

Objectifs

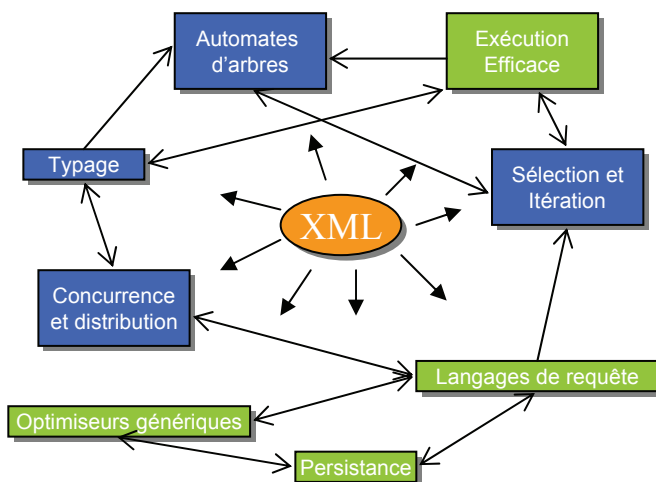
Notre projet se propose d'étudier les aspects de traitement, d'interrogation et de manipulation de grandes masses de données lorsque celles-ci sont disponibles au format XML. Nous nous intéressons plus précisément aux aspects langages de programmation et langages de requêtes.

Notre ambition est de couvrir de manière intégrée un large spectre de problématiques: de celles liées aux **aspects langages** (expressivité, typage, étude de nouvelles primitives de programmation, logiques sous-jacentes pour l'interrogation, optimisation logique), jusqu'aux aspects traitant l'**accessibilité des données** (données en *streaming*, compression, définition de modèles d'accès en mémoire secondaire, utilisation de moteurs de persistance), en passant par les problématiques liées à l'**implantation** (compilation du filtrage, optimisation physique, vérification du sous-typage, modèles d'exécution pour le *streaming*).

Partenariat

- Le Laboratoire d'Informatique de l'Ecole Normale Supérieure et le Laboratoire de Recherche Informatique de l'Université Paris Sud ont développé conjointement le langage pour documents XML CDuce: www.cduce.org.
- L'équipe Mostrare (LIFL- INRIA Futurs Lille) amène les compétences en automates d'arbres. En lien avec les langages de requêtes.
- Le projet GEMO de l'INRIA Futurs à Saclay est à la pointe de la recherche en bases données semi-structurées et en optimisation. Avec le LIAFA (Paris 7) apporte aussi des compétences en *streaming*, automates et complexité.
- Le LIF (Marseille) est spécialisé en théorie des langages: en particulier pour ce qui concerne les aspects typage et concurrence.

Dépendances



Principes de base

Trois Axes Directeurs :

- Langages de requêtes
- Traitement à la volée
- Contraintes et typage de documents

Bases Théoriques :

- Automates d'arbres, logique.
- Expressivité, complexité des langages de requêtes
- Identification de classes de requêtes évaluable en *streaming*.
- Utilisation des types statiques pour une compilation efficace et l'optimisation de requêtes.
- Théorie de la Concurrency et de la distribution.

Implantations

- CDuce
Langage de programmation fonctionnel, fortement typé, à base de patterns, orienté XML. www.cduce.org
- XML access modules
Description du stockage de données XML, optimisation des requêtes en fonction du stockage.
- REDuce, ASTuce, ...
Langages expérimentaux basés sur CDuce et sur les développements théoriques.

Quelques Références

- *A Gentle Introduction to Semantic Subtyping*. G. Castagna and A. Frisch. ICALP-PPDP '05 joint invited talk.
- *ULoad: Choosing the Right Storage for your XML Application*. A. Arion, V. Benzaken, I. Manolescu, and R. Vijay. VLDB 2005.
- *Patterns and Types for Querying XML documents*. G. Castagna. DBPL and XSym '05 joint invited talk.
- *Semantic Subtyping for the Pi-Calculus*. G. Castagna, R. De Nicola, and D. Varacca. LICS 2005
- *Expressiveness of a Spatial Logic for Trees*. I. Boneva, J.M. Talbot, and S. Tison. LICS 2005
- *Towards a Characterization of Order-Invariant Queries over Tame Structures*. M. Benedikt, L. Ségoufin. CSL 2005
- *Indexes and Path Constraints in Semistructured Data*. Y. Andre, A.C. Caron, D. Debarbieux, and Y. Roos. Workshop on Logical Aspects and Applications of Integrity Constraints. 2005
- *Counting in Trees for Free*. H. Seidl, T. Schwentick, A. Muscholl and P. Habermehl. ICALP 2004
- *N-ary Queries by Tree Automata*. J. Niehren, L. Planque, J.M. Talbot, and S. Tison. DBPL 2005
- *A Logic you Can Count On*. S. Dal Zilio, D. Lugiez and C. Meyssonier. POPL 2004
- *A First Step Towards Modelling Semistructured Data in Hybrid Multi-modal Logic*. N. Bidoit, S. Cerrito, V. Thion. JANCL 2004.

**Contact projet : CASTAGNA Giuseppe, École Normale Supérieure, 45 rue d'Ulm, Paris. Giuseppe.Castagna@ens.fr
<http://www.cduce.org/tralala.html>**