d'exécutions, en mémoire centrale efficaces.

Les perspectives de ce travail sont diverses: (i) étendre l'opération de filtrage afin de permettre également des parcours en profondeur de l'arbre (à la XPath), et (ii) offrir un mécanisme de persistance à ⒸQL ainsi que des mécanismes d'optimisation physique de requête associés.

Pattern by Example (PBE) : une interface graphique pour XML:

Il est commun que les utilisateurs d'un système de bases de données ne soient pas experts en informatique. Naturellement, le système offre une interface graphique pour permettre à de tels utilisateurs d'interroger et d'extraire de l'information. C'est pourquoi un des prolongements du travail effectué sur ⒸQL a été de définir un langage de requêtes graphique [13].

Cette interface s'inspire du langage graphique QBE (Query by Example) défini pour les systèmes relationnels et l'adapte au cas de XML. Dans un tel contexte, le fait de définir

un tel langage est crucial car non seulement la nature des données est beaucoup plus complexe mais en outre, les langages de manipulation sont basés sur le mécanisme de “pattern-matching” des langages fonctionnels qui sont difficiles à appréhender pour un utilisateur lambda. Ainsi, l'utilisateur interagit avec le système au travers de cette interface et la requête correspondante est automatiquement générée dans une forme optimisée. La génération est guidée par les types et la preuve formelle de la correction de la traduction est établie.

Données XML : aspects systèmes

Stockage natif et optimisation de requêtes

Les résultats existants en matière d'optimisation de requêtes sur des données (XML) persistantes peuvent être classés en fonction du modèle sous-jacent de stockage et de traitement de requêtes. Des travaux se sont appuyés sur des systèmes de gestion de bases de données relationnelles pour stocker des documents XML, et ont traité des requêtes XML en les traduisant en SQL. En conséquence, tous les aspects liés au placement des données, au stockage, et à l'optimisation de requêtes ont été délégués au système relationnel. Cette approche a ses limites, puisque des langages de requêtes XML plus récents (tels que XQuery et CDuce/CQL) sont bien plus complexes, et très différents de SQL.

Plus généralement, en collaboration avec Ioana Manolescu (INRIA - Futurs - Gemo), avec qui j'ai co-encadré la thèse d'A. Arion, nous proposons un optimiseur générique [17, 18, 15, 14] pouvant être paramétré en fonction des besoins et caractéristiques des applications visées.

Parallèlement, dans le cadre du travail de thèse de K. Nguyen et du post-doctorat de D. Colazzo, nous avons défini des techniques d'optimisation de requêtes (exprimées en XQuery) qui permettent grâce à une analyse statique de la requête et à la connaissance du type du document à interroger, d'élaguer ce dernier pour ne charger en mémoire centrale que la portion nécessaire à l'évaluation de la requête. Ce travail a été publié à VLDB [16] et une version étendue est actuellement soumise à la revue Journal of the ACM [52].

Workflow interactifs pour données scientifiques massives et distribuées

Les applications de gestion de données scientifiques sont, de nos jours, amenées à gérer d'énormes volumes de données dont la taille croît sans cesse. Ces données peuvent être de nature très diverse : numériques (résultats d'instruments de mesure), textuelles (corpus étudiés par des chercheurs en sciences humaines) structurées (en astronomie ou en physique) ou totalement irrégulières (dossiers médicaux).

Quelle que soit leur nature, ces données sont caractérisées par leur taille importante. Ceci résulte, d'une part, de l'utilisation de plus en plus répandue d'ordinateurs pour recueillir ces informations en lieu et place des cahiers utilisés auparavant par les chercheurs et, d'autre part, de l'utilisation d'outils puissants et automatisés pour intégrer de telles informations par exemple des télescopes, des navigateurs ciblés sur tel ou tel domaine ...

