

Traitement de données multi-sources hétérogènes dans un environnement d'objets connectés : application à la gestion interactive du réseau du transport de la société TICE.

Heterogeneous multi-source data processing in a connected objects environment: application to the interactive management of the transport network of TICE.

Ecole Doctorale : Sciences et Technologies de l'Information et de la communication

Pôle P1 : Automatique, Traitement du Signal, Traitement des Images, Robotique

Spécialité : Informatique

Unité de recherche : IBISC - Informatique, Biologie Intégrative, & Systèmes Complexes, EA 4526, Université d'Evry-Val-d'Essonne, 91020 Evry

Directeur de thèse : M. Khalifa DJEMAL - Tel : 01 69 47 75 54

Co-directeur : M. Samir OTMANE - Tel : 01 69 47 75 92

Entreprise : TICE

Co-encadrant : Karine Hallouin (TICE)

Partenaires :

Résumé :

L'objectif de cette thèse est de concevoir et réaliser un système d'interprétation et de restitution des données issue de sources hétérogènes permettant une gestion optimale et interactive des informations d'objets connectés du réseau du transport de la société TICE. Des outils d'extraction d'informations des différentes sources pour construire un entrepôt de données seront formalisés et évalués. Une approche d'apprentissage statistique sera conçue et mise en place pour l'interprétation et la restitution interactive des données.

Mots clés : Extraction de données, Entrepôt de données, Apprentissage statistique, Classification, Interprétation et restitution de données, Objets connectés,

Abstract:

The aim of this PhD thesis is to design and implement a data interpretation and restitution system coming from heterogeneous sources, which allowing an optimal and interactive management of the information from connected objects of the transport network of the company TICE. Tools for extracting information from different sources to build a data warehouse will be formalized and evaluated. A statistical learning approach will be designed and implemented for the interpretation and interactive rendering of data.

Keywords: Data extraction, data warehouse, Machine learning, Classification, Data interpretation, connected objects.

Contexte :

L'entreprise TICE gère en moyenne 80000 voyageurs par jour, ce qui engendre un volume de données très important. Face à cette masse de données dont l'évolution est exponentielle, la gestion des données nécessite des moyens technologiques et informatiques performants. Dans ce cadre, TICE souhaite optimiser et contrôler ces données pour mieux gérer ses différents services. Actuellement, l'exploitation se fait de manière ponctuelle et intuitive, ce qui ne permet pas une rentabilité contrôlée. En effet, l'exploration de nombreuses données, souvent hétérogènes, permet d'établir un modèle explicatif et de suivi du fonctionnement des différents services.

D'une manière générale, pour que cette entreprise puisse effectuer une analyse objective du bon fonctionnement de ses différents services et de leurs activités, elle doit d'abord rassembler les données provenant de toutes les sources considérées comme pertinentes, puis les transformer et les stocker dans une base de données très significative. Actuellement et devant un besoin important, l'entreprise TICE souhaite mettre en place un pôle de traitement et analyse de données pour:

- Adapter et anticiper l'offre de BUS aux besoins des personnes, comme le traitement à la demande ou le regroupement avec les données socio-économique.
- Le traitement de la fraude,
- Simuler l'impact d'une grève, etc.
- Gérer le trafic en temps réel

Dans ce cadre, les outils de l'informatique décisionnelle, en particulier les systèmes d'analyse en ligne (On-Line Analytical Processing : OLAP) [1], présentent de nombreux avantages. En effet, ces systèmes permettent une analyse sur de larges

bases de données. D'une manière générale, les approches de l'informatique décisionnelle [1, 2, 3] et de l'apprentissage statistique permettent de développer des outils d'analyse, d'interprétation et de la décision [4, 5, 6, 7].

Objectifs du travail de thèse :

En concertation avec les différents services de l'entreprise TICE, le travail commencera par la définition d'un plan de récolte de données avec une stratégie de stockage adaptée. Ensuite, autour d'un système de gestion de base de données permettant le stockage de l'entrepôt de données, la conception d'un outil d'intégration de données. Ces données souvent hétérogènes et provenant de multiples sources connectées seront intégrées après conditionnement (normalisation) dans l'entrepôt de données. Ces outils, permettent ensuite la restitution de données, et de fournir des statistiques descriptives sur les données.

