



---

## Thèses 2009

---

Laboratoire d'InfoRmatique en Image et Systèmes d'information

LIRIS UMR 5205 CNRS / INSA de Lyon / Université Claude Bernard Lyon 1 / Université Lumière Lyon 2 / Ecole Centrale de Lyon





---

# Sommaire

---

## Département Données, Connaissances, Services

---

### *Thèses de doctorat*

Interopérabilité sémantique des connaissances des modèles de produit à base de features <b>Samer Abdul ghafour</b> .....	3
A Conceptual Framework for Modelling Spatial Relations <b>Eliseo Clementini</b> .....	5
Visual Summaries of Geographic Databases by Chorems <b>Vincenzo Del fatto</b> .....	7
Application, Recherche et Intégration de Patrons de Sécurité <b>Paul El Khoury</b> .....	9
Interoperability of Technical Enterprise Application <b>Nicolas Figay</b> .....	13
A Declarative Approach for Pervasive Environments: Model and Implementation <b>Yann Gripay</b> .....	15
Extraction de concepts et de relations entre concepts à partir de documents multilingues : Approche statistique et ontologique. <b>Farah Harrathi</b> .....	19
Conception de système à base de traces numériques pour les environnements informatiques documentaires <b>Julien Laflaquière</b> .....	23
Une approche d'apprentissage distribué pour la composition de services web <b>Soufiene Lajmi</b> .....	25
Processus unifié pour la personnalisation des activités pédagogiques : méta-modèle, modèles et outils. <b>Marie Lefevre</b> .....	27

An Ontology-Based Approach to Manage Conflicts in Collaborative Design <b>Moisés Lima Dutra</b> .....	31
Mining String Data under Similarity and Soft-Frequency Constraints: Application to Promoter Sequence Analysis <b>Ieva Mitasiunaite</b> .....	33
Approche distribuée pour la simulation événementielle de réseaux de neurones impulsionsnels. Application au contrôle des saccades oculaires. <b>Anthony Mouraud</b> .....	35
Génération des interfaces adaptatives pour les environnements collaboratifs et nomadiques <b>Youssef Roummieh</b> .....	37
The Chameleon: Un Système de Sécurité pour Utilisateurs Nomades en Environnements Pervasifs et Collaboratifs. <b>Rachid Saadi</b> .....	39
RAevol : un modèle de génétique digital pour étudier l'évolutin des réseaux de régulation génétique <b>Yolanda Sanchez-Dehesa</b> .....	43
Le système ICOP : représentation, visualisation et communication de l'information à partir d'une représentation iconique des données <b>Salma Sassi</b> .....	45
Méthodologie de matching à large échelle pour des schémas XML <b>Sana Sellami</b> .....	47
Un canevas pour l'adaptation et la substitution de services Web <b>Yehia Taher</b> .....	49
Gestion des connaissances pour la maîtrise de la relation entre patrimoine documentaire et Système d'Information <b>Ovidiu Vasutiu</b> .....	51
Modèle multidimensionnel et OLAP sur architecture de grille <b>Pascal Wehrle</b> .....	53

## Département Image

---

### *Thèses de doctorat*

Optimisation multi-objectifs par colonies de fourmis : cas des problèmes de sac à dos <b>Ines Alaya</b> .....	57
implication de l'oxygène et des anti-oxydants dans le processus de radiolyse de l'eau induit par l'irradiation aux ions de haute énergie : simulations numériques pour la radiobiologie. <b>Anthony Coliaux</b> .....	59

Recherche par similarité dans les bases de données multimédia : application à la recherche par le contenu d'images	
<b>Imane Daoudi</b> .....	<b>61</b>
Simulation du Mouvement Pulmonaire pour l'amélioration des traitements par radiothérapie	
<b>Anne-Laure Didier</b> .....	<b>63</b>
Contribution à la numérisation des documents imprimés du XVIIIe siècle : Application au cas de la Gazette de Leyde	
<b>Loris Eynard</b> .....	<b>65</b>
Contributions au tri automatique de documents et de courrier d'entreprises	
<b>Djamel Gaceb</b> .....	<b>67</b>
Le document numérique : la complexité des formes et les formes de la complexité	
<b>Joël Gardes</b> .....	<b>71</b>
Caractérisation des écritures médiévales par des méthodes statistiques basées sur les cooccurrences	
<b>Koubaa Ikram Moalla</b> .....	<b>73</b>
Algorithmes et structures de données compactes pour la visualisation interactive d'objets 3D volumineux	
<b>Clément Jamin</b> .....	<b>75</b>
Analyse multirésolution des images de documents manuscrits : Application à l'analyse de l'écriture	
<b>Guillaume Joutel</b> .....	<b>77</b>
Reconstruction 3D à partir de séquences vidéo pour l'acquisition du mouvement de personnages en temps réel et sans marqueur	
<b>Brice Michoud</b> .....	<b>79</b>
Contributions à la Classification Sémantique d'Images	
<b>Alain Pujol</b> .....	<b>81</b>
Formalisme statistique pour ensembles de structures discrètes	
<b>Sébastien Rebecchi</b> .....	<b>85</b>
Mosaïque d'images multi résolution et applications	
<b>Lionel Robinault</b> .....	<b>87</b>
Algorithmes d'extraction de modèles géométriques discrets pour la représentation robuste des formes	
<b>Tristan Roussillon</b> .....	<b>89</b>
Quantization-Based Blind Watermarking of Three-Dimensional Meshes	
<b>Kai Wa ang</b> .....	<b>91</b>



**Département  
Données, Connaissances, Services**



# Interopérabilité sémantique des connaissances des modèles de produit à base de features

Samer Abdul Ghafour

Thèse

Soutenue le 09/07/2009

à l'Université Claude Bernard Lyon 1

## Jury

Pr Cauvet Corine, Université de Marseille	Rapporteur
Pr Faudot Dominique, Université de Bourgogne, Pr	Rapporteur
Pr Gardan Yvon, Université de Reims	Examineur
Pr Giraudin Jean Pierre, Université Joseph Fourier	Président
Pr Shariat Behzad, LIRIS, Université Lyon 1	Co-directeur
Dr Perna Eliane, LIRIS, Université Lyon 1	Co-directeur
Pr Ghodous Parisa, LIRIS, Université Lyon 1	Directeur

Contact : ghodous@liris.cnrs.fr

## Résumé

Dans un environnement collaboratif de développement de produit, plusieurs acteurs, ayant différents points de vue et intervenant dans plusieurs phases du cycle de vie de produit, doivent communiquer et échanger des connaissances entre eux. Ces connaissances, existant sous différents formats hétérogènes, incluent potentiellement plusieurs concepts tels que l'historique de conception, la structure du produit, les features, les paramètres, les contraintes, et d'autres informations sur le produit. Les exigences industrielles de réduction du temps et du coût de production nécessitent l'amélioration de l'interopérabilité sémantique entre les différents processus de développement afin de surmonter ces problèmes d'hétérogénéité tant au niveau syntaxique, structurel, que sémantique. Dans le domaine de la CAO, la plupart des méthodes existantes pour l'échange de données d'un modèle de produit sont, effectivement, basées sur le transfert des données géométriques. Cependant, ces données ne sont pas suffisantes pour saisir la sémantique des données, telle que l'intention de conception, ainsi que l'édition des modèles après leur échange. De ce fait, nous nous sommes intéressés à l'échange des modèles « intelligents », autrement dit, définis en termes d'historique de construction, de fonctions intelligentes de conception appelées features, y compris les paramètres et les contraintes. L'objectif de notre thèse est de concevoir des méthodes permettant d'améliorer l'interopérabilité sémantique des systèmes CAO moyennant les technologies du Web Sémantique comme les ontologies OWL DL et le langage des règles SWRL. Nous avons donc élaboré une approche d'échange basée sur une ontologie commune de features de conception, que nous avons appelée CDFO « Common Design Features Ontology », servant d'intermédiaire entre les différents systèmes CAO. Cette approche s'appuie principalement sur deux grandes étapes. La première étape consiste en une homogénéisation des formats de représentation des modèles CAO vers un format pivot, en l'occurrence OWL DL. Cette homogénéisation sert à traiter les hétérogénéités syntaxiques entre les formats des modèles. La deuxième étape consiste à définir des règles permettant la mise en correspondance sémantique entre les ontologies d'application de CAO et notre ontologie commune. Cette méthode de mise en correspondance se base principalement, d'une part, sur la définition explicite des axiomes et des règles de correspondance permettant l'alignement des entités de différentes ontologies, et d'autre part sur la reconnaissance automatique des correspondances sémantiques supplémentaires à l'aide des capacités de raisonnement fournies par les moteurs d'inférence basés sur les logiques de description. Enfin, notre méthode de mise en correspondance est enrichie par le développement d'une méthode de calcul de similarité sémantique appropriée pour le langage OWL DL, qui repose principalement sur les composants des entités en question tels que leur description et leur contexte.

## Abstract

A major issue in product development is the exchange and sharing of product knowledge among many actors. This knowledge includes many concepts such as design history, component structure, features, parameters, constraints, and more. Heterogeneous tools and multiple designers are frequently involved in collaborative product development, and designers often use their own terms and definitions to represent a product design. Thus, to efficiently share design information among multiple designers, the design intent should be persistently captured and the semantics of the modeling terms should be semantically processed both by design collaborators and intelligent systems. Regarding CAD models, most of the current CAD systems provide feature-based design for the construction of solid models. Features are devised to carry, semantically, product information throughout its life cycle. Consequently, features should be maintained in a CAD model during its migration among different applications. However, existing solutions for exchanging product information are limited to the process of geometrical data, where semantics assigned to product model are completely lost during the translation process. Current standards, such as ISO 10303, known as STEP have attempted to solve this problem, but they define only syntactic data representation so that semantic data integration is not possible. Moreover, STEP does not provide a sound basis to reason with knowledge. Our research investigates the use of Semantic Web technologies, such as ontologies and rule languages; e.g. SWRL, for the exchange of "intelligent" CAD models among different systems, while maintaining the original relations among entities of the model. Thus, we have proposed an ontological approach based on the construction of a common design features ontology, used as an Interlingua for the exchange of product data. This ontology is represented formally with OWL DL. Furthermore, axioms and mapping rules are defined to achieve the semantic integration between the applications ontologies and the common ontology. The integration process relies basically on reasoning capabilities provided by description logics in order to recognize automatically additional mappings among ontologies entities. Furthermore, the mapping process is enhanced with a semantic similarity measure in order to detect similar design features. However, this will enable data analysis, as well as manage and discover implicit relationships among product data based on semantic modeling and reasoning.

## Publications

### Revues internationales avec comité de lecture

Semantic Interoperability of Heterogeneous Semantic Resources. C Ferreira da Silva, L Médini, S Abdul Ghafour, P Hoffmann, P Ghodous. *Electronic Notes in Theoretical Computer Science* 150(2):71-85, Elsevier. 2006.

### Conférences internationales avec comité de lecture et actes

An Ontology-based Approach for "Procedural CAD Models" Data Exchange. S Abdul Ghafour, P Ghodous, B. Shariat, E Perna. Dans 13th ISPE INTERNATIONAL CONFERENCE ON CONCURRENT ENGINEERING: RESEARCH AND APPLICATIONS, Leading the Web in Concurrent Engineering (CE2006), Parisa Ghodous, Rose Dieng-Kuntz, Geilson Loureiro ed. Antibes, France. 2006.

Interoperability of heterogeneous semantic resources. C Ferreira da Silva, L Médini, S Abdul Ghafour, P Hoffmann, P Ghodous. Dans International Workshop on Database Interoperability (InterDB 2005) in connection with the 7th International Conference on Coordination Models and Languages, Philippe Thiran, Tore Risch, Djamel Benslimane ed. Namur, Belgique. pp. 74-89. 2005.

Towards an Intelligent CAD Models Sharing Based on Semantic Web Technologies. S Abdul Ghafour, P Ghodous, B. Shariat, E Perna. Dans 15th ISPE International Conference on Concurrent Engineering, Springer ed. Belfast, Northern Ireland, UK . Springer . 2008.

Using the Semantic Web for the Integration of Feature-based CAD Models Information. S Abdul Ghafour, P Ghodous, B. Shariat, E Perna. Dans 2008 International Conference on Industrial Informatics and Systems Engineering (IISE 2008), System and Information Sciences Notes ed. Glasgow, UK. Communications of SIWN . ISSN 1757-4439. 2008.

A Common Design-Features Ontology for Product Data Semantics Interoperability. S Abdul Ghafour, P Ghodous, B. Shariat, E Perna. Dans The 2007 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2007)., Silicon Valley - United States of America. 2007.

# A Conceptual Framework for Modelling Spatial Relations

Eliseo Clementini

Thèse

Soutenue le 09/06/2009

à l'INSA Lyon

## Jury

Pr. CLARAMUNT Christophe, IRENAV - Ecole Navale, Brest	Rapporteur
Pr. LAURINI Robert, LIRIS, Lyon	Directeur
Pr. LIGOZAT Gérard, LIMSI - Paris XI, Paris	Rapporteur
Pr. SCHOLL Michel, CEDRIC, CNAM, Paris	Président
Pr. SPACCAPIETRA Stefano, BD - EPFL, Lausanne, Suisse	Examineur

Contact : eliseo.clementini@alice.it

## Abstract

The models of spatial relations have stimulated great interest among researchers in spatial databases from the nineties. We can distinguish three different levels of representation where we can identify spatial relations: a purely geometric one, where objects are represented as point-sets and relations can be formally defined in mathematical terms; a computational level, where objects are represented as spatial data types and relations are calculated using spatial operators; a user level, where objects and relations correspond to the concepts of user context. From a geometric point of view, we can consider a categorization of spatial relations in three groups: topological, projective, and metric. This dissertation proposes both a general framework for modelling qualitative spatial relations and presents new developments for projective relations. By providing a formal mathematical framework to a qualitative description of relations, the models are of great research interest in geographic information science. The most important geometric property that has been taken into account is the collinearity of three points. The importance of this property is such that it influences the whole approach, making ternary relations the formal basis of the models. We have developed algorithms to calculate the relations from a vector data structure and a reasoning system on ternary projective relations. We also extended the model to the three-dimensional space and the sphere. From a user perspective, projective relations should consider context information, combining them with reference systems to avoid ambiguities in the meaning of the relation.

## Publications

### Revues internationales avec comité de lecture

Modeling and Computing Ternary Projective Relations between Regions. E. Clementini, R. BILLEN. IEEE Trans. Knowl. Data Eng. 18(6):799-814, IEEE. 2006.

### Revues nationales avec comité de lecture

Un cadre conceptuel pour modéliser les relations spatiales. E. Clementini, R. Laurini. Revue des Nouvelles Technologies de l'Information RNTI-E-14():1-17, ISSN 978.2.85428.86. 2008.

### Conférences internationales avec comité de lecture et actes

A Qualitative Model for Visibility Relations. F. TARQUINI, G. DE FELICE, P. FOGLIARONI, E. Clementini. Dans KI 2007: Advances in Artificial Intelligence, 30th Annual German Conference on AI, Joachim Hertzberg, Michael Beetz, Roman Englert ed. Osnabrück, Germany, September 10-13, 2007, pp. 510-513. SPRINGER VERLAG . ISBN 978-3-540-74564-. 2007.

Projective Relations in a 3D Environment. R. BILLEN, E. Clementini. Dans 4th International Conference, GIScience 2006, Martin Raubal, Harvey J. Miller, Andrew U. Frank, Michael F. Goodchild ed. Münster, Germany, September 20-23, 2006, pp. 18-32. SPRINGER VERLAG . ISBN 3-540-44526-9. 2006.



# Visual Summaries of Geographic Databases by Chorems

Vincenzo Del Fatto

Thèse

Soutenue le 07/04/2009

à l'INSA de Lyon

## Jury

Pr. MARTIN Hervé, LIG, Grenoble, Paris	Président
Pr. LAURINI Robert, LIRIS, Lyon	Co-directeur
Pr. VITIELLO GIULIANA, DMI, Salerno, Italie	Co-directeur
Pr. CASTANO Silvana, Univ. Milano, Milan, Italie	Rapporteur
Dr. RUAS Anne, COGIT, Paris	Rapporteur
Pr. COSTABILE Maria-Francesca, Univ. Bari, Bari, Italie	Examineur

Contact : vdelfatt@unisa.it

## Abstract

Traditional cartography is a fundamental tool to visually describe facts and relationships concerning with territory. This is a well-known and well-established approach and decision makers are usually satisfied by its expressiveness when it concerns the cartography of facts. Differently, this kind of cartography may fail when dealing with scenarios referring to heterogeneous issues, such as political, economic and demographic problems, due to the large amount of complex data to represent in a map. Then, more effective solutions in supporting users to locate facts, trends and new patterns should be investigated. In this dissertation the research carried out within an international project is presented, meant to define cartographic solutions able to better represent geographic information extracted from database contents, which refer both to geographic objects and spatio-temporal phenomena. An actual support for human activity to model and analyze the reality of interest may indeed consist of an immediate synthesis of the most relevant data, disregarding details. Such a synthesis may be based on the usage of visual metaphors, which are able to capture and reconstitute the most salient features of a scenario. Moreover, it may represent the starting point for further processing tasks aimed to derive spatial analysis data, and to support expert users in decision making, thus bridging the gap between the complexity of the adopted applications and the need for rapid and exhaustive responses expected by domain experts. The major contribution of this work along this line has been to define a methodology to visualize geographic database summaries, expressing them through "schematized representations of territories", known as chorems. In particular, two specific contributions have been produced by investigating and implementing the proposed methodology. The former consists of the formal specification of chorems in terms of visual language and structure, in order to both standardize the chorem creation and assembling process and provide a usable framework for computer systems. The latter is represented by the design and the implementation of a system which generates maps containing chorems starting from geographic database content, in a semi-automatic manner.

## Publications

### Revue internationale avec comité de lecture

A Chorem-based Approach for Visually Synthesizing Complex Phenomena. V. Del Fatto, R. Laurini, K. Lopez, M. Sebillio, G. Vitiello. *Information Visualization* 2008(7):253-264, Palgrave Macmillan. 2008.

A usability-driven approach to the development of a 3D web-GIS environment. V. Del Fatto, L. Paolino, F. Pittarello. *Journal of Visual Languages and Computing* 18(3):280-314, Academic Press, Inc., ISSN 1045-926X. 2007.

### **Conférences internationales avec comité de lecture et actes**

A Comparison between 2D and 3D Iconic Visual Arrangements for Representing Topological Relationships. V. Del Fatto, L Paolino, M Sebillio, G Tortora, G Vitiello. Dans *Visual Languages and Computing 2007 (VLC 2007)*, San Francisco, USA. 2007. (à paraître)

Guaranteeing Integrity Constraints during Interactive Map Editing with Visual Language Parsing. V. Del Fatto, V. Deufemia, L Paolino. Dans *Second IEEE International Conference on Digital Information Management (ICDIM'07)*, Youakim Badr, Richard Chbeir, Pit Pichappan ed. INSA de Lyon, France. pp. 19-24. ISBN 1-4244-1476-8. 2007.