Ces données représentent une mine d'or pour les scientifiques qui peuvent fonder leurs recherches sur une telle base d'information. La disponibilité de ces données est détermi-

nante quant à leur exploitation par une communauté scientifique. Or, le plus souvent les scientifiques utilisent, pour stocker ou manipuler de telles sources : des formats propriétaires, des systèmes ad-hoc, des processus “maison”, ce, clairement au détriment de la disponibilité et donc du partage à grande échelle.

Par ailleurs, au sein de la communauté informatique, de nombreux efforts de recherche portent, indépendamment, sur la visualisation d’informations, la gestion de processus et la gestion de données massives. Ces trois courants nous semblent indispensables à l’obtention de bonnes plateformes de gestion de données scientifiques permettant d’aider les chercheurs dans leur travail quotidien. En effet, les propriétés suivantes caractérisent la nature des applications de gestion de données scientifiques : la complexité calculatoire et le volume massif des données à traiter, l’interactivité des processus mis en jeu et la visualisation indispensable des résultats. Par conséquent, nous identifions les trois thématiques suivantes :

1. Visualisation d’information et interaction,
2. Bases de données
3. Gestion de processus et systèmes de workflow

Ces trois thématiques sont complémentaires. Si les systèmes de visualisation d’informations permettent de présenter les données de façon claire, les algorithmes et outils qu’ils utilisent sont trop complexes pour un utilisateur naïf. Les systèmes de workflow interactifs, quant à eux, ne sont pas en mesure de faire face à la taille colossale des données scientifiques. D’autre part, les systèmes de gestion de bases de données manquent de flexibilité (les modifications de schéma sont difficiles), de puissance calculatoire (les langages de requêtes ne sont pas complets au sens de Turing) et enfin ils offrent, le plus souvent, des interfaces très rudimentaires. En revanche, par la nature déclarative de leurs langages de requêtes sous-jacents et par l’état de maturité atteint par leurs techniques d’optimisation, de gestion des accès concurrents et de stockage à grande échelle, ils possèdent de nombreux atouts. C’est ainsi qu’à la confluence des bases de données et du génie logiciel, nous trouvons le domaine de la gestion de (business) processus et de “workflows”. Or, ce cadre est trop ciblé sur des application de gestion de processus métiers et les systèmes développés par Oracle, SAP ou IBM ne sont pas adaptés à la gestion de données scientifiques. Ils sont complexes à utiliser et requièrent un administrateur spécifique, leur coût est également prohibitif. En outre, ils sont adaptés au traitement de données représentées sous forme tabulaire ce qui n’est pas le cas des données scientifiques qui ont en général une structure plus riche. Enfin ils ne permettent pas d’opérer des calculs puissants indispensables en revanche dans le domaine d’applications scientifiques.

Récemment, des plateformes de gestion de données scientifiques émergent outre-atlantique (Kepler/Ptolemy). Cependant elles ne prennent pas en compte la nature interactive des processus liés au traitement des données scientifique ni n’offrent des outils puissants de visualisation des résultats. Or les deux aspects sont indispensables à l’obtention de systèmes adaptés aux besoins des chercheurs.

Dans le cadre de la thèse de Wael Khemiri que nous co-encadrons, depuis septembre 2008, avec Ioana Manolescu, nous proposons de *définir des modèles, des algorithmes ainsi que des architectures* en vue de l’obtention d’un système en mesure d’apporter aux scientifiques les moyens de tirer le meilleur parti des données qu’ils recueillent en leur permettant de les

organiser à leur guise. Les premiers résultats de ce travail sont publiés dans [35], une version plus aboutie est publiée à ICDE 2011 [11].

Ce travail de recherche s'inscrit dans le cadre du projet Digitéo EDIFLOW pour lequel collaborent trois équipes : LÉO (I. Manolescu, INRIA Saclay) AVIZ (J-D Fekete, INRIA - Saclay) et l' équipe Bases de Données du LRI. Il met en jeu trois différents domaines desquels nous espérons prendre le meilleur.

Publications depuis le passage à la première classe (2004).

