

Application BASTRI

Fiches Equipes

VALDA (SR0824OR)

Valeur à partir des données
VALDA (SR0765JR) □ VALDA

Statut: Décision signée

Responsable : Pierre Senellart

Mots-clés de "A - Thèmes de recherche en Sciences du numérique - 2024" : A3.1. Données, A3.1.1. Modélisation, représentation, A3.1.2. Gestion, interrogation et stockage, A3.1.3. Données distribuées, A3.1.4. Données incertaines, A3.1.5. Contrôle d'accès, confidentialité, A3.1.6. Optimisation de requêtes, A3.1.7. Données ouvertes, A3.1.8. Données massives (production, stockage, acheminement), A3.1.9. Bases de données, A3.1.10. Données hétérogènes, A3.1.11. Données structurées, A3.2. Connaissances, A3.2.1. Bases de connaissances, A3.2.2. Extraction de connaissances, nettoyage, A3.2.3. Inférence, A3.2.4. Web sémantique, A3.2.5. Ontologies, A3.2.6. Données liées, A3.3. Analyse de données et de connaissances, A3.3.1. Analyse en ligne, A3.3.2. Fouille de données, A3.3.3. Analyse de données massives, A3.4.3. Apprentissage par renforcement, A3.4.5. Méthodes bayésiennes, A3.5.1. Analyse de grands graphes, A4.7. Contrôle d'accès, A7.2. Logique, A7.3. Calculabilité et computabilité, A9.1. Connaissances, A9.8. Raisonnement

Mots-clés de "B - Autres sciences et domaines d'application - 2024" :

B2. Santé, B3.3. Géosciences, B4. Energie, B4.2. Production d'énergie nucléaire, B9.3. Médias, B9.5.6. Science des données, B9.6.5. Sociologie, B9.6.10. Humanités numériques, B9.7.2. Données ouvertes, B9.9. Ethique, B9.10. Confidentialité, vie privée

Domaine : Perception, Cognition, Interaction

Thème : Représentation et traitement des données et des connaissances

Période : 01/01/2018 -> 30/06/2025

Dates d'évaluation : 03/10/2019 ,

Etablissement(s) de rattachement : ENS PSL, CNRS

Laboratoire(s) partenaire(s) : DI-ENS (UMR8548)

CRI : Centre Inria de Paris

Localisation : Ecole Normale supérieure Paris

Code structure Inria : 021144-1

Numéro RNSR : 201622223R

N° de structure Inria:SR0824OR

Présentation

Les recherches de Valda portent sur les *aspects fondamentaux et systèmes de la gestion de données* complexes, en particulier les données produites par l'activité humaine. Les données qui nous intéressent sont généralement hétérogènes, massivement distribuées, d'évolution rapide, coûteuses d'accès et souvent subjectives, peut-être erronées, imprécises, incomplètes. Dans ce cadre, Valda s'intéresse à l'optimisation de ressources complexes telles que le temps et l'espace de calcul, ou des budgets de communications, financiers, ou de privauté. Le but est d'extraire de la *valeur à partir des données*, au-delà d'une simple réponse à des requêtes.

Axes de recherche

- Fondements de la gestion des données.** Cet axe couvre la théorie de la gestion des données au sens large, et en particulier les domaines de la *théorie des bases de données*, de la *représentation des connaissances* et de certains aspects *symboliques* de l'intelligence artificielle (en particulier, le *raisonnement sur les données*). L'objectif est de définir des fondements solides et de haut niveau des tâches de gestion des données (évaluation des requêtes et optimisation de diverses formes de requêtes, comptage, raisonnement, vérification des processus centrés sur les données, etc.) au moyen d'outils formels, tels que la logique (en particulier, la théorie des modèles finis), la théorie des automates, la théorie de la complexité ; nous avons également occasionnellement des contributions dans ces domaines, bien que la plupart de nos travaux soient motivés par des applications aux données. Nous nous intéressons particulièrement à obtenir des

Contact

- **Responsable :** Pierre Senellart
- **Tél :** + 33. 1. 4.4 .32. 2.0 .61
- **Secrétariat Tél :** + 33. 1. 8.0 .49. 4.0 .33

En savoir plus

- Site de l'[équipe](#)
- Site sur [inria.fr](#)
- Site du [responsable](#)
- Derniers Rapports d'Activité : [2017](#) , [2018](#) , [2019](#) , [2020](#) , [2021](#) , [2022](#) , [2023](#)

Documents sur la structure

- [Intranet](#)
- [Privés](#)

Décisions

- [12569](#) (18/12/2017) : création
- [14331](#) (30/07/2020) : prolongation
- [16994](#) (26/04/2024) : prolongation
- [17478](#) (27/11/2024) : prolongation

Localisation

- **Adresse postale :** École Normale supérieure 45 rue d'Ulm 75005 Paris France
- **Coordonnées GPS :** 48.841898, 2.345021

spécifications propres des aspects clés des systèmes de bases de données et des tâches de gestion des données (par exemple, la confidentialité, le contrôle d'accès, la robustesse), qu'il s'agisse de propriétés des données ou de langages (de requête) appropriés pour ces tâches. Nous étudions le pouvoir expressif des langages, la calculabilité et la complexité des résultats de décision ou de calcul, ainsi que la conception de structures appropriées (par exemple, des index) pour optimiser ces tâches.

2. **Incertitude, provenance et explicabilité en gestion de données.**

Cet axe de recherche porte sur la modélisation et la gestion efficace des données qui comportent de l'incertitude (distributions de probabilité, incomplétude logique, valeurs manquantes, incohérences, hypothèse de monde ouvert, etc.) et des informations de provenance (indiquant l'origine des données), ainsi que sur l'extraction d'annotations d'incertitude et de provenance à partir de données du monde réel. La provenance est également liée à l'explicabilité : déterminer d'où vient le résultat d'une tâche de gestion de données, comment et pourquoi il a été produit, aide à l'expliquer. Il est intéressant de noter que les fondements et les outils de gestion de l'incertitude reposent souvent sur des annotations de provenance. Par exemple, une manière typique de calculer la probabilité des résultats de requête dans les bases de données probabilistes est l'approche dite *intentionnelle* : générer d'abord la provenance de ces résultats de requête (dans un cadre approprié, par exemple celui des fonctions booléennes ou des semi-anneaux de provenance), puis calculer la probabilité de l'annotation de provenance résultante. Pour cette raison, nous traitons l'incertitude et la provenance de manière unifiée, et l'explicabilité comme une application de celles-ci.

3. **Découverte de connaissances à grande échelle.**

Notre dernier axe porte sur la découverte de connaissances à grande échelle. L'objectif est d'utiliser des techniques telles que l'exploration de données, l'extraction d'informations, le nettoyage de données, l'intégration d'informations et l'apprentissage automatique pour dériver des connaissances à partir de données brutes, sales, incohérentes, hétérogènes et en évolution rapide, issues de scénarios d'application réels. Nous avons l'intention de tirer parti de notre expertise en matière de gestion de données pour nous concentrer sur le passage à l'échelle des approches et outils développés. Il s'agit également, d'une certaine manière, d'un axe applicatif pour les techniques développées dans les deux autres axes ; en particulier, nous nous concentrons sur l'intensionnalité des données (c'est-à-dire le coût d'accès aux données), sur le compromis entre l'incertitude des données et son coût, sur la provenance et les explications des données.

Relations industrielles et internationales