First Findings in Chorems Discovery and Layout. R. Laurini, K. Lopez, V. Del Fatto. Dans *2nd International Symposium on Generalization of Information*, Geneva, Switzerland, October 1-3, 2007. pp. 85-104. Horst Kemers CODATA Germany. ISBN 978-3-00-022382-. 2007. (à paraître)

Chorem Editing - Visual Summary of Spatial Database Content. V. Del Fatto, R. Laurini, K. Lopez, M Sebillio, G Tortora, M. Tucci, G Vitiello. In *Proceedings of the 13th "International Conference on Distributed Multimedia Systems - Workshop on Visual Languages and Computing (VLC2007)"* San Francisco USA, 6-8 Sept. 2007, pp.256-259. 2007.

Early Performance Analysis in the Design of Spatial Databases. V. Del Fatto, M. Giordano, G. Polese, M Sebillio, G Tortora. Dans *2nd International Conference on Software and Data Technologies. (ICSOFT2007)*, Barcelona/Spain.. pp. 172-177. 2007.

Potentialities of Chorems as Visual Summaries of Geographic Databases Contents. V. Del Fatto, R. Laurini, K. Lopez, R. Loreto, F. Milleret-Raffort, M Sebillio, D. Sol-Martinez, G Vitiello. *VISUAL 2007, 9th International Conference on Visual Information Systems*, Shanghai, China, 28-29 June 2007. Edited by Qiu, G., Leung, C., Xue, X.-Y., Laurini, R., Springer Verlag LNCS, Volume 4781 "Advances in Visual Information Systems", pp. 537-548 2007.

WebMGISQL 3D - Iterating the Design Process passing through a Usability Study. V. Del Fatto, L Paolino, F. Pittarello, M Sebillio, G Tortora, G Vitiello. Dans *Proceedings of the 20th BCS HCI Group conference &#8220;Human-Computer Interaction &#8211; HCI2006&#8221;*, Londra UK, 11/09 ÷ 15/09 2006.. 2006.

Representing Topological Relationships by Using 3D objects: An Empirical Survey. V. Del Fatto, L Paolino, F. Pittarello. Dans *Proceedings of the 12 International Workshop on Visual Languages and Computing (VLC2006)*, Grand Canyon USA, 30/8 ÷ 1/9 2006. 2006.

Visual Analysis of Spatial Data through Maps of Chorems. D. DE CHIARA, V. Del Fatto, R. Laurini, M Sebillio, G Vitiello. Dans *In Proceedings of the 15th International Conference on Distributed Multimedia Systems*, San Francisco, September 10-12, 2009, Edited by Chang SK, Celentano A; and Yoshitaka A., Published by Knowledge Systems Institute, . pp. 295-300. ISBN 1-891706-25-X. 2009.

Visualizing Geographical Analysis Results From Spatial Databases Based on the Chorems. K. Lopez, R. Laurini, V. Del Fatto, D. Sol-Martinez, R. Loreto, M Sebillio, G Vitiello. Dans *International Conference on Modeling, Simulation and Visualization Methods (MSV'09: July 13-16, 2009, USA , Las Vegas, USA. 2009.* (à paraître)

### **Conférences invité**

Condividere il patrimonio cartografico tramite un sistema basato su un core di metadati geografici. V. Del Fatto, L Paolino, M Sebillio, G Tortora, G Vitiello. Conferenza Tematica AMFM2006: Informazione Geografica: Data Harmonisation, Interoperabilità e Standard, Reti e Facilities Management, Ubiquitous GIS, Roma 21 - 22 Settembre, 2006. 2006.

Soutenue le 15/12/2009

à l'Université de Lyon- Université Claude Bernard Lyon 1

## Jury

Mme. Véronique BENZAKEN Université de Paris XI	Rapporteur
Mme. Elisa BERTINO - Université de Purdue, USA	Examineur
M. Emmanuel COQUERY - Université Claude Bernard Lyon 1	Co-encadrant
M. Frédéric CUPPENS - ENST Bretagne	Examineur
Mme. Elena FERRARI - Université d'Insubria, Italie	Rapporteur
M. Mohand-Saïd HACID - Université Claude Bernard Lyon 1	Directeur de thèse
M. Volkmar LOTZ – SAP Recherche Sécurité	Examineur
M. Jean-Christophe PAZZAGLIA – SAP Recherche	Invité

## Résumé

Dans le cycle de développement d'un logiciel, l'expertise de sécurité est une qualité manquante qui doit être adressée. Les patrons, particulièrement les patrons de sécurité, sont des moyens pour rétablir le transfert de connaissances entre des experts et des non experts dans une communauté. Dans cette thèse, nous nous focalisons sur l'ingénierie de la sécurité en utilisant des patrons de sécurité. Notre investigation de la littérature a révélé des manques dans les propositions de patrons de sécurité pour établir un véritable transfert de connaissances entre

les experts de sécurité et les utilisateurs non spécialisés. En particulier, cette thèse apporte les contributions techniques suivantes :

Un nouveau format pour les patrons de sécurité pour transférer la connaissance des experts de sécurité à des utilisateurs non spécialisés. Nous présentons une étude qui expose en détail les approches pour les patrons de sécurité en les comparant aux patrons de conceptions. Cette étude identifie les inconvénients des patrons de sécurité actuels et montre que la recherche, la classification, et la composition de patrons de sécurité sont les fonctionnalités de base nécessaires pour une plus grande adoption de cette approche. De plus nous introduisons le format XML utilisé pour conserver des patrons de sécurité dans une bibliothèque de patrons de sécurité. Nous décrivons la structure de la bibliothèque de patrons de sécurité et nous décrivons les patrons de sécurité disponibles. En outre, nous présentons des exemples de patrons de sécurité qui appartiennent à différentes infrastructures du cycle de développement de logiciel. Nous présentons aussi notre travail sur les Composantes Exécutables (EC) dans le contexte du projet Européen SERENITY.

Une interface ontologique (disponible en tant que greffon Eclipse) fournit la nouvelle classification et les approches de recherche des patrons de sécurité par les experts de sécurité et les développeurs de logiciel. Nous analysons, moyennant des questionnaires et des interviews de développeurs de logiciels, les informations collectées et ensuite nous modélisons deux ontologies : la première pour les développeurs et la seconde pour décrire des patrons de sécurité. Notre méthodologie définit un processus pour les experts de sécurité qui leur permet de décrire des patrons de sécurité et leurs Composantes Exécutables pour alimenter la bibliothèque. En plus, elle fournit l'équipement nécessaire pour aider des experts de sécurité à mettre en place cette bibliothèque via le langage OWL-DL d'ontologie. Au final, on fournit un outil semi-automatique pour que les utilisateurs puissent récupérer des patrons de sécurité et des composants exécutables tout en ayant une connaissance limitée de la sécurité dont ils ont besoin.

Une approche pour l'intégration des patrons de sécurité. Pour supporter le raisonnement lié à l'intégration

de patrons de sécurité et la création de langages de patrons, nous présentons une revue détaillée sur l'intégration de solutions de sécurité pour différentes infrastructures. Nous classifions les conflits et les regroupons sur la base de leurs contextes. Nous distinguons les conflits de sécurité et les conflits fonctionnels. Par ailleurs, nous utilisons la transformation de graphes. Nous avons ainsi conçu d'un langage de patrons basés sur deux types de relations : « est en conflit » et « exige ».

L'univers de la télémédecine, est particulièrement riche en données sensibles et en infrastructures domotiques. Nous présentons une étude de cas qui se concentre sur l'assistance lointaine de patients localisés dans leurs maisons équipées (intelligentes). Nous avons développé un prototype de télémédecine pour valider les idées développées dans cette thèse.

#### **Mots-clefs :**

Patrons de sécurité, Ontologie, Intégration de patrons de sécurité, télémédecine.

#### **Publications**

##### **Revue internationale avec comité de lecture**

An XACML-based Security Pattern to achieve Socio-Technical Confidentiality in Smart Homes. P. Busnel, Paul El Khoury, S. Giroux, K. Li. *International Journal of Smart Home* Vol. 3(No. 1). 2009.

A Requirements Engineering Methodology for the Development of Security and Privacy Patterns Compliant with Legal Requirements. L. Compagna, Paul El Khoury, A. Krausová, F. Massacci, N. Zannone. *Artificial Intelligence and Law* 17():1-30, Springer, ISSN 0924-8463 (Pri. 2008).

##### **Conférences internationales avec comité de lecture et actes**

Consistency Checking of Role Assignments in Inter-Organizational Collaboration. Paul El Khoury, E Coquery, M. Hacid. Dans *Geographic Information System, Proceedings of the SIGSPATIAL ACM GIS 2008 International Workshop on Security and Privacy in GIS and LBS*, ACM ed. Irvine, California. pp. 82-88. Policies . ACM New York, NY, USA. ISBN 978-1-60558-324-2. 2008.

Achieving Socio-Technical Confidentiality using Security Pattern in Smart Homes. P. Busnel, Paul El Khoury, S. Giroux, K. Li. Dans *Future Generation Communication and Networking, 2008. FGNC '08. Second International Conference on Smart Homes*, IEEE ed. Hainan Island, China. pp. 447-452. IEEE Explore . ISBN 978-0-7695-3431-2. ISSN 10439166. 2008.

Security Patterns for Capturing Encryption-Based Access Control to Sensor Data. A. Cuevas, Paul El Khoury, L. Gomez, A. Laube. Dans *Second International Conference on Emerging Security Information, Systems and Technologies, 2008. SECURWARE '08.*, IEEE ed. Cap Esterel, France. pp. 62-67. ISBN 978-0-7695-3329-2. ISSN 10204997. 2008.

An Ontological Interface for Software Developers to Select Security Patterns. Paul El Khoury, A. Mokhtari, E Coquery, M. Hacid. Dans *2nd International Workshop on Secure systems methodologies using patterns (SPattern'08) in conjunction with the 19th International Conference on Database and Expert Systems Application, 2008. DEXA '08.*, IEEE ed. Turin, Italy. pp. 297-301. ISBN 978-0-7695-3299-8. ISSN 1529-4188. 2008.

S&D Pattern Deployment at Organizational Level: A Prototype for Remote Healthcare System. P. Busnel, Paul El Khoury, K. Li, A. Saidane, N. Zannone. Dans *4th International Workshop on Security and Trust Management. In Conjunction with IFIPTM, joint iTrust and PST Conferences on Privacy, Trust Management and Security*, Electronic Notes in Theoretical Computer Science ed. Trondheim, Norway. 2008.

XACML as a Security and Dependability (S&D) pattern for Access Control in Aml environments. F. Sanchez-Cid, A. Muñoz, Paul El Khoury, L. Compagna. Dans *Developing Ambient Intelligence Proceedings of the International Conference on Ambient Intelligence Developments (AmI.d&#8217;07)*, Springer Paris ed. . pp. 143-155. *Developing Ambient Intelligence* . ISBN 978-2-287-78543-6. 2007.

How to capture, model, and verify the knowledge of legal, security, and privacy experts: a pattern-based approach. L. Compagna, Paul El Khoury, F. Massacci, R. Thomas, N. Zannone. Dans International Conference on Artificial Intelligence and Law, Proceedings of the 11th international conference on Artificial intelligence and law, ACM ed. Stanford, California. pp. 149-153. Modelling aspects of law . ACM New York, NY, USA. ISBN 978-1-59593-680-6. 2007.

A Security Pattern for Untraceable Secret Handshakes. A. Cuevas, Paul El Khoury, L. Gomez, A. Laube, A. Sorniotti. Dans SECUREWARE, . 2009. (à paraître)

#### **Conférences invité**

Secure Remote Healthcare Environment using Serenity. Paul El Khoury, K. Li. Dans European Research towards Trusted Ambient intelligence 2nd European Workshop, Sophia Antipolis, France. 2008.

#### **Contributions à un ouvrage**

Serenity in e-Business and Smart Items Scenarios. A. Benameur, Paul El Khoury, M. Seguran, S. Kumar Sinha. Security and Dependability for Ambient Intelligence Series: Advances in Information Security , Vol. 55 Spanoudakis, George; Mana Gomez, Antonio; Spyros, Kokolakis (Eds.) 2009, Approx. 375 p. 20 illus., Hardcover ISBN: 978-0-387-88774-6 2009. (à paraître)

A Study on Recent Trends on Integration of Security Mechanisms. Paul El Khoury, M. Hacid, S. Smriti Kumar, E Coquery. Advances in Data Management, special volume of Studies in Computational Intelligence. Series Ed.: Kacprzyk, Janusz. 2009. (à paraître)



Nicolas Figay

Soutenue le 27/11/2009

à l'Université Lyon 1

### Jury

Mr. Aris. M. OUKSEL, université de l'Illinois à Chicago, USA	Rapporteur
Mr. AIT-AMEUR Yamine, Ecole Nationale Supérieure de Mécanique et d'Aérotechnique, Mr. FENG Shaw, NIST, USA	Rapporteur Examineur
Mr. JARDIM-GONCALVEZ Ricardo, nouvelle université de Lisbonne, UNINOVA-GRIS, Portugal	Examineur
Mr. GARDAN Yvon, université de Reims-Champagne-Ardennes, groupe MONGE	Examineur
Mr. MONDON Jean-Yves, Groupe EADS (Phenix)	Examineur
Mme. GHODOUS Parisa, Université Lyon 1	Directeur de thèse

Contact : nicolas.figay@eads.net

### Résumé

Dans le contexte économique actuel, les entreprises font face à de nouveaux problèmes en termes d'interopérabilité, du fait de besoins croissants de collaboration eBusiness dans les écosystèmes numériques auxquels elles appartiennent. Elles ont également besoins de pouvoir rentabiliser et faire évoluer les applications internes existantes. De plus, l'établissement rapide d'une collaboration numérique avec un membre de leur écosystème, limitée dans le temps, ne devrait pas nécessiter de modification de leurs infrastructures de communications et de leurs applications pour pouvoir échanger information et connaissance. D'un côté, les solutions actuelles sont de moins en moins adaptées pour faire face ni aux besoins croissants d'interopérabilité dans des environnements de plus en plus complexes. D'un autre côté, il n'est pas envisageable de remplacer les standards et les cadres d'interopérabilité actuellement utilisés en proposant des innovations de rupture. Les travaux de recherches réalisés dans le cadre de la thèse « Interopérabilité des applications techniques d'entreprise » concernent le développement d'une approche innovante pour construire un cadre d'interopérabilité des applications d'entreprise basé sur l'utilisation simultanée et cohérente des standards d'interopérabilité d'un écosystème et des technologies associées. L'objectif est l'interopérabilité « pragmatique ». L'approche innovante propose s'appuie sur les apports conjugués de l'ingénierie par les modèles, de la modélisation d'entreprise, des ontologies et des architectures orientées services. Elle promeut l'utilisation des commodités du WEB, basées sur des standards ouverts et gouvernés. Ce faisant, la préservation sémantique entre les standards de l'écosystème considéré, les artefacts d'ingénierie des applications et les infrastructures de communication est cruciale. Aussi l'approche innovante proposée inclut-elle le concept "d'hyper modèle étendu", qui a été développée dans le cadre de cette thèse, et dont l'usage est illustré dans le cadre particulier des applications de gestion du cycle de vie des produits industriels, au sein de l'entreprise étendue.

### Abstract

Within the current economic context, enterprises are facing new interoperability issues due to increasing needs of eBusiness Collaboration within the emerging digital ecosystems they belong to. They also need to be able to keep in pace with their heterogeneous internal legacy systems. In addition, they should not have to modify their infrastructure or applications for fast and short collaboration implying information and knowledge interchange with new partners of their ecosystem. In one hand, current solutions are less and less adapted to face increasing needs and complexity in term of interoperability. In the other hand, legacy interoperability standards and frameworks can't be replaced as it can be imagined to propose new disruptive approach and technologies. The research work undertaken for the thesis "Interoperability of Technical Enterprise Interoperability" consists in proposing an innovative approach allowing a given and

mature ecosystem to build an enterprise application interoperability framework based on simultaneous and coherent usage of eBusiness standards used by a given ecosystem, combining usage of the different relevant frameworks supporting these standards. The goal is achievement of "pragmatic" interoperability. Proposed innovative approach takes advantage of simultaneous usage of Model Driven Engineering, Enterprise Modeling, Ontology and Service Oriented Architecture. It promotes systematic usage of commodities on the WEB based on open and governed standards. Doing so, semantic preservation between ecosystem's standards, application engineering artifacts and communication infrastructures is crucial. To support semantic preservation within the context of the innovative proposed approach, the concept of "extended hypermodel" is developed and demonstrated within the context of Product Lifecycle Management within networked organizations.

## **Publications**

### **Revue internationale avec comité de lecture**

Enabling interoperability of STEP Application Protocols at meta-data and knowledge level. NFI Figay, R. Jardim-Gonçalves, A. Steiger-Garção. *Inderscience* 36(4):402-421. 2006.

### **Conférences internationales avec comité de lecture et actes**

Extended hypermodel for interoperability within the Virtual Enterprise. NFI Figay, P Ghodous. Dans SITIS2009, Tunisie. 2009. (à paraître)

Innovative Interoperability Framework for Enterprise Applications within Virtual Enterprises. NFI Figay, P Ghodous. Dans MEDES 2009, Lyon, France. 2009. (à paraître)

FLOSS as Enterprise Applications Interoperability Enabler. NFI Figay, P Ghodous. Dans SITIS2009, Tunisie. 2009. (à paraître)

Collaborative Product Development: EADS Pilot based on ATHENA results. NFI Figay, P Ghodous. Dans I-ESA 2008, Berlin, Germany. pp. 423-425. 2008.

### **Conférences invité**

Partage de données produit par le biais de processus de collaboration. Solutions proposée via le projet Athena . NFI Figay. Dans Conference MICADO INDAO 2007, Lyon. 2007. (diffusion restreinte)

OntologySummit2009 Panel Session -. NFI Figay. Dans OntologySummit2009 Panel Session - "Toward Ontology-based Standards" - Thu 26-Mar-2009, Teleconf. 2009.

### **Contributions à un ouvrage**

Unleashing the Potential of the European Knowledge Economy Value Proposition for Enterprise Interoperability. NFI Figay, L. Man-Sze, S. Crave, A. Grilo. *Unleashing the Potential of the European Knowledge Economy &#8212; Value Proposition for Enterprise Interoperability* 2008.

Collaborative Product Development: EADS Pilot Based on ATHENA. P Ghodous, NFI Figay. *Enterprise Interoperability* 2008.

The ATHENA interoperability framework. A. Berre, B. Elvesoeter, NFI Figay, C. Guglielmina, S. Johnsen, D. Karlsten, T. Knothe. *Enterprise Interoperability II New Challenges and Approaches* Volume package *Enterprise Interoperability* 2007, XVIII, 894 p. 322 illus., Hardcover ISBN: 2007.