Depuis mon passage à la première classe, en 2004, j'ai **édité les actes de la conférence BDA**, j'ai publié : **un article invité**, un article (64 pages) à la **revue Journal of the ACM (A⁺)**, **neuf articles** dans des **conférences internationales** (1 ICDE (A), 1 Data Management and Visual Analytics (C) 3 VLDB (A⁺) 1 article à la conférence ACM - PPDP (A) un article à la conférence PADL : Practical Aspects of Declarative Languages (A), un article à la conférence Flexible Query Answering (B) et un article au workshop, XIME-P, satellite de ACM Sigmod - Pods (B)) et enfin **six articles** à la conférence nationale **BDA**. Un gros effort de **transfert** du langage CDuce a été mené ce qui a conduit à son **intégration** dans les principales **distributions du système Linux**. J'ai par ailleurs contribué à la rédaction du tutoriel du langage CDuce. Enfin je compte actuellement un article en soumission.

Un article invité ; édition d'actes BDA ; un article dans la revue JACM; 9 articles dans des conférences ou WS internationaux ; 6 articles à BDA ; développement de logiciel ; un article soumis

Chapitres de livres

- [1] V. Benzaken, G. Castagna, H. Hosoya, B.C. Pierce, and S. Vansummeren. *The Encyclopedia of Database Systems*, chapter "XML Typechecking". Springer, 2009. **Article invité**.

Édition d'ouvrages collectifs

- [2] Véronique Benzaken, editor. *21èmes Journées Bases de Données Avancées, BDA 2005, Saint Malo, 17-20 octobre 2005, Actes (Informal Proceedings)*, 2005.

Articles dans des revues internationales avec comité de lecture

- [3] Alain Frisch, Giuseppe Castagna and Véronique Benzaken. Semantic subtyping: Dealing set-theoretically with function, union, intersection, and negation types. *Journal of the ACM*, 55(4):1–64, 2008.

Communications dans des conférences internationales avec comité de lecture

- [4] Véronique Benzaken, Jean-Daniel Fekete, Pierre-Luc Hémerly, Wael Khemiri, and Ioana Manolescu. Ediflow: data-intensive interactive workflows for visual analytics. In *International Conference on Data Engineering (ICDE)*, 2011. Etend la version [34].
- [5] Ioana Manolescu, Wael Khemiri, Veronique Benzaken, and Jean-Daniel Fekete. Reaviz: Reactive workflows for visual analytics. In *Data Management & Visual Analytics workshop*, 2009.

- [6] V. Benzaken, G.Castagna, D. Colazzo, and C. Miachon. Pattern by example: Type-driven visual programming of XML queries". In *10th International ACM SIGPLAN Symposium on Principles and Practice of Declarative Programming*, Valencia, Spain, July 2008. ACM. Etend la version [36]
- [7] A. Arion, V. Benzaken, I. Manolescu, and Y. Papakonstantinu. Structured materialized views for XML queries. In *33rd International Conference on Very Large Databases (VLDB 2007)*, September 2007.
- [8] A. Arion, V. Benzaken, I. Manolescu, and Y. Papakonstantinu. Algebra-based identification of tree patterns in xquery. In Henrik Legind Larsen, Gabriella Pasi, Daniel Ortiz-Arroyo, Troels Andreasen, and Henning Christiansen, editors, *Flexible Query Answering Systems: 7th International Conference, FQAS 2006*, volume 4027 of *LNCS*, pages 13 – 25, Milano, June 2006. Springer Verlag.
- [9] V. Benzaken, G.Castagna, D. Colazzo, and K. Nguyen. Type-based XML projection. In *32nd International Conference on Very Large Databases (VLDB 2006)*, September 2006. Etend la version[37].
- [10] Véronique Benzaken, Giuseppe Castagna, and Cédric Miachon. A full pattern-based paradigm for XML query processing. In *PADL, International Symposium on Practical Aspects of Declarative Languages*, pages 235–252, 2005.
- [11] Andréi Arion, Véronique Benzaken, Ioana Manolescu, and Ravi Vijay. Uload: Choosing right storage for your XML application. In *Demonstration at the 31st International Conference on Very Large Databases*, 2005.
- [12] Andréi Arion, Véronique Benzaken, and Ioana Manolescu. XML access modules: Towards physical data independence in XML databases. In *Proceedings of the Second International Workshop on XQuery Implementation, Experience and Perspective colocated with ACM Sigmod Pods*, 2005.