Les recherches dans cette thèse porteront principalement sur la conception et la réalisation d'un système d'interprétation et de restitution des données provenant de sources hétérogènes permettant une gestion optimale et interactive des informations issues d'objets connectés du réseau du transport de la société TICE. Pour ces objectifs, un modèle d'apprentissage statistique sera développé et mis en place pour l'interprétation et la restitution interactive des données. Les différentes fonctionnalités réalisées seront évaluées dans le cadre d'une architecture OLAP avec un système d'information décisionnel qui pourra apporter des solutions technologiques pertinentes et innovantes à l'entreprise TICE. Des modalités de visualisation de données et d'interaction seront également étudiées afin d'offrir une interface homme machine multimodale et multiplateformes.

Profil candidat(e) :

- Diplômé(e) de Master Recherche (ou équivalent) en Informatique, Recherche opérationnelle, Apprentissage et intelligence artificielle.
- Compétences en développement logiciel, base données et bons bagages scientifiques.
- Intérêt pour la conception et le prototypage rapide, les tests, et l'évaluation avec les utilisateurs finaux.
- Bonne maîtrise de la communication en français et en anglais (oral/écrit).
- Qualités recherchées : grande motivation, autonomie, rigueur, force de proposition, ouverture aux approches pluridisciplinaires

Modalités pratiques :

Cette thèse sera effectuée alternativement au sein du laboratoire IBISC (équipe IRA2) à Evry, France et dans l'entreprise TICE à Evry.

Le candidat retenu sera inscrit en thèse à l'École Doctorale Sciences et technologies de l'information et de la communication (STIC) de l'Université Paris-Saclay.

La thèse sera placée sous la direction de Khalifa DJEMAL (UEVE), codirigée par Samir Otmane (UEVE) et co-encadrée par Karine Hallouin de la société TICE.)

Candidature sur la plateforme ADUM : <https://www.adum.fr/index.pl>

Contact : Khalifa DJEMAL : khalifa.djemal@univ-evry.fr

Références

[1] Lucile Sautot, Conception et implémentation semi-automatique des entrepôts de données : application aux données Ecologiques. Thèse de doctorat de l'Université de Bourgogne, 09/06/2015.

[2] Michaël Briot, Étude sur la mise en œuvre d'un entrepôt de données et conception d'un prototype en vue d'une intégration au sein de France Billet, Mémoire, CNAM, 2013.

[3] Alexis Lechervy, Apprentissage interactif et multi-classes pour la détection de concepts sémantiques dans des données multimédia, Thèse de doctorat de l'Université de Cergy-Pontoise, 6 décembre 2012.

[4] Rostom Kachouri, Khalifa Djemal, Hichem Maaref, Multiple kernel weighting based SVM for heterogeneous image recognition system, International Journal of Signal and Imaging Systems Engineering (IJSISE), Vol. 4, n° 2, pages: 60-70, inderscience, DOI: 10.1504/IJSISE.2011.041600, 2011.

[5] Rostom Kachouri, Khalifa Djemal and Hichem Maaref, Multi-model classification method in heterogeneous image databases, Pattern Recognition (Elsevier), Vol. 43, n° 12, Pages 4077-4088, December 2010.

[6] Khalifa Djemal and Hichem Maaref, Intelligent Information Description and Recognition in Biomedical Image Databases, In: Computational Modeling and Simulation of Intellect: Current State and Future Perspectives, Book Edited by Boris Igelnik, pages: 52-80, Publisher IGI Global, ISBN: 978-1-60960-551-3, February 2011.

[7] Khalifa Djemal, Hichem Maaref and Rostom Kachouri, Image Retrieval System in Heterogeneous Database, In: Automation Control - Theory and Practice, Book Edited by: A. D. Rodic, pages: 327-350, publisher INTECH, ISBN: 978-953-307-039-1, December 2009.