# A Declarative Approach for Pervasive Environments: Model and Implementation

Yann Gripay

Thèse

Soutenue le 10/12/2009

à l'INSA de Lyon

## Jury

Pr Benzaken Véronique, Université de Paris XI	Président
Pr Elmagarmid Ahmed K., Université de Purdue, Etats-Unis	Rapporteur
Pr Grumbach Stéphane, INRIA / Chinese Academy of Sciences, Chine	Rapporteur
Dr Laforest Frédérique, INSA de Lyon	Co-directeur
Dr Manolescu Ioana, INRIA Saclay	Rapporteur
M. Menga David, EDF R&D	Invité
Pr Petit Jean-Marc, INSA de Lyon	Co-directeur

Contact : yann.gripay@liris.cnrs.fr

## Résumé

Les environnements informatiques évoluent vers ce qu'on appelle des systèmes pervasifs : ils ont tendance à être de plus en plus hétérogènes, décentralisés et autonomes. D'une part, les ordinateurs personnels et autres terminaux mobiles sont largement répandus et occupent une grande place dans les systèmes d'information. D'autre part, les sources de données et fonctionnalités disponibles peuvent être réparties sur de larges espaces grâce à des réseaux allant du réseau mondial Internet jusqu'aux réseaux locaux pair-à-pair pour les capteurs. Elles sont de plus dynamiques et hétérogènes: bases de données avec des mises à jour fréquentes, flux de données provenant de capteurs logiques ou physiques, et services fournissant des données stockées ou provenant de capteurs, transformant des données ou commandant des actionneurs. Les environnements pervasifs posent de nouveaux défis pour exploiter leur plein potentiel, en particulier la gestion d'interactions complexes entre ressources réparties. Il est cependant difficile de gérer ces sources de données et fonctionnalités hétérogènes avec les systèmes actuels, ce qui constitue un frein pour le développement d'applications pervasives. Il est ainsi nécessaire de combiner au sein de développements ad hoc des langages de programmation impératifs (C++, Java...), des langages de requêtes classiques pour les bases de données (SQL...) et des protocoles réseau (JMX, UPnP...). Ce n'est cependant une solution ni pratique ni adéquate sur le long terme. Les approches déclaratives offrent l'avantage de fournir une vue logique des ressources qui abstrait les problématiques d'accès physique et permet la mise en œuvre de techniques d'optimisation. Les requêtes SQL sur les bases de données relationnelles en sont une illustration typique et bien connue. C'est pourquoi la définition déclarative de requêtes sur des sources de données et des fonctionnalités est reconnue comme un défi majeur dans le but de simplifier le développement d'applications pervasives. Actuellement, les extensions des SGBDs (Système de Gestion de Bases de Données) permettent d'avoir une vue homogène et d'effectuer des requêtes sur des bases de données et des flux de données (notamment les SGFDs, Système de Gestion de Flux de Données). La notion de service est un moyen courant de représenter les fonctionnalités réparties d'un système informatique, mais n'est pas encore pleinement intégrée au sein des SGBDs. Malgré de nombreuses propositions, une compréhension claire des interactions entre données, flux de données et services manque toujours, ce qui constitue un frein majeur pour la définition déclarative des applications pervasives, en lieu et place des actuels développements ad hoc. Dans cette thèse, nous proposons un framework définissant une vue orientée données des environnements pervasifs: la notion classique de base de données est étendue pour construire une notion plus large, l'environnement pervasif relationnel, qui intègre les sources de données à la fois conventionnelles et non-conventionnelles, à savoir données, flux de données et services. Cette notion permet le développement d'applications pervasives de manière déclarative en utilisant des requêtes continues orientées service qui combinent ces sources de données. Dans ce framework, nous proposons un

modèle de données pour les environnements pervasifs, appelé SoCQ (pour Service-oriented Continuous Query), qui prend en compte leur hétérogénéité, dynamique et répartition. Nous définissons la structure de notre modèle de données avec la notion de relation dynamique étendue (eXtended Dynamic Relation, ou XD-Relation) représentant les sources de données. Nous définissons également un langage algébrique pour notre modèle de données avec l'algèbre Serena (Service-enabled algebra), à partir de laquelle un langage de type SQL a été défini. Ce langage permet d'exprimer de manière déclarative des requêtes sur les environnements pervasifs. Afin d'implémenter ce framework, nous avons conçu une architecture de système de gestion d'environnements pervasifs (Pervasive Environment Management System, ou PEMS) qui prend en charge notre modèle de données. Un PEMS est un système de gestion dynamique de données et de services qui gère de manière transparente les problématiques liées au réseau telles que la découverte de services et les interactions à distance. Il supporte l'exécution de requêtes ponctuelles et continues orientées service que les développeurs d'applications peuvent aisément concevoir pour développer des applications pervasives. Un prototype de PEMS a été implémenté, avec lequel des expérimentations ont été réalisées.

## Abstract

Computing environments evolve toward what is called pervasive systems: they tend to be more and more heterogeneous, decentralized and autonomous. On the one hand, personal computers and other handheld devices are largely widespread and take a large part of information systems. On the other hand, available data sources and functionalities may be distributed over large areas through networks that range from a world-wide network like the Internet to local peer-to-peer connections like for sensors. They are dynamic and heterogeneous: distributed databases with frequent updates, data streams from logical or physical sensors, and services providing data from sensors or storage units, transforming data or commanding actuators. Pervasive environments pose new challenges in order to exploit their full potential, in particular through the management of complex interactions between distributed resources. Their heterogeneous data sources and functionalities are not homogeneously manageable in today's systems. This is a big issue when building pervasive applications. Imperative programming languages (e.g., C++, Java), classical query languages for databases (e.g., SQL), and network protocols (e.g., JMX, UPnP) must be combined in ad hoc developments, which is neither convenient nor suitable as a long-term solution. Declarative approaches offer the advantage of providing a logical view on resources that abstracts physical access issues and enables optimization techniques. SQL queries over relational databases are a typical and well-known illustration of those approaches. Therefore, querying data sources and functionalities in a declarative way is recognized as a major issue in pervasive environments in order to simplify the development of applications. Currently, extensions of DBMSs (DataBase Management Systems) provide a homogeneous view and query facilities for both relational data and data streams (e.g., DSMSs (Data Stream Management Systems), CEP (Complex Event Processing), ESP (Event Stream Processing)). Services are a common way to represent distributed functionalities in a computing environment, but are not yet fully integrated with DBMSs. Despite a lot of propositions, a clear understanding of the interplays between relational data, data streams and services is still lacking and is the major bottleneck toward the declarative definition of pervasive applications, instead of the current ad hoc development of such applications. In this thesis, we propose a framework that defines a data-centric view of pervasive environments: the standard notion of database is extended to come up with a broader notion, defined as relational pervasive environment, integrating both conventional and non-conventional data sources, namely data, streams and services. It enables the development of applications for pervasive environments using declarative service-oriented continuous queries combining those data sources. Within this framework, we propose a data model for pervasive environments, namely the SoCQ data model (standing for Service-oriented Continuous Query), that takes into account their heterogeneity, dynamicity and distribution. We define the structure of our data model with the notion of eXtended Dynamic Relation (XD-Relation) representing data sources. We also define an algebraic language for our data model with the Service-enabled algebra (Serena algebra), from which a SQL-like language (the Serena SQL) has been devised. This language enables the expression of declarative queries over pervasive environments. In order to implement this framework, we have designed a Pervasive Environment Management System (PEMS) that supports our data model. A PEMS is a service-enabled dynamic data management system that seamlessly handles network issues like service discovery and

remote interactions. It supports the execution of service-oriented one-shot and continuous queries that application developers can easily devise to build pervasive applications. A prototype of PEMS has been implemented, on which experimentations have been conducted.

## **Publications**

### **Revues nationales avec comité de lecture**

Towards service-oriented continuous queries in pervasive systems. Y. Gripay, F. Laforest, J-M Petit. RSTI - ISI 13(5):33-57. 2008.

### **Conférences internationales avec comité de lecture et actes**

Service-oriented Continuous Queries for Pervasive Systems. Y. Gripay. Dans EDBT 2008 PhD Workshop (unofficial proceedings), Nantes, France. pp. 1-7. 2008.

Context-Sensitive Security Framework for Pervasive Environments. C.E. Pigeot, Y. Gripay, V. Scuturici, J. Pierson. Dans ECUMN2007, IEEE ed. Toulouse. pp. 391-400. 2007.

A Simple (yet Powerful) Algebra for Pervasive Environments. Y. Gripay, F. Laforest, J-M Petit. Dans EDBT 2010, 13th International Conference on Extending Database Technology, Lausanne, Switzerland. pp. 1-12. 2010. (à paraître)

SoCQ: A Framework for Pervasive Environments. Y. Gripay, F. Laforest, J-M Petit. Dans ISPAN 2009, 10th International Symposium on Pervasive Systems, Algorithms and Networks, Kaohsiung, Taiwan, R.O.C.. pp. 1-6. 2009. (à paraître)

### **Conférences nationales avec comité de lecture et actes**

Vers une algèbre relationnelle étendue aux services. Y. Gripay, F. Laforest, J-M Petit. Dans BDA 2008, Guilhaumand-Granges (Ardèche), France. pp. 1-20. 2008. (diffusion restreinte)

Towards Action-Oriented Continuous Queries in Pervasive Systems. Y. Gripay, F. Laforest, J-M Petit. Dans Bases de Données Avancées 2007 (BDA'07), Marseille. pp. 1-20. 2007. (diffusion restreinte)

Une architecture pervasive sécurisée : PerSE. Y. Gripay, J. Pierson, C.E. Pigeot, V. Scuturici. Dans UbiMob'06, ACM ed. Paris. pp. 147-150. 2006.

### **Autres conférences**

Données en environnement pervasif : support pour la proactivité. V. Scuturici, Y. Gripay, F. Conil. (XXVème congrès INFORSID, Atelier GEDSIP : Gestion de données dans les systèmes d'information pervasifs), , Perros-Guirec. 2007.

### **Contributions à un ouvrage**

Managing Pervasive Environments through Database Principles: A Survey. Y. Gripay, F. Laforest, J-M Petit. Advances in Data Management 2009.

### **Brevets et Logiciels**

SoCQ: un système de gestion de données et de services en environnement pervasif. Y. Gripay, L. La, F. Laforest, J-M Petit. Dans BDA'09, 25èmes journées Bases de Données Avancées, Namur, Belgique. pp. 1-5. 2009.

### **Rapports de recherche/technique**

Context-Sensitive Security in a Pervasive Environment. C.E. Pigeot, Y. Gripay, J. Pierson, V. Scuturici. Rapport de recherche RR-LIRIS-2006-017 2006.



# Extraction de concepts et de relations entre concepts à partir de documents multilingues : Approche statistique et ontologique.

**Farah Harrathi**

**Thèse**

Soutenue le 28/09/2009

à l'INSA de Lyon

## **Jury**

M. Gargouri Faïez, MIRACL , Tunisie	Rapporteur
M. Simonet Michel, TIMC	Rapporteur
M. Benhamadou Abdelmajid, MIRACL , Tunisie	Président
M. Gammoudi Mohamed Mohsen, VIAREC, Tunisie	Directeur
Mle Calabretto Sylvie, LIRIS	Directeur
M. Jean-Marie Pinon, LIRIS	Examineur
Mle Roussey Catherine, LIRIS	Co-directeur

Contact : farah.harrathi@insa-lyon.fr

## **Résumé**

Les travaux menés dans le cadre de cette thèse se situent dans la problématique de recherche- indexation des documents et plus spécifiquement dans celle de l'extraction des descripteurs sémantiques pour l'indexation. Le but de la Recherche d'Information (RI) est de mettre en œuvre un ensemble de modèles et de systèmes permettant la sélection d'un ensemble de documents satisfaisant un besoin utilisateur en termes d'information, exprimé sous la forme d'une requête. Un Système de Recherche d'Information (SRI) est composé principalement de deux processus. Un processus de représentation et un processus de recherche. Le processus de représentation, appelé indexation, permet de représenter les documents et la requête par des descripteurs ou des indexs. Ces descripteurs reflètent le contenu des documents. Le processus de recherche consiste à comparer les représentations des documents à la représentation de la requête. Dans les SRI classiques, les descripteurs utilisés sont des mots (simples ou composés). Ces SRIs considèrent le document comme étant un ensemble de mots, souvent appelé « sac de mots ». Dans ces systèmes, les mots sont considérés comme des graphies sans sémantique. Les seules informations exploitées concernant ces mots sont leurs fréquences d'apparition dans les documents. Ces systèmes ne prennent pas en considération les relations sémantiques entre les mots. Par exemple, il est impossible de trouver des documents représentés par un mot M1 synonyme d'un mot M2, dans le cas où la requête est représentée par M2. Aussi, dans un SRI classique un document indexé par le terme « bus » ne sera jamais retrouvé par une requête indexée par le terme « taxi », bien qu'il s'agisse de deux termes qui traitent le même thème « moyen de transport ». Afin de remédier à ces limites, plusieurs travaux se sont intéressés à la prise en compte de l'aspect sémantique des termes d'indexation. Ce type d'indexation est appelé indexation sémantique ou conceptuelle. Ces travaux passent du niveau des mots au niveau des concepts (les sens des mots) ; ainsi, les descripteurs d'un document sont des concepts. Dans ces travaux les termes dénotant les concepts sont extraits à partir du document en utilisant des techniques statistiques ou/et linguistiques. Ces termes sont par la suite projetés sur une ressource sémantique (ontologie, thésaurus...) afin d'extraire les concepts associés. Les approches d'indexation sémantique existantes ont été principalement appliquées aux corpus monolingues. Dans ces corpus tous les documents sont écrits dans une même langue. Ces approches utilisent un analyseur morphosyntaxique (lemmatiseur et étiqueteur) de la langue du corpus pour l'indexer. De ce fait, ils ne s'appliquent pas à des corpus multilingues où les documents du corpus sont écrits dans plus d'une langue. L'objectif de notre travail de thèse et de proposer une approche d'indexation sémantique adaptée aux corpus multilingues. Dans ce cadre nous proposons une approche statistique et ontologique d'indexation adaptée aux documents multilingues. Nous proposons une technique statistique exploitant les fréquences de mots afin d'extraire les termes des

documents. Par la suite, le modèle d'ontologie est utilisé afin d'associer les termes à leurs concepts. Ce modèle est aussi utilisé pour extraire les relations entre les concepts à partir des documents. Nous proposons des algorithmes indépendants de la langue des textes pour reconnaître des concepts et des relations de l'ontologie dans les textes. De ce fait, l'approche est « robuste » et indépendante de la langue et du domaine du corpus. Concernant la validation, nous appliquons notre approche sur un corpus réel, le corpus médical de la campagne d'évaluation CLEF'2007, en utilisant le méta-thésaurus UMLS.

## **Abstract**

The research work of this thesis is related to the problem of document search and indexing, and more specifically the extraction of semantic descriptors for document indexing. An Information Retrieval System (IRS) is a set of models and systems for selecting a set of documents satisfying user needs in terms of information expressed as a query. In IR, a query is composed mainly of two processes for representation and retrieval. The representation process is called indexing, it allows to represent documents and query descriptors, or indexes. These descriptors reflect the contents of documents. The retrieval process consists in comparing document representations and query representation. In classical IRS, the descriptors used are words (simple or compound). These IRS consider the document as a set of words, often called a "bag of words". In these systems, words are considered as graphy without semantics. The only information used for these words is their occurrence frequency in the documents. These systems do not take into account the semantic relationships between words. For example, it is impossible to find documents represented by a word M1synonyms of word M2, where the query is represented by M2. Also, in a classic IRS document indexed by the term "bus" will never be found by a query indexed by the word "taxi", although these two words deal with the same subject "means of transportation." To address these limitations, several studies were interested in taking into account the semantic indexing terms. This type of indexing is called semantic or conceptual indexing. These works take into account the notion of concept in place of the notion of word. In this work the terms denoting concepts are extracted from the document by using statistical techniques. These terms are then projected onto semantics resources such as: ontology, thesaurus and so on to extract the concepts involved. Existing approaches for semantic indexing have been applied mainly to monolingual corpus. These approaches use a morphosyntactic analyzer for indexing. As a result, they do not apply to multilingual corpora. The aim of this thesis is to propose an approach of semantic indexing adapted to a multilingual corpus. In this context we propose a statistical and ontological indexing approach adapted to multilingual documents. A statistical technique use the frequency of words in order to extract the terms of the documents. The ontology model is used to associate the words to concepts. This model is also used to extract the relations between concepts from documents. We propose algorithms independent of the language of the texts to identify concepts and relations of the ontology in the texts. Hence the approach is "robust" and independent of the language and the domain corpus. Regarding validation, we apply our approach to a real corpus, the corpus of medical evaluation campaign CLEF'2007 using the UMLS meta-thesaurus.

## **Publications**

### **Conférences internationales avec comité de lecture et actes**

Multilingual Extraction of Semantic Indexes. C. Roussey, S. Calabretto, F. Harrathi. Dans Proceedings of the 2007 international workshop on Semantically aware document processing and indexing 2007, ACM ed. Montpellier, France.. pp. 1-8. International Conference Proceeding Series 259. ISBN 978-1-15159-668-. 2007.

Multilingual Indexing Based on Ontologies. C. Roussey, S. Calabretto, F. Harrathi, M. Gammoudi. Dans 13th ISPE International Conference On Concurrent Engineering: Research And Applications, Parisa Ghodous, Rose Dieng-Kuntz, Geilson Loureiro ed. Antibes. pp. 418-425. Frontiers in Artificial Intelligence and Applications 143. IOS Press Nieuwe Hemweg 6B 1013 BG Amsterdam Netherlands. ISBN 1-58603-651-3. ISSN 0922-6389. 2006.

### **Conférences nationales avec comité de lecture et actes**

Indexation semi automatique de corpus multilingues basée sur une ontologie. F. Harrathi, S. Calabretto, C. Roussey. Dans Colloque Indice, Index, indexation, Ismail TIMIMI, Susan KOVACS ed. Lille. pp. 203-219.

Sciences et Techniques de l'information . ADBS 25 rue Claude Tillier 75012 Paris. ISBN 10:2-84365-088-7. ISSN 1762-8288. 2005.

Indexation sémantique des documents multilingues. F. Harrathi, C. Roussey, S. Calabretto, L. Maisonnasse, M. Gammoudi. Dans Atelier RISE associé au 27ème Congrès INFORSID 2009, INFORSID ed. Toulouse. pp. 31-50. 2009.

#### **Conférences invité**

Recherche d'information sémantique multilingue. S. Calabretto, C. Roussey, F. Harrathi. Dans 7ème colloque du chapitre français de l'ISKO. Intelligence collective et organisation des connaissances., M. Hassoun et M. El Hachani ed. Lyon, 24-26 juin. 2009.

#### **Autres conférences**

Natural Language Processing Method for Multilingual Semantic Indexing. C. Roussey, S. Calabretto, F. Harrathi. Dans 12th International Conference on Applications of Natural Language to Information Systems, CNAM, Paris, France. 2007.