**Communications dans des conférences nationales
avec comité de lecture**

- [13] V. Benzaken, J-D Fekete, W. Khemiri, and I. Manolescu. EDIFLOW: Data-intensive interactive workflows for visual analytics. In *Bases de données avancées – BDA*, Toulouse, 2010.
- [14] Ioana Manolescu, Wael Khemiri, Véronique Benzaken, and Jean-Daniel Fekete. Reactive workflows for visual analytics. In *Bases de données avancées – BDA*, Namur, 2009.
- [15] Véronique Benzaken, Giuseppe Castagna, Dario Colazzo, and Cédric Miachon. Pattern by example: type-driven visual programming of XML queries. In *BDA, 2007*
- [16] I. Manolescu, V. Benzaken, A. Arion, and Y. Papakonstantinou. Structured materialized views for XML queries. In D. Laurent, editor, *BDA 2006*, Lille, October 2006.
- [17] Véronique Benzaken, Giuseppe Castagna, Dario Colazzo, and Kim Nguyen. Type-based projectors. In *BDA, 2006*.
- [18] V. Benzaken, G. Castagna, and C. Miachon. CQL a pattern-based query language for XML. In *In Proceedings 20th Bases de Données Avancées (BDA'04)*, 2004.

Notices descriptives, manuels d'initiation ou de référence de logiciels ou langages

- [19] V. Benzaken, G. Castagna, and A. Frisch. *CDuce Tutorial*, 2002-2010. Available on line at <http://www.cduce.org/tutorial.html>.

Développement de logiciels, brevets

- [20] *Langage CDuce*, 2008. **Intégré dans les distributions Linux** (Debian, Ubuntu, Fedora, Mandriva).

Articles soumis ou en préparation

- [21] V. Benzaken, G. Castagna, D. Colazzo, and K. Nguyen. Optimizing XML querying using type-based document projection (60 pages). Under submission (JACM).

Liste complète des publications.

Livres

- [1] V. Benzaken and A. Doucet. *Bases de données orientées objet : origines et principes*. Armand-Colin, 1993.

Chapitres de livres

- [2] V. Benzaken, G. Castagna, H. Hosoya, B.C. Pierce, and S. Vansummeren. *The Encyclopedia of Database Systems*, chapter “XML Typechecking”. Springer, 2009. **Article invité.**
- [3] V. Benzaken, C. Delobel, and G. Harrus. Clustering strategies in o_2 : an overview. In F. Bancilhon, C. Delobel, and P. Kanellakis, editors, *Building an Object-Oriented Database System: the Story of O_2* , pages 385–410. Morgan Kaufman, 1992.

Édition d'ouvrages collectifs

- [4] Véronique Benzaken, editor. *21èmes Journées Bases de Données Avancées, BDA 2005, Saint Malo, 17-20 octobre 2005, Actes (Informal Proceedings)*, 2005.
- [5] M. Adiba and V. Benzaken, editors. *Numéro spécial Bases de données*, volume 17 of *TSI*. Hermès, 1998.
- [6] M. Atkinson, V. Benzaken, and D. Maier, editors. *Persistent Object Systems 6*. Workshops in Computing. Springer-Verlag, Tarascon, France, 5-9 September 1994.

