



Master Informatique, Mathématiques, Multimédia & Télécommunications
Spécialité « informatique »
Spécialité « Réseaux de Télécommunications, Multimédia et Automatique »

Proposition de sujet de Stage Recherche 2017-2018

Titre : Exploitation de Dark Test Data

Laboratoire d'Informatique et d'Automatique pour les Systèmes (LIAS/ISAE-ENSMA)

Encadrant(s) :

- **Ladjel BELLATRECHE** (bellatreche@ensma.fr, 05 49 49 80 77)
- **Yassine OUHAMMOU** (ouhammoy@ensma.fr)
- **Selma KHOURI** (selma.khour@ensma.fr)

Mots clés : Besoins non fonctionnels, Données de Test, Benchmark

Description du sujet

Toutes les communautés et les organismes de recherche et d'industrie consacrent beaucoup d'efforts, pour tester leurs produits afin d'évaluer leurs qualités. Au-delà de la performance fonctionnelle, cette évaluation se base essentiellement sur l'utilisation des métriques associées à des besoins non fonctionnels (par exemple le temps de réponse, la consommation d'énergie, la qualité de service, etc.). Dans les systèmes complexes tels que les systèmes de gestion de bases de données, la maîtrise de complexité nécessite des tests d'analyse allant de la conception à la vérification et la validation.

En ce qui concerne notre domaine d'étude, les bases de données, nous constatons que chaque produit est associé à des tests, chacun étant matérialisé par un ensemble de résultats. En conséquence, une mine de tests existe couvrant toutes les phases de cycle de vie de conception : l'identification des besoins, la modélisation conceptuelle, la modélisation logique, la phase de déploiement, la modélisation physique, et la phase d'exploitation. En analysant de près les résultats de tests, nous remarquons qu'ils incluent plusieurs paramètres associés à l'environnement de tests couvrant des dimensions comme : les données utilisées pour les tests, les algorithmes utilisés, la plateforme d'exécution, les besoins non fonctionnels, les requêtes, les hypothèses, les structures d'optimisation utilisées, etc. L'évolution de la technologie de bases de données, des plateformes de déploiement (matérielles et logicielles) contribue à l'explosion exponentielle des informations autour de tests. De plus, les nouveaux produits dérivés de cette évolution doivent également être testés et validés. Face à cette situation, nous nous sommes interrogés sur la localisation digitale de cette mine d'information afin d'être exploitée par des tiers. Les tests et leurs environnements sont généralement localisés dans les articles scientifiques publiés par les chercheurs ou dans des sites Web spécialisés comme le TPC (Transaction Processing Council). Cette mine d'information est partiellement exploitée à des fins de comparaison des résultats lorsqu'un nouveau produit est lancé. Vu leurs caractéristiques, les données de tests représentent un bon exemple de données obscures (Dark Data). Ces dernières sont l'ensemble d'informations que les organisations collectent, traitent et stockent pendant leurs activités régulières, mais qu'elles n'arrivent pas à les réutiliser. La proposition d'une démarche d'explicitation et de persistance de toutes les dimensions ainsi que les résultats de tests dans le contexte des bases de données contribue à les sortir de l'obscurité. Une fois

SP2MI, Bvd Marie et Pierre Curie, Téléport 2, B.P. 30179, 86960 FUTUROSCOPE
Téléphone : 05 49 49 74 71, Télécopie : 05 49 49 65 70
E-mail : Sylyvie.Duclaud@sic.sp2mi.univ-poitiers.fr

Master Informatique, Mathématiques, Multimédia & Télécommunications
Spécialité « informatique »
Spécialité « Réseaux de Télécommunications, Multimédia et Automatique »

persistées, ces données accompagnées de leurs environnements peuvent être exploitées pour d'autres usages tels que : la reproductibilité des résultats, la recommandation, l'analyse OLAP, l'analyse multicritère, la détection des fraudes pour fournir plus de transparence aux tests et leur réutilisation.

Objectif du stage

L'objectif de ce stage est de construire un entrepôt dédiée aux données de tests. Une fois conçu, des techniques issues de *machine learning* l'exploiteront pour des fins de recommandation et personnalisation.

Technologies impliquées

- **Entrepôt de Données NoSQL** : MongoDB
- **Ingénierie Dirigée par les Modèles**
- **Deep Dive**
- **Tensor-flow**
- **Langages de programmation** : Java
- **Architecture logicielle** : web service, microservice et broker de message (RabbitMQ)
- **Framework web** : Vue.js (pour fournir un démonstrateur graphique)

Références Bibliographiques

- Lahcene Brahimi, Yassine Ouhammou, Ladjel Bellatreche, Abdelkader Ouared, More Transparency in Testing Results: Towards an Open Collective Knowledge Base, 10th IEEE International Conference on Research Challenges in Information Science (RCIS 2016), edited by IEEE, 2016, pp. 315-320.
- Lahcene Brahimi, Ladjel Bellatreche, Yassine Ouhammou, A Recommender System for DBMS Selection Based on a Test Data Repository, 20th East-European Conference on Advances in Databases and Information Systems (ADBIS 2016), LNCS, edited by Springer, 2016, pp. 15-20.
- Lahcene Brahimi, Ladjel Bellatreche, Yassine Ouhammou, Coupling Multi-Criteria Decision Making and Ontologies for Recommending DBMS, 22nd International Conference on Management of Data (COMAD 2017).
- Abdelkader Ouared, Yassine Ouhammou, Ladjel Bellatreche: MetricStore repository: on the leveraging of performance metrics in databases. ACM SAC 2017 : 1820-1825
- Abdelkader Ouared, Yassine Ouhammou, Ladjel Bellatreche: CostDL: A Cost Models Description Language for Performance Metrics in Database. ICECCS 2016: 187-190
- Abdelkader Ouared, Yassine Ouhammou, Amine Roukh: A Meta-advisor Repository for Database Physical Design. MEDI 2016 : 72-87