# Conception de système à base de traces numériques pour les environnements informatiques documentaires

Thèse

Julien Laflaquière

Soutenue le 08/12/2009

à l'Université de Technologie de Troyes

## Jury

Dr Bachimont Bruno, Heudyasic (UTC), Compiègne	Rapporteur
Pr Leleu-Merviel Sylvie, Dream (UVHC), Valenciennes	Rapporteur
Dr Aussenac-Gilles Nathalie, IRIT, Toulouse	Examineur
Pr Zreik Kaldhoun, Paragraphe (Paris8), Paris	Président
Pr Mille Alain, LIRIS (Lyon), Lyon	Invité
Pr Zacklad Manuel, Tech-cico (ICD-UTT), Troyes	Directeur
Dr Prié Yannick, LIRIS, Lyon	Directeur

Contact : julien.laflaquiere@liris.cnrs.fr

## Résumé

Cette thèse traite de la problématique générale des traces numériques liées à la réalisation d'une activité documentaire instrumentée. Elle s'inscrit dans le cadre de l'approche des Systèmes à Base de Traces modélisée (SBTm) pour laquelle elle développe pour la première fois une méthodologie complète de modélisation des traces numériques qui lui est propre. L'approche des SBTm propose de définir a priori les objets de l'observation automatique dans un modèle de trace. La conception de la méthodologie proposée repose sur des fondements théoriques tirés de domaines de recherche impliquant des formes de modélisation de l'activité instrumentée (IHM, IC, CSCW), et sur une mise en pratique dans deux situations d'activités réelles, l'une d'apprentissage collaboratif à distance, l'autre de production de contenus de formation. Dans le cadre du second terrain d'application, un développement technique de visualisation de traces a permis la tenue d'une expérimentation spécialement conçue pour évaluer le potentiel des traces modélisées en tant que support de réflexivité d'une activité documentaire complexe. Les premiers résultats de ce travail exploratoire dans des situations d'activités collectives ont conduit à étendre le cadre conceptuel des SBTm. L'apport de la notion de trace conjointe, qui désigne des traces générées et réutilisées par des collectifs d'utilisateurs, offre de nouvelles perspectives d'usages étendant la problématique générale des traces modélisées.

## Abstract

This thesis addresses the general topic of digital traces related to the achievement of user activity in a digital documentary space. This work is part of the modelised Trace Based Systems approach (mTBS). This work contributes to the approach by creating a specific digital trace modelization methodology. The mTBS approach aims to make some reusable digital traces in user activity context. In order to produce digital trace as qualitative activity representation, it proposes to use a trace model which predefine observed objects. The proposed methodology design is based firstly on theoretical foundations drawn from research domains involving human-computer activity modeling (CHI, KM, CSCW), and secondly based on an implementation work in two actual activities, on the one hand a collaborative elearning activity, on the other hand a activity of professional training content production. During this second applied research study, we developed a trace specific display tool. This tool has been used for support an experiment designed to assess the modeled traces potential as an activity reflectivity medium. Early results of this exploratory work in collective activity situations have led us to propose to extend the conceptual prime framework of the mTBS approach. We suggested the "conjoined trace" which means traces generated and used by groups of users. It offers new collaborative practice prospects extending the general problem of trace-based approaches.

## **Publications**

### **Conférences internationales avec comité de lecture et actes**

Des traces modélisées, un nouvel objet pédagogique ?. J. Laflaquière, Y. Prié, A. Mille. Dans 4th annual scientific conference - LORNET, Montréal. 2007.

A trace-based System Framework for Experience Management and Engineering. J. Laflaquière, L. S Settouti, Y. Prié, A. Mille. Dans Second International Workshop on Experience Management and Engineering (EME 2006) in conjunction with KES2006, Bournemouth UK. 2006.

### **Conférences nationales avec comité de lecture et actes**

Facilitation de tâches informatiquement médiées : une approche centrée sur la réflexivité de l'utilisation. J. Laflaquière, A Ciaccia. Dans 6ème colloque des jeunes chercheurs en sciences cognitives, Bordeaux, France. pp. 117-117. 2005.

Approche de modélisation de l'expérience : utilisation de systèmes complexes pour l'assistance aux tâches de veille informatiquement médiées. J. Laflaquière, P-A. Champin, Y. Prié, A. Mille. Dans ISKO-France 2005, Amos David ed. INIST-CNRS, Vandoeuvre-lès-Nancy, France. pp. 209-230. Presse Universitaire de Nancy . ISBN 2-86480-817-X. 2005.

Modélisation d'utilisation de système pour une assistance à base de trace : une application de Musette à la tâche de veille documentaire. J. Laflaquière, Y. Prié. Dans Workshop Traces, Interactions, co-constructions collectives et relations à la cognition. CoMETE AS du RTP38, Paris. 2003.

L'expérience tracée des activités conjointes instrumentées. J. Laflaquière, Y. Prié. Dans Atelier Interaction, Contextes, Traces, Université de Caen. 2009.

Ingénierie des traces numériques d'interaction comme inscriptions de connaissances. J. Laflaquière, Y. Prié, A. Mille. Dans 19èmes Journées Francophones d'Ingénierie des Connaissances, Nancy. pp. 183-195. 2008.

Traces d'utilisation et réflexivité : la question du collectif. J. Laflaquière, Y. Prié. Dans 11ème colloque de l'ARCo - Arco07, . 2007.

### **Autres conférences**

Traces et inscriptions de connaissances. J. Laflaquière, L. S Settouti, Y. Prié, A. Mille. Dans 18e journées francophones d'Ingénierie des Connaissances, Grenoble. pp. 329-330. 2007.

### **Contributions à un ouvrage**

Musette : Modéliser les Usages et les Tâches pour Tracer l'Expérience. J. Laflaquière, Y. Prié. Réutilisation de l'expérience : modèles et applications 2009. (à paraître)

### **Rapports de recherche/technique**

Un environnement pour gérer des traces comme inscriptions de connaissances. J. Laflaquière, L. S Settouti, Y. Prié, A. Mille. Rapport de recherche RR-LIRIS-2007-022 , Soumis à RFIA2008. 2007.

# Une approche d'apprentissage distribué pour la composition de services web

Soufiene Lajmi

Thèse

Soutenue le 26/11/2009

à l'Université de Lyon- Université Claude Bernard Lyon 1

## Jury

Nadine Cullot	rapporteur
Mohamed Mohsen Gammoudi	rapporteur
Vincent Hilaire	examineur
Faiez Gargouri	examineur
Khaled Ghedira	directeur de thèse
Chirine Ghedira	Co-directrice de thèse
Djamal Benslimane	Co-directeur de thèse

## Résumé

Ces dernières années, les services web ont eu un intérêt important au niveau de la recherche ainsi qu'à l'industrie. Le présent sujet de thèse entre dans le cadre de la composition de services web et de leur intégration. En effet, l'utilisation simple des services web existants peut ne pas répondre au besoin du demandeur de service. Une solution possible est de combiner un ensemble de services afin d'avoir une réponse satisfaisante.

L'objectif global de ce sujet est d'améliorer le processus de découverte et d'intégration de services web. Dans ce travail, nous présentons une solution pour la composition de service web basée essentiellement sur le raisonnement à partir de cas. Ensuite, nous proposons d'améliorer le processus de recherche des cas similaires par la classification des cas. L'étape suivante consiste à définir une stratégie de recherche pour explorer la base de cas. Pour ce faire, nous avons défini deux stratégies. La première est basée sur l'exploration d'un arbre de décision pour la recherche de la classe contenant les cas les plus similaires à un nouveau problème donné par l'utilisateur. La deuxième stratégie utilise le concept de treillis de Galois. Enfin, nous proposons de distribuer notre architecture afin d'exploiter les solutions fournies par d'autres sites.

**Mots clés :** Service Web, Composition de Services Web (CSW), Treillis de Galois, Arbre de décision, Web Sémantique, SMA, Raisonnement à Partir de Cas (RàPC).

## Publications

### Revues internationales avec comité de lecture

CBR Method for Web Service Composition. S. Lajmi, C. Ghedira, K Ghedira. Lecture Notes in Computer Science 4879():314-326, Springer. 2009. (à paraître)

### Conférences internationales avec comité de lecture et actes

Agent-Based framework for Web services composition. S. Lajmi, C. Ghedira, K Ghedira. Dans The Fifth European workshop On Multi-Agent Systems (EUMAS'2007), Hamamet, Tunisia. 2007.

How to apply CBR methods in Web service composition?. S. Lajmi, C. Ghedira, K Ghedira. Dans THE INTERNATIONAL CONFERENCE ON SIGNAL-IMAGE TECHNOLOGY & INTERNET-BASED SYSTEMS (SITIS'2006), Hammamet-Tunisia. pp. 230-239. 2006.

WeSCo\_CBR: How to Compose Web Services via Case Based Reasoning. S. Lajmi, C. Ghedira, K Ghedira, D.

Benslimane. Dans The IEEE International Symposium on Service-Oriented Applications, Integration and Collaboration held with the IEEE International Conference on e-Business Engineering (ICEBE 2006), Shanghai, China. pp. 618-622. 2006.

**Conférences nationales avec comité de lecture et actes**

Une méthode d'apprentissage pour la composition de services web. S. Lajmi, C. Ghedira, K Ghedira. Dans Atelier Intelligence Artificielle et Web Intelligence (IAWI'07), Grenoble, France. 2007.

# Processus unifié pour la personnalisation des activités pédagogiques : méta-modèle, modèles et outils.

Thèse

Marie Lefevre

Soutenue le 01/12/2009

à l'Université Claude Bernard Lyon 1

## Jury

Mme Grandbastien Monique, LORIA, Nancy	Rapporteur
M. Vidal Philippe, IRIT, Toulouse	Rapporteur
M. Nodenot Thierry, LIUPPA, Anglet	Président
M. Marty Jean-Charles, SYSCOM, Chambéry	Examineur
M. Mille Alain, LIRIS, Villeurbanne	Directeur
Mme Jean-Daubias Stéphanie, LIRIS, Villeurbanne	Co-directeur
Mme Guin Nathalie, LIRIS, Villeurbanne	Co-directeur

Contact : marie.lefevre@gmail.com

## Résumé

Cette thèse en informatique se situe dans le domaine des Environnements Informatiques pour l'Apprentissage Humain (EIAH). Dans ce cadre, nous avons abordé la question de la personnalisation de l'apprentissage. Nos travaux de recherche consistaient à identifier un processus qui permette à la fois de personnaliser des séances de travail sur papier et des séances de travail sur des logiciels pédagogiques. Nous souhaitons que ce processus permette de prendre en compte les spécificités de chaque apprenant en s'appuyant sur son profil, mais qu'il prenne également en compte les buts et les habitudes pédagogiques des enseignants. Enfin, nous souhaitons que ce processus soit implémentable dans un système externe aux logiciels à personnaliser. Notre problématique s'est donc décomposée en trois points : comment exploiter les profils d'apprenants pour prendre en compte les individualités des apprenants ? Comment adapter une activité pour prendre en compte les besoins et habitudes pédagogiques d'un enseignant ? Et enfin, comment attribuer une activité à un apprenant ? Pour répondre au premier point, nous avons proposé le modèle cPMDL. Ce complément du langage de modélisation des profils PMDL permet de contraindre les profils des apprenants afin de sélectionner ceux ayant les caractéristiques requises par les contraintes. cPMDL nous permet donc d'exploiter les informations contenues dans les profils au sein du processus de personnalisation. Pour répondre au deuxième point, nous avons proposé l'approche GEPPETO. Cette approche s'appuie sur des modèles et des processus génériques permettant d'adapter les activités en fonction des intentions pédagogiques des enseignants grâce à la définition de contraintes sur les activités. Nous avons décliné cette approche pour permettre l'adaptation des activités papier (GEPPETOP), ainsi que pour l'adaptation des activités logicielles et de la configuration des environnements qui les supportent (GEPPETOS). Pour répondre au troisième point, nous avons proposé le modèle PERSUA2 qui permet de lier les contraintes sur profils de cPMDL aux contraintes sur activités de GEPPETO. Ces liens, nommés règles d'affectation, sont ensuite hiérarchisés selon leur degré de priorité pour former une stratégie pédagogique qui sera associée à un ou plusieurs contextes d'utilisation. Nous avons mis en œuvre ces différentes contributions théoriques dans Adapte, un module de l'environnement informatique du projet PERLEA. Le rôle de cet environnement est d'assister l'enseignant dans la gestion de profils créés par l'enseignant lui-même ou issus de logiciels pédagogiques. Adapte est l'une des exploitations possibles des profils : le module réalisé permet de fournir à chaque apprenant des activités adaptées à son profil tout en respectant les choix pédagogiques de son enseignant. Ces activités peuvent être des activités papier proposées par le système ou des activités logicielles personnalisées par Adapte, mais effectuées dans un autre EIAH. Ce module, pleinement opérationnel, a montré la faisabilité technique de nos contributions théoriques et nous a permis de conduire des mises à l'essai auprès d'enseignants.

## Abstract

This thesis in computer science belongs to the field of Interactive Learning Environments (ILE). In this context, we have addressed the issue of personalization of learning. Our research has consisted in identifying a process allowing one to personalize both paper working sessions and working sessions on educational software. Our goal was to design a process able to take into account the specificities of each learner, based on their profiles, but also to take into account the pedagogical goals and habits of teachers. Moreover, we had to design this process such as it could be easily implemented in a software external to the system being personalized. Our problem was therefore decomposed into three points: how to use learners' profiles to take the individuality of learners into account? How to adapt a pedagogical activity to take the teaching needs and habits of a teacher into account? And finally, how to assign an activity to a learner? To answer the first point, we have proposed the cPMDL model. This complement of the profiles modelling language PMDL allows one to constrain the learners' profiles to select those with the characteristics required by the constraints. cPMDL allows us to exploit the information contained in the profiles during the personalization process. To answer the second point, we have proposed the GEPPETO approach. This approach relies on generic models and processes to adapt activities according to the teachers' intentions, by defining constraints on activities. We have instantiated this approach to enable on the one hand the adaptation of paper activities (GEPPETOp) and on the other hand to enable the adaptation of software activities and the adaptation of configuration of environments that support them (GEPPETOs). To address the third point, we have proposed the PERSUA2 model which links cPMDL constraints on profiles with GEPPETO constraints on activities. Next, these links, called assignment rules, are organized into a hierarchy according to their priority degree in order to form a pedagogical strategy. This pedagogical strategy is then associated with one or more contexts of use. We have implemented these different theoretical contributions in *Adapte*, a module of the environment associated to the PERLEA project. The role of this environment is to assist teachers in the management of profiles created by themselves or coming from pedagogical software. *Adapte* is one of the possible uses of profiles: the module developed provides each learner with activities suited to their profiles, while respecting the teaching choices of the teacher. These activities may be paper activities proposed by the system or software activities personalized by *Adapte* but made within another ILE. This module, fully operational, has demonstrated the technical feasibility of our theoretical contributions and has allowed us to conduct experiment with teachers.

## Publications

### Conférences internationales avec comité de lecture et actes

Supporting Acquisition of Knowledge to Personalize Interactive Learning Environments through a Meta-Model. M. Lefevre, S. Jean-Daubias, N. Guin. Dans ICCE 2009 (17th International Conference on Computers in Education), Hong Kong. pp. 439-446. 2009.

A Meta-Model to Acquire Relevant Knowledge for Interactive Learning Environments Personalization. M. Lefevre, A. Mille, S. Jean-Daubias, N. Guin. Dans ADAPTIVE 2009, Athènes, Grèce. pp. 78-85. 2009.

Generation of pencil and paper exercises to personalize learners' work sequences: typology of exercises and meta-architecture for generators. M. Lefevre, S. Jean-Daubias, N. Guin. Dans E-Learn 2009 - World Conference on E-Learning in Corporate, Government, Healthcare & Higher Education, Vancouver, Canada. pp. 2843-2848. 2009.

Generation of exercises within the PERLEA project. S. Jean-Daubias, M. Lefevre, N. Guin. Dans 2nd Workshop on Question Generation, AIED'09 - Artificial Intelligence in Education, Brighton, Great Britain. pp. 38-42. AIED 2009 Workshops Proceedings 1. 2009.

*Adapte*, a Tool for the Teacher to Personalize Activities. M. Lefevre, N. Guin, S. Jean-Daubias. Dans ITS 2008, Montreal. pp. 699-701. 2008.

### Conférences nationales avec comité de lecture et actes

Uniformisation de la structure de profils d'apprenants issus de sources hétérogènes. S. Jean-Daubias, C. Eyssautier-Bavay, M. Lefevre. Dans EIAH 2009 - Environnements Informatique pour l'Apprentissage

Humain, Le Mans, France. pp. 77-84. ISBN 978-2-7342-1149-5. 2009.

Adapte, un outil pour proposer des activités personnalisées au sein du projet PERLEA. M. Lefevre. Dans RJC EIAH 2008, Lille. pp. 145-146. 2008.

#### **Autres conférences**

Génération de feuilles d'exercices adaptées aux profils d'apprenants dans le cadre du projet PERLEA. M. Lefevre, S. Jean-Daubias. Dans Poster, EIAH 2007 - Environnements Informatique pour l'Apprentissage Humain, Lausanne (Suisse). 2007.

Personnaliser des séquences de travail à partir de profils d'apprenants. M. Lefevre, S. Jean-Daubias, N. Guin. Dans Poster, EIAH 2009 - Environnements Informatiques pour l'Apprentissage Humain, Le Mans, France. 2009.

#### **Rapports de recherche/technique**

Generation of pencil and paper activities within the PERLEA project. S. Jean-Daubias, M. Lefevre, N. Guin. Rapport de recherche RR-LIRIS-2009-016 2009.

Generation of exercises within the PERLEA project. M. Lefevre, S. Jean-Daubias, N. Guin. Rapport de recherche RR-LIRIS-2008-011 , Soumis à ICALT 2008. 2008.

Génération d'exercices au sein du projet PERLEA. M. Lefevre, S. Jean-Daubias, N. Guin. Rapport de recherche RR-LIRIS-2007-030 2007.



# An Ontology-Based Approach to Manage Conflicts in Collaborative Design

Moisés Lima Dutra

Thèse

Soutenue le 27/11/2009

à l'Université Claude Bernard Lyon 1

## Jury

Pr Ouksel Aris, Université d'Illinois, USA	Rapporteur
Pr Gardan Yvon, Université de Reims	Rapporteur
Pr Ait Ameer Yamine, Université de Poitier	Président
Dr Feng Shaw, NIST, USA	Examineur
Pr Gonçalves Ricardo, UNINOVA, Université New Lisbonne, Portugal	Co-directeur
Pr Ghodous Parisa, LIRIS	Directeur

Contact : ghodous@liris.cnrs.fr

## Abstract

Abstract Today's complex design projects require teams of designers to come together to facilitate the sharing of their respective expertise in order to produce effective design solutions. Due to the increasing need for exchanging knowledge, modern design projects are more structured to work with distributed virtual teams that collaborate over computer networks to achieve global optima in design. Nevertheless, in the collaborative design process, the integration of multidisciplinary virtual teams – involving exchange and sharing of knowledge and expertise – frequently generates a lot of conflicting situations. Different experts' viewpoints and perspectives, in addition to several ways of communicating and collaborating at the knowledge level, make all this process very hard to tackle. In order to achieve an optimal scenario, some problems must firstly be solved, such as: requirement specification and formalization, ontology integration, and conflict detection and resolution. Specifying and formalizing the knowledge demands a great effort towards obtaining representation patterns that aggregate several disjoint knowledge areas. Each expert should express himself so that the others can understand his information correctly. It is necessary, therefore, to use a flexible and sufficiently extensive data representation model to accomplish such a task. Some current models fall short of providing an effective solution to effective knowledge sharing and collaboration on design projects, because they fail to combine the geographical, temporal, and functional design aspects with a flexible and heterogeneous knowledge representation model. This work proposes an architecture for collaborative design that intends to be synchronous, heterogeneous, service-oriented, agent-based, and ontology-based. Particular representation models are transformed into ontology instances and merged together in order to accomplish the final product design. It is a synchronous approach because the merging process is undertaken at the same time that the interaction among the designers takes place. It is heterogeneous because it provides the users with two approaches for ontology integration: the use of a generic ontology and the harmonization process. Our proposal focuses on collaborative design conflicts and makes use of Web Ontology Language (OWL) and Web Services, the former as a tool for knowledge representation and the latter as a technological support for communication.