Articles dans des revues internationales avec comité de lecture

- [7] Alain Frisch, Giuseppe Castagna, and Véronique Benzaken. Semantic subtyping: Dealing set-theoretically with function, union, intersection, and negation types. *Journal of the ACM*, 55(4):1–64, 2008.
- [8] V. Benzaken, S. Cerrito, and S. Praud. Static verification of dynamic constraints: a semantics-based approach. *Network and Information Systems Journal (NIS)*, 2(5-6):549–571, 1999.
- [9] V. Benzaken and A. Doucet. Thémis: A Database Programming Language Handling Integrity Constraints. *Very Large Databases (VLDB) Journal*, 4(3):493–518, 1995.

1 Livre; 2 chapitres de livre dont 1 article invité ; 3 éditions d'ouvrages ; 3 revues internationales ; 1 revue nationale ; 23 articles dans des conférences internationales ; 14 BDA ; 1 développement de logiciel

**Articles dans des revues nationales
avec comité de lecture**

- [10] V. Benzaken, A. Doucet, and X. Schaefer. Integrity Constraint Checking Optimization based on Abstract Databases Generation and Program Analysis. *Journal de l'Ingénierie des Systèmes d'Information*, 1(3):9–29, 1995.

Communications dans des conférences internationales avec comité de lecture

- [11] Véronique Benzaken, Jean-Daniel Fekete, Pierre-Luc Hémary, Wael Khemiri, and Ioana Manolescu. Ediflow: data-intensive interactive workflows for visual analytics. In *International Conference on Data Engineering (ICDE)*, 2011. Etend la version [34].
- [12] Ioana Manolescu, Wael Khemiri, Veronique Benzaken, and Jean-Daniel Fekete. Reaviz: Reactive workflows for visual analytics. In *Data Management & Visual Analytics workshop*, 2009.
- [13] V. Benzaken, G.Castagna, D. Colazzo, and C. Miachon. Pattern by example: Type-driven visual programming of XML queries". In *10th International ACM SIGPLAN Symposium on Principles and Practice of Declarative Programming*, Valencia, Spain, July 2008. ACM. Etend la version [36].
- [14] A. Arion, V. Benzaken, I. Manolescu, and Y. Papakonstantinu. Structured materialized views for XML queries. In *33rd International Conference on Very Large Databases (VLDB 2007)*, September 2007.
- [15] A. Arion, V. Benzaken, I. Manolescu, and Y. Papakonstantinu. Algebra-based identification of tree patterns in xquery. In Henrik Legind Larsen, Gabriella Pasi, Daniel Ortiz-Arroyo, Troels Andreasen, and Henning Christiansen, editors, *Flexible Query Answering Systems: 7th International Conference, FQAS 2006*, volume 4027 of *LNCIS*, pages 13 – 25, Milano, June 2006. Springer Verlag.
- [16] V. Benzaken, G.Castagna, D. Colazzo, and K. Nguyen. Type-based XML projection. In *32nd International Conference on Very Large Databases (VLDB 2006)*, September 2006. Etend la version[37].
- [17] Andréï Arion, Véronique Benzaken, and Ioana Manolescu. XML access modules: Towards physical data independence in XML databases. In *Proceedings of the Second International Workshop on XQuery Implementation, Experience and Perspective colocated with ACM Sigmod Pods 2005*, 2005.
- [18] Andréï Arion, Véronique Benzaken, Ioana Manolescu, and Ravi Vijay. Uload: Choosing right storage for your XML application. In *Demonstration at the 31st International Conference on Very Large Databases*, 2005.
- [19] Véronique Benzaken, Giuseppe Castagna, and Cédric Miachon. A full pattern-based paradigm for XML query processing. In *PADL, International Symposium on Practical Aspects of Declarative Languages*, pages 235–252, 2005.
- [20] V. Benzaken, M. Burelle, and G. Castagna. Information flow security for XML transformations. In V. Saraswat, editor, *Eight Asian Computing Science Conference (ASIAN'03)*, Lecture Notes in Computer Science, Mumbai, India, 10-13 December 2003. Springer-Verlag.