## Publications

### Revues internationales avec comité de lecture

A Generic and Synchronous Ontology-based Architecture for Collaborative Design. M Lima Dutra, P Ghodous, O. Kuhn, T. Minh. Concurrent Engineering, Research and Applications (), Sage. 2010. (à paraître)

### Conférences internationales avec comité de lecture et actes

Resolving Collaborative Design Conflicts Through an Ontology-based Approach. M Lima Dutra, P Ghodous, R. Jardim-Gonçalves. Dans 15th ISPE Conference on Concurrent Engineering, Springer-Verlag London ed. Belfast, Irlande du Nord, Août 2008. pp. 11-19. Advanced Concurrent Engineering . Springer London . ISBN

978-1-84800-971-4. ISSN 1865-5440. 2008.

Collaborative Architecture Based on Web-Services. O. Kuhn, M Lima Dutra, P Ghodous, T. Dusch, P. Collet. Dans 15th ISPE Conference on Concurrent Engineering, Springer-Verlag London ed. Belfast, Irlande du Nord, Août 2008. pp. 53-62. Advanced Concurrent Engineering . Springer London . ISBN 978-1-84800-971-4. ISSN 1865-5440. 2008.

Using an Inference Engine to Detect Conflicts in Collaborative Design. M Lima Dutra, C Ferreira da Silva, P Ghodous, R. Jardim-Gonçalves. Dans 14th International Conference on Concurrent Enterprising (ICE 2008), Lisbonne, Portugal, Juin 2008. pp. 133-140. Advances in Concurrent Engineering . ISBN 978 0 85358 244 1. 2008.

Collaborative Environment for the Detection and the Follow-up of Legionella. Yousse Roummieh, P Ghodous, P. Vanehms, C. Verdier, M Lima Dutra, M.C. NICOLLE. Dans Methods and Design models for Health Information Systems, workshops in conjunction with IEEE ICDIM'2007, Lyon, France. pp. 609-614. ISBN 978-1-4244-1475-8. 2007.

A Reasoning Approach for Conflict Dealing in Collaborative Design. M Lima Dutra, P Ghodous. Dans 14th ISPE Conference on Concurrent Engineering, São José dos Campos, Brazil. pp. 495-502. Complex Systems Concurrent Engineering . Springer London . ISBN 978-1-84628-975-0. 2007.

EXPRESS to OWL morphism: making possible to enrich ISO10303 Modules.. M Lima Dutra, C. Agostinho, R. Jardim-Gonçalves, P Ghodous, A. Steiger-Garção. Dans 14th ISPE Conference on Concurrent Engineering, São José dos Campos, Brazil. pp. 391-402. Complex Systems Concurrent Engineering . Springer London . ISBN 978-1-84628-975-0. 2007.

Distributed Artificial Intelligent System for Collaborative Design. M Lima Dutra, K. Slimani, P Ghodous. Dans 13th ISPE Conference on Concurrent Engineering, Juan Les Pins, France. pp. 128-135. FRONTIERS IN ARTIFICIAL INTELLIGENCE AND APPLICATIONS . IOS Press . ISSN 0922-6389. 2006.

# Mining String Data under Similarity and Soft-Frequency Constraints: Application to Promoter Sequence Analysis

Ieva Mitasiunaite

Thèse

Soutenue le 19/05/2009

à l'INSA de Lyon

## Jury

Pr Berthold Michael R., University of Konstanz, Constance, Allemagne	Examineur
Pr Boulicaut Jean-Francois, INSA Lyon, LIRIS UMR CNRS 5205, Lyon	Directeur
Dr Gandrillon Olivier, UCBL, CGMC UMR CNRS 5534, Lyon	Examineur
Pr King Ross D., University of Wales, Aberystwyth, Pays de Galles, Royaume-Uni	Examineur
Dr Mouchiroud Dominique, UCBL, LBBE UMR CNRS 5558, Lyon	Examineur
Pr Siebes Arno, Universiteit Utrecht, Utrecht, Pays-Bas	Rapporteur
Dr Tesseire Maguelonne, CEMAGREF, Montpellier, Montpellier	Rapporteur

Contact : ieva.mitasiunaite@gmail.com

## Abstract

An inductive database is a database that contains not only data but also patterns. Inductive databases are designed to support the KDD process. Recent advances in inductive databases research have given rise to a generic solvers capable of solving inductive queries that are arbitrary Boolean combinations of anti-monotonic and monotonic constraints. They are designed to mine different types of pattern (i.e., patterns from different pattern languages). An instance of such a generic solver exists that is capable of mining string patterns from string data sets. In our main application, promoter sequence analysis, there is a requirement to handle fault-tolerance, as the data intrinsically contains errors, and the phenomenon we are trying to capture is fundamentally degenerate. Our research contribution to fault-tolerant pattern extraction in string data sets is the use of a generic solver, based on a non-trivial formalisation of fault-tolerant pattern extraction as a constraint-based mining task. We identified the stages in the process of the extraction of such patterns where state-of-art strategies can be applied to prune the search space. We then developed a fault-tolerant pattern match function *InsDels* that generic constraint solving strategies can soundly tackle. We also focused on making local patterns actionable. The bottleneck of most local pattern extraction methods is the burden of spurious patterns. As the analysis of patterns by the application domain experts is time consuming, we cannot afford to present patterns without any objective clue about their relevancy. Therefore we have developed two methods of computing the expected number of patterns extracted in random data sets. If the number of extracted patterns is strongly different from the expected number from random data sets, one can then state that the results exhibits local associations that are a priori relevant because they are unexpected. Among others applications, we have applied our approach to support the discovery of new motifs in gene promoter sequences with promising results.

## Publications

### Revue internationale avec comité de lecture

Extracting Signature Motifs from Promoter Sets of Differentially Expressed Genes. I Mitasiunaite, C. Rigotti, S. Schicklin, L. Meyniel, J-F. Boulicaut, O Gandrillon. In *Silico Biology* 8(0043):17-39. 2008.

### Conférences internationales avec comité de lecture et actes

Parameter Tuning for Differential Mining of String Patterns. J Besson, C. Rigotti, I Mitasiunaite, J-F. Boulicaut. Dans Proc. 2nd International Workshop on Domain Driven Data Mining DDDM'08 co-located with ICDM'08, Pisa, Italy. pp. 77-86. IEEE Computer Society . 2008.

About softness for inductive querying on sequence databases. I Mitasiunaite, J-F. Boulicaut. Dans Proc. 7th

Int. Baltic Conf. on Databases and Information Systems DB&IS'06, Vilnius, Lithuania. pp. 77-82. IEEE . 2006.

#### **Autres conférences**

New mechanisms of v-ErbA oncogene action revealed by SAGE analysis. C. Bresson, C. Keime, C. Faure, Y. Letrillard, I Mitasiunaite, C. Rigotti, J-F. Boulicaut, A. Bendjeladi, N. Benhra, O Gandrillon, S. Gonin-Giraud. Dans Integrative Post-Genomics, IPG'06, Lyon, France. 2006.

Using pattern discovery techniques in computational biology to discover new putative transcription factor binding sites in the promoter sequences of insulin-regulated genes. I Mitasiunaite, J Besson, E. Meugnier, J-F. Boulicaut, H. Vidal, S Rome. Dans Integrative Post-Genomics IPG'05, Lyon, France. 2005.

#### **Contributions à un ouvrage**

Introducing softness into inductive queries on string databases. I Mitasiunaite, J-F. Boulicaut. Databases and Information Systems IV. Volume 155 Frontiers in Artificial Intelligence and Applications, O. Vasilecas et al.(Eds.), pp. 117-132. IOS Press. 2007.

# Approche distribuée pour la simulation événementielle de réseaux de neurones impulsifs. Application au contrôle des saccades oculaires.

Thèse

**Anthony Mouraud**

Soutenue le 29/06/2009

à l'Université Antilles Guyane

## Jury

M. Alexandre Frédéric, INRIA Lorraine, Nancy	Rapporteur
M. Tarroux Philippe, LIMSI, Orsay	Rapporteur
M. Lefevre Laurent, INRIA Rhône-Alpes, Lyon	Examineur
M. Guillaume Alain, Mouvement & Perception, Marseille	Examineur
Mme Paugam-Moisy Hélène, LIRIS, Lyon	Co-directeur
M. Puzenat Didier, GRIMAAG, Pointe-à-Pitre	Directeur

Contact : anthony.mouraud@gmail.com

## Résumé

Pour les réseaux de neurones impulsifs, effectuer les simulations en mode événementiel (event-driven) permet de réduire le temps d'exécution séquentiel par rapport aux méthodes clock-driven. D'autre part, l'utilisation d'un support parallèle permet de profiter d'un plus grand nombre de ressources matérielles pour optimiser les performances des simulations. Ce travail de thèse propose un simulateur événementiel, multithreadé et distribué pour la simulation de réseaux de neurones impulsifs de grande taille. Il est dénommé par le sigle DAMNED, qui signifie Distributed And Multithreaded Neural Event-Driven simulation framework. Répartissant le réseau de neurones sur les ressources matérielles synchronisées par une méthode décentralisée de gestion du temps virtuel, DAMNED introduit également un fonctionnement multithread. DAMNED permet d'accélérer les calculs et de simuler des réseaux de plus grande taille qu'en séquentiel. DAMNED offre la possibilité d'exploiter de nombreux modèles de réseaux et de neurones impulsifs et la plupart des supports matériels sont exploitables. Nous présentons l'utilisation de DAMNED sur un modèle simple de réseau pour différentes valeurs de tailles, connectivités et dynamiques. Ensuite, nous proposons une application directe de DAMNED dans une modélisation du contrôle des saccades oculaires. Ce modèle, entièrement basé sur des neurones impulsifs, étudie les circuits neuronaux du système saccadique restreints au tronc cérébral. On montre, à l'aide de ce modèle, que l'hypothèse selon laquelle une somme vectorielle ("vector summation") des activités de la carte motrice du colliculus supérieur coderait pour l'amplitude de la saccade correspond davantage aux données obtenues pour le modèle exécuté sur DAMNED qu'à l'hypothèse d'un moyennage de vecteurs ("vector average"). Même si des évolutions et optimisations sont envisagées, l'originalité de ce travail, parmi les premiers simulateurs distribués de réseaux de neurones impulsifs, réside dans le couplage d'une stratégie événementielle, d'un multithreading interne aux processus logiques et une architecture physique distribuée. Le simulateur DAMNED constitue donc une avancée dans le domaine des réseaux de neurones impulsifs, et particulièrement dans celui des réseaux de grande taille. Les expériences réalisées sur le modèle connexionniste du contrôle des saccades oculaires, et les résultats qu'elles apportent à la communauté neuroscientifique confirment les perspectives d'utilisation de ce travail de thèse.

## Abstract

Simulating spiking neuron networks with a sequential event-driven approach consumes less computation time than clock-driven methods. On the other hand, a parallel computing support provides a larger amount of material resources for optimizing simulation performance. This PhD dissertation proposes an event-driven, multithreaded and distributed framework for simulating large size spiking neuron networks. The name of the simulator is the acronym DAMNED, for Distributed And Multithreaded Neural Event-Driven

simulation framework. DAMNED distributes the neurons and connections of the network on the material resources synchronized through a decentralized global virtual time. DAMNED also couples local multithreaded processing to the distributed hardware. DAMNED allows to speed up the simulation and to manage wider neural networks than sequential processing. DAMNED is suited to run many models of spiking neurons and networks, and most material supports are workable. Using DAMNED is presented first on simple networks for different sizes, connectivities and activities. Next, DAMNED is applied to modelling the control of saccadic eye movements. Completely based on spiking neurons, the model studies interactions between the neural circuits of the saccadic system located in the brainstem. The model helps validating the hypothesis that the saccade amplitude could be encoded by a vector summation of the activities in the superior colliculus motor map rather than a vector average, from comparison with data obtained in the simulation. Even if further developments and improvements may be forecasted, the originality of the work is to couple event-driven and distributed programming. Moreover, among the parallel simulators for spiking neuron networks, DAMNED is the first one taking advantage of an event-driven strategy internal multithreading of the logic processes and a distributed architecture of physical processes. Hence DAMNED is an advance in the area of simulating spiking neuron networks, mainly for wide size networks. Experiments on simulating the control of saccadic eye movements by a spiking neural network model and their contributions for the neuroscience community confirm the perspectives for further uses of the present work.

## **Publications**

### **Conférences internationales avec comité de lecture et actes**

Simulation of Large Spiking Neural Networks on Distributed Architectures, The "DAMNED" Simulator. A Mouraud, D. Puzenat. Dans Proc. Int. Conf. on Engineering Applications of Neural Networks, EANN'09, London, England. pp. 359-370. Springer . 2009. (à paraître)

### **Conférences nationales avec comité de lecture et actes**

DAMNED, un simulateur parallèle et événementiel, pour réseaux de neurones impulsifs. A Mouraud, H. Paugam-Moisy. Dans Actes NeuroComp 2006, Pont-à-Mousson, France. pp. 120-123. 2006.

# Génération des interfaces adaptatives pour les environnements collaboratifs et nomadiques

Youssef Roummieh

Thèse

Soutenue le 09/12/2009

## Jury

M. KOUKAM ABDER, SeT, MONTBELIARD	Rapporteur
M. TATA SAMIR, TELECOM SUDPARIS, Paris	Rapporteur
Mme JAULENT MARIE-CHRISTINE, INSERM, Paris	Examineur
Mme VERDIER CHRISTINE , LIG, Grenoble	Examineur
M. VANHEMS PHILIPPE , Hopital Edouard Herriot, Lyon	Examineur
Mme GHODOUS PARISA, LIRIS, Lyon	Directeur

Contact : [youssef.roummieh@liris.cnrs.fr](mailto:youssef.roummieh@liris.cnrs.fr)

## Résumé

Le contexte de cette recherche se situe dans la perspective d'une meilleure compréhension de l'ingénierie de l'Interaction Homme-Machine pour les environnements collaboratifs et nomadiques. Les systèmes d'informations d'aujourd'hui sont de plus en plus collaboratifs et nomadiques. Cette transformation impose deux contraintes à prendre en considération pendant la conception d'interfaces utilisateur. D'une part, dans une application collaborative, les acteurs interagissent les uns avec les autres afin de réaliser un projet en commun. Ces acteurs interviennent dans différentes disciplines et chacun d'eux a son propre savoir-faire. Ainsi, ils ne partagent pas les mêmes informations relatives au projet. D'autre part, les utilisateurs sont mobiles et désirent avoir accès au système quel que soit sa localisation géographique, tout en conservant la confidentialité des informations définies dans le système. Cette mobilité requiert l'intégration de terminaux mobiles dans le système d'information. Cependant, pour des raisons de taille et de poids, les terminaux mobiles disposent de ressources moins importantes que celles offertes par des stations fixes. Devant ce constat, notre objectif est de construire une interface pour une application collaborative et mobile, personnalisée et adaptée en fonction du profil de l'utilisateur et des ressources d'interactions du terminal utilisé. En particulier, notre contribution est d'une part, la proposition d'une architecture pour un système coopératif, et d'autre part, l'introduction de nouveaux formalismes, basés sur les environnements collaboratifs, pour la modélisation des tâches. Celle-ci joue un rôle essentiel dans la conception d'interfaces utilisateur. Enfin, nous proposons un cadre à base de modèles, pour la génération automatique d'interfaces utilisateur personnalisées aux acteurs et adaptables aux caractéristiques de plateformes, basé sur quatre niveaux d'abstraction. L'applicabilité de notre proposition est vérifiée par des exemples d'application dans le domaine de la santé. Nous avons développé un prototype de système collaboratif pour le suivi des infections nosocomiales. L'étude de cas menée à l'aide de ce système a été rendue possible grâce à la collaboration de l'hôpital Édouard Herriot à Lyon qui nous a transmis son savoir-faire. Ainsi, nous avons développé l'outil AICEStudio afin de montrer la faisabilité de notre approche pour la génération des interfaces utilisateur personnalisées aux acteurs et adaptables aux caractéristiques de plateformes.

## Publications

### Conférences internationales avec comité de lecture et actes

Adaptive Interface for Collaborative Environment: AICE. Youssef Roummieh, P Ghodous. Dans The Sixth International Conference on Ubiquitous Intelligence and Computing, Brisbane, Australia. 2009.

Adaptive Architecture for Collaborative Environment. Yousse Roummieh, P Ghodous. Dans 16th ISPE International Conference on Concurrent Engineering, Taipei, Taiwan. 2009.

Abstract UserInterface for Collaborative Environments. Yousse Roummieh, P Ghodous. Dans The 2009 International Symposium on Collaborative Technologies and Systems, Baltimore, Maryland, USA. 2009.

Collaborative Environment for the Detection and the Follow-up of Legionella. Yousse Roummieh, P Ghodous, P. Vanehms, C. Verdier, M Lima Dutra, M.C. NICOLLE. Dans Methods and Design models for Health Information Systems, workshops in conjunction with IEEE ICDIM'2007, Lyon, France. pp. 609-614. ISBN 978-1-4244-1475-8. 2007.

#### **Autres conférences**

Outil informatique pour la détection et le suivi du risque de Légionellose nosocomiale. Yousse Roummieh, M.C. NICOLLE, P Ghodous, C. Verdier, P. Vanhems. Dans XIXe congrès de la SFHH et de la SIIHFF, Paris, France. 2008.

# The Chameleon: Un Système de Sécurité pour Utilisateurs Nomades en Environnements Pervasifs et Collaboratifs.

**Rachid Saadi**

**Thèse**

Soutenue le 17/06/2009

à l'INSA de Lyon

## **Jury**

Pr Hacid Mohand-Said, LIRIS, Lyon	Président
Pr Cuppens Frédéric, SERES, Bretagne	Rapporteur
Pr Damiani Ernesto, SESAR, Milan, Italy	Rapporteur
Dr Balbiani Philippe, IRIT, Toulouse	Examineur
Dr Fayolle Jacques, DIOM, Saint-Etienne	Examineur
Dr Serhrouchni Ahmed, NMS, Paris	Examineur
Pr Brunie Lionel, LIRIS, Lyon	Directeur
Pr Pierson Jean-Marc, IRIT, Toulouse	Co-directeur

Contact : rachid.saadi@univ-pau.fr

## **Résumé**

Le partage et la distribution des ressources et des services constituent un défi important de l'informatique moderne. Ainsi, le développement des systèmes distribués vient comme une réponse à des besoins liés à la propagation et à l'acquisition d'informations et de services sans aucune contrainte ni barrière. L'informatique mobile connaît de nos jours un essor fulgurant, en raison de l'apparition d'une nouvelle classe d'utilisateurs à caractère nomade liée à la création de dispositifs portables capables d'offrir de plus en plus de ressources et de services, communiquant grâce à des infrastructures légères et sans fil. Ainsi, la convergence entre les infrastructures terrestres et mobiles permet aux usagers d'avoir à disposition un réseau très vaste d'informations et de services accessibles quels que soient le lieu et le moment. L'avènement de l'informatique mobile a fait évoluer les systèmes distribués, en permettant la mise en place d'environnements intelligents et communicants offrant à l'utilisateur la possibilité de faire interagir ses équipements (téléphone, PDA, etc.) avec son environnement de manière aisée et transparente. Cette évolution a abouti au concept d'Informatique pervasive qui, en plaçant l'utilisateur nomade au coeur de l'informatique moderne, a ouvert de nouvelles perspectives théoriques et applicatives. La mobilité de l'utilisateur et la multiplication des dispositifs légers et autonomes accentuent et rendent plus complexes les problèmes de sécurité. En effet, la mise en place d'environnements pervasifs permet à l'utilisateur nomade de solliciter ou de communiquer avec d'autres utilisateurs, des ressources ou des services au sein d'environnements ou de domaines qui lui sont inconnus et réciproquement. Les mécanismes et les solutions existants sont inadéquats pour pallier les nouveaux challenges issus principalement des problèmes d'authentification, de contrôle d'accès et de protection de la vie privée. Dans un système aussi fortement distribué et imprédictible, l'existence d'une autorité centralisatrice n'est en effet pas possible. Il est donc nécessaire de rendre la décision d'accès plus autonome et donc répartie. Dans cette thèse, nous défendons l'idée que l'implémentation d'un système de sécurité basé sur la confiance constitue une solution particulièrement adaptée aux environnements pervasifs. Notre travail a conduit à l'élaboration d'une architecture générique de sécurité prenant en charge l'utilisateur depuis l'identification jusqu'à l'attribution de droits d'accès dans des milieux inconnus. Nommée « Chameleon », celle-ci prend en charge les problèmes issus de l'authentification et du contrôle d'accès dans les environnements mobiles et ubiquitaires. La mise en place de notre approche a nécessité (i) la définition d'un modèle de confiance inter-organisations (T2D) dont le but est d'étendre le champ d'accès des utilisateurs mobiles aux ressources environnantes; (ii) l'élaboration d'un nouveau format de certificat « X316 »; (iii) l'implémentation d'un nouveau processus de signature (FeMoS) permettant l'adaptation des documents signés au contexte de

l'utilisateur.

## **Abstract**

Users of Information Technologies are somehow contradictory! On one hand, they want their life easy, so they prefer to access transparently to the large set of information, appliances and devices being in their environment. On the other hand, they want to have a clear understanding of what they access and they want to ensure that only the minimum information about themselves is delivered to third parties. They want to be trusted, but are reluctant to trust. In our thesis, we explore this contradiction and try to give way to mechanisms that allow a certain balance reachable between paranoia and naiveness. This goal is large, and we will certainly not address all kind of issues this statement opens. We focus our work on trust management in collaborative distributed environments. They witness a growing interests with the meet of web-enhanced information technologies and wireless devices. The anywhere anytime access to information and services is nowadays a must. All the approaches rely on a kind of trust to establish the necessary regulations for authentication, authorization and access control. While the trust is easy to set up between the known participants of a communication, the evaluation of trust becomes a challenge when confronted with unknown environment. It is more likely to happen that the collaboration in the mobile environment will occur between totally unknown parties. An approach to handle this situation has long been to establish some third parties that certify the identities, roles and/or rights of both participants in a collaboration. In a completely decentralized environment, this option is not sufficient. To decide upon accesses one prefer to rely only on what is presented to him by the other party and by the trust it can establish, directly by knowing the other party or indirectly, and vice-versa. Hence a mobile user must for example present a set of certificates known in advance and the visited site may use these certificates to determine the trust he can have in this user and thus potentially allow an adapted access. In this schema the mobile user must know in advance where she wants to go and what she should present as identifications. This is difficult to achieve in a global environment. Moreover, the user likes to be able to have an evaluation of the site she is visiting to allow limited access to her resources. And finally, an user does not want to bother about the management of her security at fine grain while preserving her privacy. Ideally, the process should be automatized. Our work was lead to define the Chameleon architecture. Thus the nomadic users can behave as chameleons by taking the "colors" of their environments enriching their nomadic accesses. It relies on a new T2D trust model which is characterized by support for the disposition of trust. Each nomadic user is identified by a new morph certification model called X316. The X316 allows to carry out the trust evaluation together with the roles of the participants while allowing to hide some of its elements, preserving the privacy of its users and adapting to the trustfulness of the environment.

## **Publications**

### **Revues internationales avec comité de lecture**

(Dis)trust Certification Model for Large Access in Pervasive Environment. R Saadi, J. Pierson, L. Brunie. Journal of Pervasive Computing and Communications (). 2005.

### **Conférences internationales avec comité de lecture et actes**

Access Pass Certificate, Distrust Certification Model for Large Access in Pervasive Environement. R Saadi, J. Pierson, L. Brunie. Dans IEEE International Conference on Pervasive Services 2005 ICPS'05, IEEE ed. Santorini Greece. pp. 00-00. 2005.

Establishing Trust Beliefs based on a Uniform Disposition to Trust. R Saadi, O. Hasan, J. Pierson, L. Brunie. Dans IEEE/ACM SITIS 2007, IEEE Computer Society Press ed. Shanghai, China. 2007.

Context Adapted Certificate Using Morph Template Signature for Pervasive Environments.. R Saadi, J. Pierson, L. Brunie. Dans Fourth International Symposium on Ubiquitous Computing Systems (UCS 2007), Springer's Lecture Notes in Computer Science (LNCS) series ed. Akihabara Convention Hall, Tokyo, Japan. pp. 17-31. 2007. (à paraître)

X316 Security Toolbox for New Generation of Certificate. R Saadi, J. Pierson, L. Brunie. Dans 4th International Conference on Trust, Privacy & Security in Digital Business, Springer LNCS serie ed. Regensburg, Germany. pp. 248-258. 2007.

A New Certificate Signature for Pervasive Environment. R Saadi, J. Pierson, L. Brunie. Dans 3rd International

Workshop on Security, Privacy and Trust in Pervasive and Ubiquitous Computing , IEEE Computer Society Press ed. Istanbul, Turkey. pp. 43-48. ISBN 0-7695-2863-5. 2007.

Morph Digital Signature A New Signature Framework for e-Documents in Pervasive Environments. R Saadi, J. Pierson, L. Brunie. Dans ICPS'07 : IEEE International Conference on Pervasive Services, IEEE Computer Society Press ed. Istanbul, Turkey. 2007.

The Chameleon: A Pervasive Grid Security Architecture. R Saadi, J. Pierson, L. Brunie. Dans ICNS 2007 The Third International Conference on Networking and Services, IEEE computer society ed. Athens Greece. 2007.

Authentication and Access Control Using Trust Collaboration in Pervasive Grid Environment.. R Saadi, J. Pierson, L. Brunie. Dans GPC 2007 Grid and Pervasive Computing, Springer Verlag ed. Paris France. pp. 348-361. Lecture Notes in Computer Science (LNCS) . ISSN 0302-9743. 2007.

#### **Conférences nationales avec comité de lecture et actes**

Authentification par la méfiance dans les systèmes pervasifs. R Saadi, J. Pierson, L. Brunie. Dans UbiMob05 Deuxièmes Journées Francophones: Mobilité et Ubiquité 2005, ACM ed. . pp. 00-00. 2005.

#### **Rapports de recherche/technique**

X316 Morph Access Pass: A New Extensible Model of Certification. R Saadi, L. Brunie. Rapport de recherche RR-LIRIS-2006-022 , Soumis à IEEE Symposium on Security and Privacy 2007. 2006.



# RAevol : un modèle de génétique digital pour étudier l'évolutin des réseaux de régulation génétique

Yolanda Sanchez-Dehesa

Thèse

Soutenue le 19/11/2009

à l'INSA de Lyon

## Jury

Florence d'Alché-Buc  
Endika Bengoetxea-Castro  
Guillaume Beslon  
Jean-François Boulicaut  
José-María Peña  
Jean-Daniel Zucker

rapporteur  
examineur  
directeur de thèse  
examineur  
examineur  
rapporteur

Contact : yolandasdf@gmail.com

## Résumé

Le réseau de régulation génique est un élément central de toute cellule vivante capable de réguler la production de protéines en fonction des besoins et/ou des conditions extérieures. De ce fait, la mesure, l'analyse et, in fine, la compréhension des mécanismes de régulation sont au coeur de la biologie contemporaine et, plus particulièrement, de la biologie des systèmes. Or, lorsqu'on étudie parallèlement plusieurs réseaux de régulation correspondant à des organismes différents, on constate que ces réseaux possèdent des caractéristiques structurelles communes. Ainsi, chez les organismes bactériens, il a été montré que la topologie globale du réseau suivait une topologie scale-free, que les réseaux présentaient une répartition en motifs spécifiques ou que la distribution du nombre de noeuds du réseau (les Facteurs de Transcription) suivait une loi de puissance d'exposant plus élevé que la distribution du nombre de gènes dans les mêmes organismes. L'origine de ces structures reste encore mal connue, c'est pourquoi une approche de modélisation semble être appropriée pour émettre des hypothèses quant au rôle joué par les différentes pressions évolutives (mutation, sélection, dérive, ...). Dans ce but, nous avons développé RAEvol, un modèle de génétique digitale dédié à l'étude de l'évolution des réseaux de régulation. RAEvol est un modèle intégré, au sens où un organisme n'inclut pas seulement un réseau de régulation mais aussi un génome et un phénotype. En réunissant une population de tels individus, en introduisant un mécanisme mutationnel au niveau des génomes et un mécanisme sélectif au niveau des phénotypes, nous pouvons alors d'étudier l'évolution de ces réseaux dans un contexte cohérent. Nous présentons ici le cadre général de notre étude, les réseaux de régulation procaryotes et leur structure, ainsi que les choix de modélisation qui nous ont conduit à développer RAEvol. Le modèle lui-même sera décrit de façon exhaustive et une première série d'expérimentation sera décrite en détail pour montrer son potentiel. Nous présenterons enfin deux travaux utilisant notre modèle et publiés dans des revues ou conférences internationales. Ces travaux illustrent la capacité du modèle à produire des données puis des hypothèses pertinentes pour la biologie évolutive et la biologie des systèmes.

## Abstract

Regulation and regulation networks are core elements of any living cell. This network enables the cell to regulate its protein production depending on its needs. Yet, measure, analysis and comprehension of regulation mechanisms are at the heart of contemporary biology and, particularly, of systems biology. Now, when studying different regulation networks coming from different species, some regularities emerge and networks appear to share some structural properties. In prokaryotic organisms, it has been shown that regulation networks are scale-free, that their motif distribution is not random or that the number of regulation nodes in the networks of different species follows a power-law which exponent is larger than the exponent of genes numbers distribution in the same organisms. To explain such regularities and to

understand how evolution has structured these networks, we designed a digital genetic model, RÆvol, which is dedicated to the study of the evolutionary history of regulation networks. RÆvol enables us to study how the different evolutive pressures (mutation, selection, drift...) act on the network and which one(s) gives the networks their universal features. Moreover, RÆvol is an integrated model, meaning that, in RÆvol, apart from their regulation networks, organisms own a genome and a phenotype. By creating of population of such organisms, by introducing mutations at the genetic level and selection at the phenotypic level, we are able to study the evolution of regulation networks in a coherent context. Here, we present the general context of our study (prokaryotic regulation networks and their structure) together with the main principle of our model. RÆvol is then fully described and our first experiments are presented, showing how the model can be used. Finally, two complementary appendix present two publications, illustrating that RÆvol is able to produce valuable data and hypothesis for evolutionary biology and systems biology.

## **Publications**

### **Revue internationale avec comité de lecture**

Scaling Laws in Bacterial Genomes: A Side-Effect of Selection of Mutational Robustness. G. Beslon, D.P. Parsons, Y. Sanchez-Dehesa, J.M. Pena, C. Knibbe. *BioSystems* (). 2010. (à paraître)

From Digital Genetics to Knowledge Discovery: Perspectives in Genetic Network Understanding. G. Beslon, D.P. Parsons, J.M. Pena, C. Rigotti, Y. Sanchez-Dehesa. *Intelligent Data Analysis Journal* (). 2010. (à paraître)

Modelling Evolution of Regulatory Networks in Artificial Bacteria. Y. Sanchez-Dehesa, D.P. Parsons, J.M. Pena, G. Beslon. *Mathematical Modelling of Natural Phenomena* 3(2):27-66, ISSN 0973-5348. 2008.

### **Conférences internationales avec comité de lecture et actes**

Modeling evolution of genetic regulation in artificial bacteria. G. Beslon, Y. Sanchez-Dehesa, J.M. Pena. (European Conference on Complex Systems), , Jérusalem. 4p. 2008.

Artificial Regulatory Networks Evolution. Y. Sanchez-Dehesa, L. Cerf, J.M. Pena, J-F. Boulicaut, G. Beslon. Dans Proc 1st Int Workshop on Machine Learning for Systems Biology MLSB 07 , Evry, France. pp. 47-52. 2007.

Scaling Laws in Digital Organisms. G. Beslon, Y. Sanchez-Dehesa, D.P. Parsons, J.M. Pena, C. Knibbe. Dans Proc. Information Processing in Cells and Tissues IPCAT'09, Ascona, Switzerland. pp. 111-114. 2009.

### **Conférences nationales avec comité de lecture et actes**

RAevol, un modèle de génétique digitale des réseaux de régulation. Y. Sanchez-Dehesa, J.M. Pena, G. Beslon. Dans Réseaux d'Interactions : Analyse, Modélisation, Simulation (RIAMS), Lyon. 2007.

Evolution of scaling laws in artificial regulation networks. D.P. Parsons, G. Beslon, C. Knibbe, Y. Sanchez-Dehesa, J.M. Pena. Dans Integrative Post-Genomics, Lyon. pp. 22-22. 2009.

# Le système ICOP : représentation, visualisation et communication de l'information à partir d'une représentation iconique des données

Salma Sassi

Thèse

Soutenue le 08/10/2009

à l'INSA de Lyon et ENSI de Manouba (thèse en cotutelle)

## Jury

Pr. Ben Ahmed Mohamed , RIADI -ENSI , Manouba, Tunisie	Co-directeur
Pr. Cauvet Corine , Université Aix Marseille, Marseille	Rapporteur
Pr. Claude-Yves Chrisment, Irit, Toulouse	Rapporteur
Pr. Flory andré , LIRIS -INSA de lyon , Lyon	Co-directeur
Pr. Rubel Paul , INSA de Lyon , Lyon	Président
Pr. Verdier Christine , IMAG- Université Joseph Fourier , Grenoble	Co-directeur

Contact : salma.sassi@insa-lyon.fr

## Résumé

Les systèmes d'informations connaissent un essor constant depuis leur création. Leur usage laisse de nouvelles possibilités d'accès et de traitement des informations, en appui à l'action. Or, le constat général est toujours le même : accumulation de données non décrites, non référencées, sans origine, difficultés d'accès à des données mises à jour... On ne sait plus qui fait quoi ni où se trouve l'information que l'on cherche. Le temps et le nombre de personnes intermédiaires, nécessaires à la recherche d'une donnée sont un frein à la circulation de l'information. Ceci se vérifie dans tous les domaines, y compris le domaine médical. La mise en partage d'une information communiquée et temporalisée constitue un enjeu primordial. Notre étude des systèmes existants a permis de révéler trois limites principales qui accroissent la dispersion des informations d'un même domaine. L'un des tout premiers freins réside dans le fait que les interfaces existantes ne correspondent ni à l'attente ni au fonctionnement des utilisateurs. Le deuxième problème réside dans le fait que certains systèmes ne sont pas communicants ce qui rend impossible une vue globale des informations reliées à un projet bien déterminé et finalement le troisième problème concerne l'accès à l'information qui nécessite l'accès à diverses sources d'informations. Ces dernières sont généralement hétérogènes, que ce soit au niveau syntaxique ou sémantique. Des ontologies complexes contenant des milliers de termes sont créées pour résoudre les conflits sémantiques alors que le problème de la syntaxe et de la structure unique des données restent un problème difficile à résoudre. Notre contribution consiste principalement à faire coopérer les systèmes d'information hétérogènes. Nous proposons pour cela une architecture de médiation sémantique. Des méta-ontologies de domaine et de tâches y sont associées afin d'assurer une convergence des sources d'informations. Nous avons recours aux annotations et aux métadonnées qui facilitent la description de ces sources, afin d'établir des correspondances entre elles, résoudre des conflits et exploiter ensuite les données elles-mêmes. Le deuxième volet de notre contribution concerne un nouvel outil de visualisation graphique et chronologique. Ce système permet de tracer sur une ligne temporelle les informations relatives à un domaine pour n'apparaître à l'utilisateur dont il a besoin et ce qu'il est autorisé à consulter. On développe ces thèmes en les illustrant sur un terrain d'application qui présente de nombreux facteurs de complexité : les systèmes d'information de santé. Nos propositions ont été validées par le développement de deux prototypes : le prototype OR (Object Reconstruction) et le prototype TravelIn qui soulignent la faisabilité de notre approche.

## Abstract

Information systems are in continuous development since their creation. Using these systems offers the possibility to information access and treatment. Although, the general operation is always the same: gathering non-described, unreferenced, and unoriginal data, as well as the difficulty in accessing updated data... We do not know where the searched information is, or who created it. The time and the number of intermediate persons

that are necessary for data search, reduce the circulation of information. This is true in all domains including the medical domain. The communication of shared and temporal information remains an important problem. Our study on current systems showed three main limitations that increase the information dispersal on the same domain. The first problem is that the current interfaces do not correspond to user needs and work. The second problem is that some of information systems do not communicate. This makes impossible to generate an overall view of the information which are connected to the same project. Finally the third problem concerns the information access that requires the access to diverse resources. These last ones are generally heterogeneous within the syntax or semantic level. Complex ontologies containing thousands of terms are created to resolve the semantic conflicts. Nevertheless, the syntax and the unique data structure remain a difficult problem to be resolved. Essentially, our contribution consists in cooperating heterogeneous information systems. For this reason, we propose semantic mediation architecture. Domain meta-ontology and task metaontology are associated to assure the information sources convergence. We also use annotations and metadata that facilitate the information resources description in order to make correspondences between them, to resolve conflicts and finally to exploit the data themselves. The second part of our contribution concerns a new tool of graphic and chronological visualization. This system allows to represent on a temporal component the information related to a given domain, and also to show the needed and the authorized information to the user. We develop these proposals by illustrating them in an application domain that presents many complexity factors: medical information systems. Our proposals were validated throughout two prototypes development: the OR (Object Reconstruction) prototype and the Travel'In prototype.

## **Publications**

### **Reuves internationales avec comité de lecture**

Collaborative tasks: reorganize the information representation and communication in the project management. S Sassi, A. Flory, C. Verdier. International Electronic Journal of Digital Enterprise (). 2009. (à paraître)

### **Reuves nationales avec comité de lecture**

un médiateur sémantique pour une représentation unifiée et chronologique du dossier médical.. S Sassi. Santé et systématique 10(1-2/2):105-136, Hermes-Lavoisier, ISBN 978-2-7462-1912, ISSN 10.1-2.105-136. 2007.

### **Conférences internationales avec comité de lecture et actes**

Approche sémantique multi-niveaux pour la représentation et le partage des informations dans la gestion de projet. S Sassi, C. Verdier, A. Flory. Dans International Conference on Information Systems and Economic Intelligence (SIIE'2009), Hammamet, Tunisia.. 2009.