-
- [21] V. Benzaken, G. Castagna, and A. Frisch. CDuce: an XML-centric general-purpose language. In *ACM International Conference on Functional Programming (ICFP'03)*, 2003.
 - [22] V. Benzaken, G. Castagna, and A. Frisch. CDuce: a white paper. In B. C. Pierce and P. Wadler, editors, *International workshop on Programming Languages for XML (PLANX'02)*, Pittsburgh, USA, 2002.
 - [23] A. Frisch, G. Castagna, and V. Benzaken. Semantic subtyping. In *Seventeenth Annual IEEE Symposium on Logic in Computer Science (LICS'2002)*, pages 137–146, Copenhagen, Denmark, 2002.
 - [24] F. Wattez, S. Cluet, V. Benzaken, and C. Fiegel. Benchmarking queries over trees: Learning the hard truth the hard way (short paper). In *ACM SIGMOD International Conference*, Dallas, May 2000. ACM.
 - [25] V. Benzaken, N. Hugon, H. Klaudel, E. Pelz, and R. Rieman. A mnet calculus-based semantics for triggers. In Springer Verlag, editor, *International Conference on Theory and Applications of Petri Nets*, LNCS, Lisbon, Portugal, June 1998.
 - [26] V. Benzaken and X. Schaefer. Static management of integrity in object-oriented databases: Design and implementation. In Hans-Jörg Schek, editor, *6th International Conference on Extending Database Technology (EDBT'98)*, Valencia, Spain, March 23–27 1998. Springer Verlag.
 - [27] V. Benzaken and X. Schaefer. Static integrity constraint management in object-oriented database programming languages via predicate transformers. In Aksit and Matsuoka, editors, *European Conference on Object-Oriented Programming (ECOOP'97)*, number 1024 in Lecture Notes in Computer Science, pages 60–84. Springer-Verlag, 1997.
 - [28] V. Benzaken and X. Schaefer. Ensuring efficiently the integrity of persistent object systems via abstract interpretation. In S. Nettles and R. Connor, editors, *Proceedings of Persistent Object Systems 7*, pages 72–87, Cape May, U.S.A., June 1996. Morgan Kaufman.
 - [29] V. Benzaken and A. Doucet. Thémis: a database programming language with integrity constraints. In C. Beeri, A. Ohori, and D. Shasha, editors, *Proceedings of Database Programming Languages 4*, Workshop in Computing, pages 243–262, New York City, USA, September 1993. Springer-Verlag.
 - [30] V. Benzaken, C. Lécluse, and P. Richard. Enforcing Integrity Constraints in Database Programming Languages. In R. Morrison A. Albano, editor, *Proceedings of Persistent Object Systems 5*, Workshop in Computing, pages 282–299, San Miniato, Italy, September 1–5 1992. Springer-Verlag.
 - [31] V. Benzaken. An Evaluation Model for Clustering Strategies in the O₂ Object-Oriented Database System. In S. Abiteboul and P. Kanellakis, editors, *International Conference on Database Theory (ICDT)*, number 470 in Lecture Note in Computer Science, pages 126–140, Paris, France, December 1990. Springer Verlag.
 - [32] V. Benzaken and C. Delobel. Enhancing Performance in a Persistent Object Store: Clustering Strategies in O₂. In *Proceedings of Persistent Object Systems 4*. Morgan Kaufmann, September 1990.

- [33] F. Bancilhon, G. Barbedette, V. Benzaken, C. Delobel, S. Gamerman, C. Lécluse, P. Pfeffer, P. Richard, and F. Velez. The design and implementation of o₂, an object-oriented database system. In *Advances in Object-Oriented Database Systems*, Lecture Notes in Computer Science. Springer-Verlag, September 1988.