### **Conférences nationales avec comité de lecture et actes**

Proposition d'un méta-modèle pour une structuration collective et partagée de la mémoire de l'entreprise. S Sassi. Dans Conférence Internationale sur les "Systèmes d'Information et Intelligence Economique ", Hammamet-Tunisie.. 2008.

nouvelles Interfaces pour la représentation de l'information médicale. A. Flory, C. Verdier, S Sassi. Dans INFORSID, Tunisie Hammamet. 2006.

### **Conférences invité**

A new system for project management: the I2MS interface. S Sassi, C. Verdier, A. Flory. Dans Dans TIGERA, Tunisie Hammamet. 2007.

### **Autres conférences**

le médiateur sémantique SGI2: Système de Gestion des Informations Iconisées. S Sassi. INFORSID 2007, Perros Guirec, La Bretagne 2007.

Life Line Management System for Medical Records. C. Verdier, S Sassi, A. Flory. Dans MedInfo, Australie 2007.

# Méthodologie de matching à large échelle pour des schémas XML

Sana Sellami

Thèse

Soutenue le 30/11/2009

à l'INSA de Lyon

## Jury

Dr Benharkat Aïcha-Nabila, LIRIS, Lyon

Pr Amghar Youssef, LIRIS, Lyon

Pr Defude Bruno, TELECOM, Paris

Dr Morvan Franck, IRIT, toulouse

Pr Verdier Christine, LIG, grenoble

Pr Akoka Jacky, Cédric, Paris

Pr Boussaid Omar, ERIC, Lyon

Co-directeur

Directeur

Rapporteur

Rapporteur

Examineur

Examineur

Examineur

Contact : sana.sellami@insa-lyon.fr

## Résumé

De l'intégration des schémas de bases de données jusqu'à l'alignement d'ontologies, la problématique qui a suscité le plus de points ardus à résoudre cette dernière décennie est la recherche des correspondances (entre schémas de bases de données, XML ou ontologies). Mais aujourd'hui, il existe des centaines voire même des milliers de schémas disponibles sur le Web qui nécessitent d'être rassemblés, organisés et mis en correspondance. Cette profusion des données affecte aussi bien les processus de gestion des données que de mapping et de matching des données. Nous nous intéressons plus particulièrement au matching, qui est un processus qui vise à découvrir les correspondances sémantiques entre différents formats de données tels que les schémas, les ontologies, interfaces Web,... Plusieurs travaux de matching ont été proposés comme solution à la problématique de l'intégration des données. La plupart de ces travaux se sont focalisés sur le matching des schémas simples et de petite dimension (50-100 éléments). Cependant, dès que l'on passe à un contexte à large échelle, plusieurs problèmes se posent tels que des problèmes de performance en termes de temps d'exécution et de qualité des résultats. Les schémas e-Business par exemple, sont des schémas répartis sur plusieurs fichiers et leur taille varie d'une centaine à un millier d'éléments. L'objectif de notre travail est de relever le challenge du matching pour de tels schémas. En particulier, nous proposons une méthodologie de matching à large échelle, basée sur une approche hybride et structurée en trois phases (pré-matching, matching et post-matching), qui vise à optimiser le matching en s'appuyant notamment sur une phase préalable de prétraitement. Ce prétraitement se base sur des techniques d'analyse, de traitement linguistique des éléments des schémas et une approche de décomposition des schémas. La décomposition de schémas que nous proposons est une approche holistique qui consiste à diviser les schémas en sous schémas tout en identifiant ceux qui sont linguistiquement similaires. L'utilisation d'une telle approche nécessite des techniques qui passent à l'échelle et qui permettent un traitement d'un grand nombre de données en une seule fois. Pour cela, nous avons choisi d'utiliser une technique d'extraction d'arbres fréquents. La méthodologie proposée est supportée par une plateforme nommée PLASMA (*\_P\_latform for \_LA\_rge \_S\_chema \_MA\_tching*) que nous avons développée pour des besoins d'évaluation et d'expérimentations. Pour réaliser ces évaluations, nous avons défini une méthode d'évaluation et nous l'avons appliquée à l'évaluation de PLASMA. Nous avons pu ainsi démontrer grâce aux expérimentations réalisées que notre système offre des résultats fiables pour des schémas très volumineux et nombreux, et que grâce à la méthodologie mise au point, nous avons réussi à améliorer les performances du matching en temps d'exécution.

## Publications

Revues internationales avec comité de lecture

Towards a more scalable schema matching: a novel approach. S Sellami, A. Benharkat, Y. Amghar. International Journal of Distributed Systems and Technologies (IJ DST) (.). 2010. (à paraître)

Quality of large scale matching in E-Business domain. S Sellami, A. Benharkat, Y. Amghar. International Electronic Journal of Digital Enterprise . (.). 2009. (à paraître)

PLASMA: A PLATform for Schema MATching and Management. A. Benharkat, R. Rifaieh, S Sellami, M Boukhebouze, Y. Amghar. IBIS-International Journal of Interoperability in Buisness Information Systems 2(2):9-20, ISSN 1862-6378. 2007.

#### **Conférences internationales avec comité de lecture et actes**

Vers une plateforme de gestion des schémas XML. S Sellami, A. Benharkat, R. Rifaieh, Y. Amghar. Dans DECOR en conjonction avec EGC 2007, Namur Belgique. pp. 28-43. 2007.

Schema Matching for Document Exchange: A Constraint Based Approach. S Sellami, A. Benharkat, R. Rifaieh, Y. Amghar. Dans THE INTERNATIONAL CONFERENCE ON SIGNAL-IMAGE TECHNOLOGY & INTERNET&#8211;BASED SYSTEMS (SITIS' 2006), Hammamet TUNISIE. pp. 299-309. springer Verlag, LNCS series . ISSN 2-9525435-1. 2006.

Pre-matching: Large XML schemas decomposition approach. S Sellami, A. Benharkat, Y. Amghar. Dans 8th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE 2009), Algarve, Portugal. pp. 9-10. Springer . 2009.

Quality of Matching in large scale scenarios. S Sellami, A. Benharkat, Y. Amghar. Dans International Conference on Information Systems and Economic Intelligence (SIIE'2009), Hammamet, Tunisia. 2009.

From simple to large scale matching: a hybrid approach. S Sellami, A. Benharkat, Y. Amghar. Dans 23rd International Symposium on Computer and Information Sciences ( ISCIS 2008 ), Istanbul, Turkey. pp. 1-4. ISBN 978-1-4244-2880-9. 2008.

#### **Conférences nationales avec comité de lecture et actes**

Quelle vision pour le matching à large échelle ?. S Sellami, A. Benharkat, Y. Amghar. Dans 10th Maghrebian Conference on Information Technologies (MCSEAI'08), Algérie-Oran. pp. 356-361. 2008.

Méthodologie de Matching à large Echelle de schémas XML. S Sellami, A. Benharkat, Y. Amghar. Dans 27ème congrès INFORSID, Toulouse. pp. 277-292. 2009.

#### **Contributions à un ouvrage**

Large Scale matching issues and advances. S Sellami, A. Benharkat, Y. Amghar. Ontology Theory, Management and Design: Advanced Tools and Models 2009. (à paraître)

# Un canevas pour l'adaptation et la substitution de services Web

Yehia Taher

Thèse

Soutenue le 24/07/2009

à l'Université Claude Bernard Lyon 1

## Jury

M Aït-Ameur Yamine, LISI, Poitiers

M Léonard Michel, MATIS, Genève, Suisse

M Boissier Olivier, G2I, Saint-Etienne

M Boughanem Mohand, IRIT, Toulouse

M Benslimane Djamel, LIRIS, Lyon

Mme Fauvet Marie-Christine, LIG, Grenoble

Rapporteur

Rapporteur

Examineur

Examineur

Directeur

Co-directeur

Contact : yehia.taher@liris.cnrs.fr

## Résumé

Les services Web ont émergé comme un support de développement et d'intégration d'applications ou de systèmes d'information. Dans ce cadre, les interactions entre deux applications consommateur et fournisseur, encapsulées par des services Web se font par échanges de messages. Ces échanges s'appuient sur la notion d'interfaces, qui décrivent les interactions dans lesquelles un service peut s'engager et les dépendances entre ces interactions. Dans le Web actuel, il arrive très fréquemment que de nombreux services répondent à un même ensemble de besoins fonctionnels. Ces services sont souvent offerts par le biais d'interfaces différentes. Des nombreuses raisons, telles que la panne du service fournisseur, peuvent amener un consommateur à substituer son fournisseur habituel par un autre fournisseur qui offre la même fonctionnalité. Cette substitution provoque des incompatibilités entre l'interface du service consommateur et celle du service fournisseur substitut. Cela est dû au fait que le service consommateur n'a pas été fait en fonction de ce nouveau service fournisseur. Les recherches que nous menons dans cette thèse visent à résoudre le problème des incompatibilités dans des interactions entre deux services consommateur et fournisseur substitut. En particulier, notre contribution s'étend, tant sur le plan théorique que sur le plan pratique. Il s'agit d'une part d'un canevas pour la génération automatique des adaptateurs des interactions entre deux services. D'autre part, nous proposons une architecture logicielle multicouche fournissant un cadre permettant une substitution transparente et flexible d'un service fournisseur par un autre service vis-à-vis du consommateur du premier. Dans notre canevas pour la génération automatique des adaptateurs, une modélisation des interfaces de services en des automates a été adoptée. Puis, une étape de détection des incompatibilités entre ceux-ci est réalisée. Ensuite, un adaptateur des interactions entre les deux services est généré automatiquement sur la base d'incompatibilités détectées. La génération de l'adaptateur est guidée par le modèle d'automates. Cela permet de modéliser l'adaptateur indépendamment de son implémentation cible, permettant ainsi une bonne réutilisation des modèles. Une fois généré, l'automate de l'adaptateur suffisamment détaillé est projeté sur la technologie CEP (Complexe Event Processing). Cette projection est réalisée à l'aide des composants cartouches (en anglais : Templates) que nous avons mis en œuvre. Chaque cartouche étant conçue pour générer du code exécutable (en termes de requêtes continues) pour la technologie CEP. Notre architecture proposée pour la substitution de services Web s'intègre d'une part la notion de communauté de services, et d'autre part un progiciel nommé OSC (Open Service Connectivity). Une communauté de service est perçue comme un moyen d'exposer des descriptions communes d'un caractère fonctionnel désiré sans explicitement se référer à un service spécifique. L'OSC est un composant logiciel dont l'objectif est d'appliquer le principe de pilotes ODBC et JDBC dans un environnement à base de services Web. Plus précisément, il est responsable de gérer les interactions entre les consommateurs des communautés d'une part et les communautés d'autre part, en fournissant des fonctions permettant la sélection et la substitution de services. Mots clés : Services

Web, interface, substitution, incompatibilités structurelles et comportementales, détection, résolution, adaptateurs, automate, CEP, communauté de services, OSC.

## **Abstract**

Web services have emerged as a support for development and integration of applications and information systems. In this context, the interactions between two consumer and supplier applications, encapsulated by Web services are done by exchanging messages. These exchanges are based on the concept of interfaces, which describe the interactions in which a service can handle and dependencies between these interactions. In the current Web, it is very often that many services meet the same set of functional requirements. These services are often delivered through different interfaces. For many reasons, such as failure of the service provider, the consumer has to replace his usual supplier with another supplier that offers the same functionality. This substitution leads to incompatibilities between the interfaces of the service consumer and new service provider. This is because the customer service was not been developed according to the new service provider. The research we conduct in this thesis aims to solve the problem of incompatibility in the interaction between two services; consumer and new provider. In particular, our contribution extends both in theory and in practice. It is a part of a framework for the automatic generation of adapters for interactions between two services. On the other hand, we propose a multi-layer software architecture providing a framework for transparent and flexible substitution of a service provider by another with respect to an existed consumer. In our framework for automatic generation of adapters, services interfaces modeling using automata has been adopted. Then, a step of detecting incompatibilities between them is achieved. After that, an adapter of the interactions between the two services is generated automatically based on the detected incompatibilities. The generation of the adapter is based on the automata model. The generated adapter automaton contains a sufficient detail for the projected technology CEP (Complex Event Processing). This projection is performed using components templates that are implemented by us. Each template is designed to generate executable code (in terms of continuous requests) for the CEP technology. Our proposed architecture for the substitution of integrated Web services is an approach for deploying communities of Web services. A community promotes the dynamic binding of Web services through software named OSC (Open Service Connectivity). A community service is seen as a way to expose common descriptions of a desired functionality without explicitly referring to a specific service. OSC is a software component whose objective is to apply the principle of ODBC and JDBC in an environment based on Web services. Specifically, it is responsible for managing interactions between communities of consumers on the one hand and communities on the other hand, providing functions for the selection and substitution of services. Keywords: Web, interface, substitution, structural and behavioral incompatibilities, detection, resolution, adapters, automata, CEP, community, OSC.

## **Publications**

### **Conférences internationales avec comité de lecture et actes**

Diagnosing Incompatibilities in Web service Interactions for Automatic Generation of Adapters. Y. Taher, A. Aït-Bachir, M.C. Fauvet, D. Benslimane. AINA '09: Proceedings of the 23st International Conference on Advanced Networking and Applications, Bradford, UK. IEEE Computer Society Los Alamitos, CA, USA. 2009.

Using CEP Technology to Adapt Messages Exchanged by Web Services. Y. Taher, M.C. Fauvet, M. Dumas, D. Benslimane. WWW '08: Proceeding of the 17th international conference on World Wide Web, Beijing, China. pp. 1231-1232. ACM New York, NY, USA. ISBN 978-1-60558-085-2. 2008.

A Multi-Layer and Multi-Perspective Approach to Compose Web Services. D. Benslimane, Z. Maamar, Y. Taher, M. Lahkim, M.C. Fauvet, M. Mrissa. AINA '07: Proceedings of the 21st International Conference on Advanced Networking and Applications, Niagara Falls, Canada. pp. 31-37. IEEE Computer Society Los Alamitos, CA, USA. ISBN 0-7695-2846-5. ISSN 1550-445X. 2007.

Towards an Approach for Web services Substitution. Y. Taher, D. Benslimane, M.C. Fauvet, Z. Maamar. IDEAS '06: Proceedings of the 10th International Database Engineering and Applications Symposium, Delhi, India. pp. 166-173. IEEE Computer Society Washington, DC, USA.

# Gestion des connaissances pour la maîtrise de la relation entre patrimoine documentaire et Système d'Information

**Ovidiu Vasutiu**

**Thèse**

Soutenue le 3/12/2009

à l'INSA de Lyon

## **Jury**

Youssef Amghar, INSA de Lyon

Jacques Faveeuw, Caisse Nationale des Allocations Familiales

Eric Gaussier, Université J. Fourier Grenoble

David Jouve, Caisse Nationale des Allocations Familiales

Pierre Maret, Université de Saint Etienne

Jean-Marie Pinon, INSA de Lyon

Ioan Roxin, Université de Franche-Comté

## **Résumé**

L'activité des organisations qui ont pour mission la mise en application de la législation est fondée sur l'analyse et la manipulation des textes réglementaires. Par ailleurs, au vu de la quantité massive d'information traitée au sein de ces organisations, le travail dépasse très largement les capacités humaines, les amenant à solliciter l'assistance du Système d'Information. Une des principales caractéristiques des textes réglementaires est que ceux-ci sont en perpétuelle évolution. Afin d'assurer la cohérence et la conformité des ressources de l'organisation, leur mise en oeuvre au sein du Système d'Information doit être réexaminée à chacune de ces évolutions. Pour cela nous proposons un modèle commun pour la mise en relation des textes réglementaires et des composantes du système d'information. Ce modèle commun est ensuite enrichi des mécanismes d'analyse des dépendances et d'étude d'impact permettant d'identifier et d'évaluer les conséquences d'une modification réglementaire sur les textes réglementaires et sur le système d'information.

Ces problématiques, issues du contexte de la Caisse Nationale des Allocations Familiales (Cnaf), n'ont toutefois pas été traitées uniquement en fonction du contexte spécifique de travail, mais bien en les considérant comme révélateurs d'une problématique plus large concernant de manière générale la gestion documentaire et la gestion des Systèmes d'Information. Ainsi ces propositions ont été validées lors de leur intégration dans le Système d'Information Documentaire de la Cnaf mais leur champ d'application peut être élargi à d'autres domaines.

## **Publications**

### **Conférences nationales avec comité de lecture et actes**

Gestion des changements et étude d'impact dans un système d'information réglementaire. V.O Vasutiu, D. Jouve, Y. Amghar. Dans INFORSID'2006 - Informatique des ORganisations et Systèmes d'Information et de Décision 2006, Inforsid ed. Hammamet, Tunisie, du 31 mai au 3 juin 2006. pp. 1007-1022. Actes du XXIVème Congrès . Association Inforsid Les Bastidonnnes - 13122 Ventabren. ISBN 2-906855-22-7. 2006.

### **Edition scientifique d'ouvrages**

XML based Legal Document Drafting Environement. V.O Vasutiu, D. Jouve, Y. Amghar, J.M. Pinon. (20th Aniversary Annual JURIX Conference, Workshop on Legislative XML), , Leiden, Netherlands. 4p. Foundation for Legal Knowledge Based Systems . 2007.



Soutenue le 05/01/2009

à l'INSA de Lyon

### Jury

KOSCH Harald, Université de Passau,  
RAVAT Franck, Université de Toulouse  
DARMONT Jérôme, Université Lyon 2,  
KOSCH Harald, Université de Passau,  
MIQUEL Marvonne, INSA de Lyon,  
MELAB Nouredine, Université Lille 1,  
RAVAT Franck, Université de Toulouse,  
TCHOUNIKINE Anne, INSA de Lyon

Rapporteurs  
Rapporteurs

### Résumé

Les entrepôts de données et les systèmes OLAP (OnLine Analytical Processing) permettent un accès rapide et synthétique à de gros volumes de données à des fins d'analyse. Afin d'améliorer encore les performances des systèmes décisionnels, une solution consiste en la mise en oeuvre d'entrepôts de données sur des systèmes répartis toujours plus puissants. Les grilles de calcul en particulier offrent d'importantes ressources de stockage et de traitement. Le déploiement d'un entrepôt sur une infrastructure décentralisée de grille nécessite cependant l'adaptation du modèle multidimensionnel et des processus OLAP pour tenir compte de la répartition et de la réplication des données et de leurs agrégats. Nous introduisons un modèle d'identification des données de l'entrepôt réparti et une méthode d'indexation des données sous forme de blocs multidimensionnels. Cette structure d'index s'appuie sur des index spatiaux en X-tree et des treillis de cuboïdes, et permet la localisation des données matérialisées ainsi que des agrégats calculables sur les différents noeuds de la grille. Nous proposons une méthode de traitement de requêtes OLAP visant à construire un plan d'exécution optimisé à partir de la liste des blocs candidats contribuant au résultat de la requête. Enfin, nous définissons une architecture de services de grille GIROLAP (Grid Infrastructure for Relational OLAP), intégrée à l'intergiciel Globus, et déployée dans le cadre du projet GGM (Grille Géno-Médicales) de l'ACI « Masse de Données ».

### Publications

#### Revue internationale avec comité de lecture

GGM Efficient Navigation and Mining in Distributed Geno-Medical Data. J. Pierson, J. Gossa, P. Wehrle, Y. Cardenas, E.S. Mahmoud, C. Sebastien, L. Brunie, C. Dhaenens, A. Hameurlain, N. Melab, M. Miquel, F. Morvan, T. El Gazali, A. Tchounikine. IEEE Transactions on NanoBioscience 6(2):110-116, IEEE, ISSN 1536-1241. 2007.