**Communications dans des conférences nationales
avec comité de lecture**

- [34] V. Benzaken, J-D Fekete, W. Khemiri, and I. Manolescu. Data-intensive interactive workflows for visual analytics. In *Bases de données avancées – BDA*, Toulouse, 2010.
- [35] Ioana Manolescu, Wael Khemiri, Véronique Benzaken, and Jean-Daniel Fekete. Reactive workflows for visual analytics. In *Bases de données avancées – BDA*, Namur, 2009.
- [36] Véronique Benzaken, Giuseppe Castagna, Dario Colazzo, and Cédric Miachon. Pattern by example: type-driven visual programming of XML queries. In *BDA*, 2007.
- [37] V. Benzaken, G.Castagna, D. Colazzo, and K. Nguyen. Type-based XML projection. In D. Laurent, editor, *BDA 2006*, October 2006.
- [38] Véronique Benzaken, Giuseppe Castagna, Dario Colazzo, and Kim Nguyen. Type-based projectors. In *BDA*, 2006.
- [39] I. Manolescu, V. Benzaken, A. Arion, and Y. Papakonstantinou. Structured materialized views for XML queries. In D. Laurent, editor, *BDA 2006*, Lille, October 2006.
- [40] V. Benzaken, G. Castagna, and C. Miachon. CQL a pattern-based query language for XML. In *In Proceedings 20th Bases de Données Avancées (BDA'04)*, 2004.
- [41] V. Benzaken, S. Cerrito, and S. Praud. Static verification of dynamic constraints: a semantics-based approach. In *Actes de la 15ième Conférence sur les Bases de Données Avancées*, Bordeaux, France, 1999.
- [42] V. Benzaken and X. Schaefer. Ensuring efficiently the integrity of persistent object systems via abstract interpretation. In *Actes de la 12 ième Conférence sur les Bases de Données Avancées*, Cassis, August 1996. La version étendue se trouve dans [28].
- [43] V. Benzaken, A. Doucet, and X. Schaefer. Integrity constraint checking optimization based on abstract interpretation. In *Actes de la 10ième Conférence sur les Bases de Données Avancées*, Clermont-Ferrand, France, August 1994.
- [44] V. Benzaken, A. Doucet, and P.Y. Policella. Définition et gestion de contraintes d'intégrité dans le langage Thémis. In *Actes de la 9ième Conférence sur les Bases de Données Avancées*, Toulouse, France, October 1993.
- [45] V. Benzaken and A. Doucet. Enforcement Tests Generation for Integrity Constraints Checking Based on Simplification Methods in Object-Oriented Database Systems. In *Actes de la 8ième Conférence sur les Bases de Données Avancées*, Trégastel, France, September 15-18 1992.
- [46] V. Benzaken. Un modèle d'évaluation de stratégies de regroupement dans un système de bases de données orienté-objet. In *Actes de la 6ième Conférence sur les Bases de Données Avancées*, Montpellier, France, September 1990.

-
- [47] V. Benzaken and C. Delobel. Regroupement d'objets sur disque dans un système de bases de données orienté-objet. In *Actes de la 5ième Conférence sur les Bases de Données Avancées*, Genève, Suisse, September 1989.
- [48] V. Benzaken, C. Delobel, and J.B. N'dala. Gestionnaires de mémoires et d'objets. In *Actes de la 4ième Conférence sur les Bases de Données Avancées*, Bénodet, France, June 1988.

Thèses

- [49] V. Benzaken. Langages de programmation pour bases de données et contraintes d'intégrité. Habilitation à diriger des recherches, 1996.
- [50] V. Benzaken. Regroupement d'objets sur disque dans un système de bases de données orienté-objet. PhD thesis, Université de Paris-Sud, 1990.

Notices descriptives, manuels d'initiation ou de référence de logiciels ou langages

- [51] V. Benzaken, G. Castagna, and A. Frisch. *CDuce Tutorial*, 2002-2010. Available online at <http://www.cduce.org/tutorial.html>.

Articles soumis ou en préparation

- [52] V. Benzaken, G. Castagna, D. Colazzo, and K. Nguyen. Optimizing XML querying using type-based document projection. Under submission.

Développement de logiciels, brevets

- [53] *Langage CDuce*, 2003. **Intégré dans les distributions Linux** (Debian, Ubuntu, Fedora, Mandriva).