#### Conférences internationales avec comité de lecture et actes

Grid Services for Efficient Decentralized Indexation and Query Execution on Distributed Data Warehouses. P. Wehrle, A. Tchounikine, M. Miquel. Dans Forum Proceedings of the 19th International Conference on Advanced Information Systems Engineering (CAISE'07), J. Eder, S. L. Tomassen, A. L. Opdahl, and G. Sindre ed. Trondheim, Norway. pp. 13-16. ISSN 1503-416X. 2007.

A Grid Services-Oriented Architecture for Efficient Operation of Distributed Data Warehouses on Globus. P. Wehrle, M. Miquel, A. Tchounikine. Dans The IEEE 21st International Conference on Advanced Information Networking and Applications (AINA-07), Niagara Falls, Ontario, Canada. pp. 994-999. IEEE Computer Society . ISBN 0-7695-2846-5. ISSN 1550-455X. 2007.

GeWolap: A Web Based Spatial OLAP Proposal. S Bimonte, P. Wehrle, A. Tchounikine, M. Miquel. Dans Second International Workshop on Semantic-based Geographical Information Systems, Robert Meersman, Zahir Tari, Pilar Herrero et al ed. Montpellier, France. pp. 1596-1605. Springer-Verlag . 2006.

A Model for Distributing and Querying a Data Warehouse on a Computing Grid. P. Wehrle, M. Miquel, A. Tchounikine. Dans 11th International Conference on Parallel and Distributed Systems, ICPADS 2005, IEEE Computer Society ed. Fukuoka, Japan. pp. 203-209. ISBN 0-7695-2281-5. ISSN 1521-9097. 2005.

#### **Autres conférences**

Entrepôts de données sur grilles de calcul. P. Wehrle, M. Miquel, A. Tchounikine. Dans Premier atelier sur Extraction et Gestion Parallèles Distribuées de Connaissances (EGPDC), 5èmes Journées d'Extraction et de Gestion des Connaissances (EGC'05), Paris. 2005.

#### **Contributions à un ouvrage**

Entrepôts de données sur grilles de calcul. P. Wehrle, M. Miquel, A. Tchounikine. Revue des Nouvelles Technologies de l'Information, RNTI-E-5 "Extraction des connaissances : Etat et perspectives", Editions Cépaduès, pp. 195-198, ISBN 2.85428.707.X, ISSN 1764-1667 2005.

# Département Image



# Optimisation multi-objectifs par colonies de fourmis : cas des problèmes de sac à dos

Ines Alaya

Thèse

Soutenue le 05/05/2009

à l'Université Claude Bernard Lyon 1 et ENSI Tunis

## Résumé

Dans cette thèse, nous nous intéressons à l'étude des capacités de la métaheuristique d'optimisation par colonie de fourmis (Ant Colony Optimization - ACO) pour résoudre des problèmes d'optimisation combinatoires multi-objectifs. Dans ce cadre, nous avons proposé une taxonomie des algorithmes ACO proposés dans la littérature pour résoudre des problèmes de ce type. Nous avons mené, par la suite, une étude expérimentale de différentes stratégies phéromonales pour le cas du problème du sac à dos multidimensionnel mono-objectif. Enfin, nous avons proposé un algorithme ACO générique pour résoudre des problèmes d'optimisation multi-objectifs. Cet algorithme est paramétré par le nombre de colonies de fourmis et le nombre de structures de phéromone considérées. Il permet de tester et de comparer, dans un même cadre, plusieurs approches. Nous avons proposé six variantes de cet algorithme dont trois présentent de nouvelles approches et trois autres reprennent des approches existantes. Nous avons appliqué et comparé ces variantes au problème du sac à dos multidimensionnel multi-objectifs. Après une étude comparative des différentes variantes suivant différentes métriques de performance, nous avons trouvé qu'une nouvelle variante proposée trouve globalement les meilleurs résultats. Cette variante utilise une idée relativement nouvelle puisqu'elle utilise une seule colonie de fourmis et une structure de phéromone pour chaque objectif, et les fourmis considèrent aléatoirement à chaque étape de construction un objectif à optimiser.

## Publications

### Revues nationales avec comité de lecture

Optimisation par colonies de fourmis pour le problème du sac-à-dos multi-dimensionnel. I. Alaya, C. Solnon, K Ghedira. Techniques et Sciences Informatiques (TSI) 26(3-4):371-390, Hermès-Lavoisier. 2007.

### Conférences internationales avec comité de lecture et actes

Ant Colony Optimization for Multi-objective Optimization Problems. I. Alaya, C. Solnon, K Ghedira. Dans 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI), Patras, Greece. pp. 450-457. IEEE Computer Society . 2007.

Ant algorithm for the multidimensional knapsack problem. I. Alaya, C. Solnon, K Ghedira. Dans Proceedings of international conference on Bioinspired Methods and their Applications (BIOMA 2004), Slovenia. pp. 63-72. 2004.

### Conférences nationales avec comité de lecture et actes

Algorithme fourmi avec différentes stratégies phéromonales pour le sac à dos multidimensionnel. I. Alaya, C. Solnon, K Ghedira. Dans Méthodologies et Heuristiques pour l'Optimisation des Systèmes Industriels (MHOSI 2005), . 2005.

### Edition scientifique d'ouvrages

Ant colony optimization for the k-graph partitioning problem. I. Alaya, O Sammoud, M. Hammami, K Ghedira. Tunisia Japan Symposium on Sciences and Technologies (TJASST03) 2003.

### Rapports de recherche/technique

Optimisation par colonies de fourmis pour le problème du sac-à-dos multidimensionnel. I. Alaya, C. Solnon, K Ghedira. Rapport de recherche RR-LIRIS-2006-005 , Soumis à TSI. 2006.



# implication de l'oxygène et des anti-oxydants dans le processus de radiolyse de l'eau induit par l'irradiation aux ions de haute énergie : simulations numériques pour la radiobiologie.

**Anthony Coliaux**

**Thèse**

Soutenue le 15/12/2009

à l'Université de Lyon- Université Claude Bernard Lyon 1

## **Jury**

Mr Douki Thierry

Mr Gervais Benoît

Mr Balosso Jacques

Mr Gauduel Yann

Mr Montarou Gérard

Mr Beuve Michaël

Mme Rodriguez-Lafrasse

Rapporteur

Rapporteur

Examineur

Examineur

Examineur

Directeur de thèse

Co-directrice

## **Publications**

### **Revues internationales avec comité de lecture**

Different mechanisms of cell death in radiosensitive and radioresistant p53 mutated head and neck squamous cell carcinomas exposed to carbon ions and x-rays. M. Maalouf, G. Alphonse, A. Coliaux, M. Beuve, S. Trajkovic-Bodennec, P. Battiston-Montagne, I. Testard, O. Chapet, M. Bajard, G. Taucher-Scholz, C. Fournier, C. Rodriguez-Lafrasse. *Int J Radiat Oncol Biol Phys* 74(1):200-209. 2009.

Statistical effects of dose deposition in track-structure modelling of radiobiology efficiency. M. Beuve, A. Coliaux, D. Dabli, D. Dauvergne, R. Gervais, G. Montarou, E. Testa. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and* 267(6):983-988. 2009.

Parameters and Local Effect Model predictions for head and neck squamous carcinoma cells exposed to High Linear Energy Transfert ions. M. Beuve, G. Alphonse, M. Maalouf, A. Coliaux, P. Battiston-Montagne, P. Jalade, E. Balanzat, A. Demeyer, M. Bajard, C. Rodriguez-Lafrasse. *International Journal of Radiation Oncology\*Biology\*Physics* 71(2):635-642. 2008.



# Recherche par similarité dans les bases de données multimédia : application à la recherche par le contenu d'images

**Imane Daoudi**

**Thèse**

Soutenue le 17/07/2009

à l'INSA de Lyon, Faculté des sciences de Rabat

## Jury

Pr Baskurt Atilla, LIRIS, Lyon	Directeur
Pr Daoudi Mohamed, Lifi, lille	Rapporteur
PH Ibn El Haje El Hassan, traitement images, Rabat, Maroc	Rapporteur
Pr Aboutajdine Driss, Lirit, Rabat, Maroc	Président
Pr Idrissi Khalid, LIRIS, Lyon	Examineur
PH Ouatik Said, LSIQ, Fes, Maroc	Examineur

Contact : khalid.idrissi@insa-lyon.fr

## Résumé

L'émergence des données numériques multimédia ne cesse d'augmenter. L'accès, le partage, le stockage et la recherche de ces données sont devenus des besoins réels et les problématiques sont nombreuses et variées. Afin de rendre cette masse d'information facilement exploitable pour tout utilisateur, il est nécessaire de disposer de techniques d'indexation et de recherche rapides et efficaces. Mes travaux de thèse s'inscrivent dans le domaine des données multimédia et plus précisément des images fixes. L'objectif principal est de développer une méthode performante d'indexation et de recherche des plus proches voisins () qui soit adaptée à la recherche d'image par le contenu et aux propriétés des descripteurs d'images (grand volume, grande dimension, hétérogénéité, etc.). Il s'agit d'une part, d'apporter des réponses aux problèmes de passage à l'échelle et de la malédiction de la dimension et d'autre part de traiter les problèmes de mesure de similarité qui se posent, et qui sont liés à la nature des données manipulées. Notre première proposition consiste en l'utilisation d'une structure d'indexation multidimensionnelle basée sur l'approximation ou filtrage, par une amélioration de la méthode RA-Blocks. Elle repose sur un algorithme de découpage de l'espace de données qui améliore notablement la capacité de stockage de l'index ainsi que le temps de la recherche. Dans une deuxième temps, nous proposons une méthode d'indexation multidimensionnelle adaptée à des données hétérogènes (couleur, texture, forme). Notre méthode combine une technique non linéaire de la réduction de la dimension à une structure d'indexation multidimensionnelle basée sur l'approche approximation. Cette combinaison permet de répondre à travers, un formalisme unique, aux différents verrous que nous nous sommes fixés de lever.

## Publications

### Revue internationale avec comité de lecture

An efficient High-Dimensional Indexing Method for Content-Based Image Retrieval in Large Image Databases. I. Daoudi, K. Idrissi, S. Ouatik, A. Baskurt, D. Aboutajdine. EURASIP Journal, Signal processing : Image communication 24():775-790, Elsevier. 2009.

Vector Approximation based Indexing for High-Dimensional Multimedia Databases . I. Daoudi, S. Ouatik, A. El Kharraz, K. Idrissi, D. Aboutajdine. Engineering Letters 16(2):210-218, ISSN 1816-0948. 2008.

### Conférences internationales avec comité de lecture et actes

Kernel Region Approximation Blocks For Indexing Heterogonous Databases. I. Daoudi, K. Idrissi, S. EL Alaoui Ouatik. Dans IEEE International Conference on Multimedia & Expo, Hannover - Germany. pp. 1237-1240. ISBN 978-1-4244-2571-. 2008.

Kernel Based Approach for High Dimensional Heterogeneous Image Features Management in CBIR Context.

I. Daoudi, K. Idrissi, S. EL Alaoui Ouatik. Dans Advanced Concepts for intelligent Vision Systems, . 2008.

A semi supervised metric learning for content-based image retrieval,. I. Daoudi, K. Idrissi, S. EL Alaoui Ouatik. Dans International Conference on Signal-Image Technology & Internet-Based Systems, Marrakech. 2009. (diffusion restreinte)

#### **Conférences nationales avec comité de lecture et actes**

Une mesure de similarité par une approche noyau pour l'indexation et la recherche par le contenu dans les bases de données hétérogènes. I. Daoudi, K. Idrissi, S. Ouatik. (CORESA (COmpression et Representation des Signaux Audiovisuels)), , Toulouse. 5p. 2009.

#### **Rapports de recherche/technique**

Recherche par similarité dans les grandes bases de données multimédias : application à la recherche par le contenu dans les bases d'images. I. Daoudi. Rapport de recherche RR-LIRIS-2007-013 2007.

# Simulation du Mouvement Pulmonaire pour l'amélioration des traitements par radiothérapie

Anne-Laure Didier

Thèse

Soutenue le 30/03/2009

à l'Université de Lyon- Université Claude Bernard Lyon 1

## Jury

Pr Neveu Matc, LE2I, Dijon

MC Promayon Emmanuel, TIMC, Grenoble

BEUVE Michaël, Université Claude Bernard LYON 1

CROSNIER André, Professeur des Universités, Université de MONTPELLIER

GIRAUD Philippe, Hôpital Européen Georges Pompidou – PARIS

SHARIAT Behzad, Université Claude Bernard LYON 1

Rapporteur

Rapporteur

Co-directeur

Examineur

Examineur

Directeur

Contact : bshariat@liris.cnrs.fr

## Résumé

Avec plus de 900 000 nouveaux cas chaque année chez les hommes et 330 000 chez les femmes, le cancer du poumon est actuellement l'un des cancers les plus répandus dans le monde. La majorité des patients atteints de ce cancer subiront un traitement par radiothérapie. La radiothérapie ainsi que l'hadronthérapie nécessitent un contrôle précis sur la position du volume tumoral durant le traitement, afin d'éviter l'irradiation des tissus sains. Lorsque la tumeur se trouve sur un organe en mouvement, la difficulté majeure est de cibler la tumeur pendant le traitement. Ce facteur limitant est particulièrement marqué dans le cas des tumeurs pulmonaires, ce qui explique les résultats cliniques encore insatisfaisants. Les stratégies actuelles utilisées en clinique, permettant de prendre en compte le mouvement de la tumeur au cours de la respiration, présentent toutes des inconvénients majeurs. Par exemple, certaines consistent à augmenter les marges d'irradiation ce qui entraîne un dépôt de dose important sur les tissus sains avoisinants, d'autres, vont consister à suivre des implants positionnés dans la tumeur ce qui rend la thérapie très invasive et enfin, certaines vont irradier la tumeur à un instant précis du cycle respiratoire, faisant ainsi l'hypothèse contestée d'une reproductibilité du mouvement tumoral au cours du cycle respiratoire. Notre objectif est de mettre en place un modèle biomécanique des poumons et de leur environnement afin de prédire la position de la tumeur à partir de paramètres mesurables directement en externe tels que la quantité d'air inspirée ou encore le mouvement de la surface externe de la peau. Ce modèle s'appuie sur les données anatomiques du patient (géométriques, physiologiques et biomécaniques) et permet de reproduire la variabilité du mouvement respiratoire de façon non-invasif. La variabilité du mouvement pulmonaire s'expliquant, en particulier, par l'action assez indépendante des muscles intercostaux et du diaphragme lors de la respiration, le modèle biomécanique développé au cours de cette thèse comprend la modélisation de ces éléments anatomiques. Des études sur la cinématique des côtes ainsi que sur le mouvement du diaphragme et le comportement des tissus mous ont été menées et ont permis, par la suite, d'établir un modèle de l'environnement pulmonaire précis et personnalisé. Le rôle important de la plèvre lors de la respiration a été démontré et celle-ci a donc été intégrée au modèle biomécanique du système respiratoire. Les simulations ont été réalisées à l'aide d'un logiciel de calcul par éléments finis et une relation entre le mouvement externe de la peau et le mouvement interne des côtes et donc des poumons a pu être établie. Mots clés : traitements oncologiques, anatomie, traitements d'images médicales, méthode des éléments finis, cinématique des côtes, rôle de la plèvre

## Publications

### Conférences internationales avec comité de lecture et actes

A chest wall model based on rib kinematics. D.A-L Didier, P.F. Villard, J. Saadé, JM. Moreau, M. Beuve, B. Shariat. Dans 6th International Conference on Biomedical Visualization, IEEE ed. Pompeu Fabra University,

Barcelona, Spain. 2009. (à paraître)

Mechanical role of pleura on lung motion during breathing . D.A-L Didier, P.F. Villard, M. Beuve, B. Shariat. Dans XXXIIème Congrès Annuel de la SOCIÉTÉ DE BIOMÉCANIQUE, Lyon (France). 2007.

Breathing Thorax Simulation based on Pleura Behaviour and Rib Kinematics. D.A-L Didier, P.F. Villard, J-Y Bayle, M. Beuve, B. Shariat. Dans Information Visualisation - MediVis, IEEE ed. Zurich (Suisse). pp. 35-40. 2007.

# Contribution à la numérisation des documents imprimés du XVIIIe siècle : Application au cas de la Gazette de Leyde

Loris Eynard

Thèse

Soutenue le 18/12/2009

## Jury

Mme Likforman-Sulem Laurence, TELECOM ParisTech, Paris	Rapporteur
M Cubaud Pierre, CEDRIC, Paris	Rapporteur
M Florenzano Michel, MAP, Marseille	Président
M Marinai Simone, Università di Firenze, Florence, Italie	Examineur
Mme Mercier-Faivre Anne-Maie, LIRE, Lyon	Examineur
M Emptoz Hubert, LIRIS, Villeurbanne	Directeur

Contact : loris\_eynard@yahoo.fr

## Résumé

La Gazette de Leyde est un journal politique parmi les plus influents du XVIIIème siècle et représente un corpus de plus de 140 000 pages réparti sur plus d'un siècle de parution. Le projet de numérisation, valorisation et exploitation a été monté en collaboration avec les chercheurs en Sciences Humaines de l'UMR LIRE, son objectif était la meilleure utilisation des potentiels des STIC pour améliorer leur travail (d'historiens) et répondre à leurs besoins. Permettre une navigation rapide et intelligente dans le corpus de la Gazette de Leyde est apparu comme le défi à relever en priorité. Dans cet objectif nous avons travaillé sur les structures physique et logique (mise en page, typographie) en exploitant notamment les particularités du style des mots et des titres des articles. Cette démarche nous a permis de faire la différenciation entre les styles italique et romain. Couplée à l'élaboration d'une technique de repérage des titres de rubriques, (qui reposait sur leur typographie et placement dans les colonnes de texte), elle nous a conduit à la création automatique d'un index intelligent de la Gazette de Leyde, en mode image. La seconde partie de cette thèse consiste en une étude de faisabilité de l'adaptation de la méthode de Transcription Assistée par Ordinateur qui se présente comme une alternative aux logiciels commerciaux O.C.R, inefficaces sur les documents anciens plus ou moins bien conservés. La transcription consiste à regrouper les caractères en classes par similarité de formes. Ces classes seront ensuite étiquetées par un spécialiste du document ce qui conduit à un résultat de bonne qualité.

## Abstract

The Gazette of Leyde is a political newspaper among the most influential from the eighteenth century. It consists of a corpus containing over 140.000 pages spread over a century of publication. The project of digitization, development and exploitation has been planed in collaboration with Humanities researchers from the UMR LIRE. The goal of this project was to derive the best of CSIT's potential in order to improve historians' work and to meet their needs better. Being able to perform quick and intelligent surfing across the corpus of the Gazette of Leyde has become a high-priority challenge. For that reason, we have worked on the gazette's physical and logical structures (layout, typography) by exploiting the words style and titles of articles particularities. This enabled us to create a method to differenciante between the Roman and italic styles. Coupled with the development of a technique for identifying the titles of articles, (based on typography and their placement in the columns of text), it has led to the automatization of the creation of an intelligent index of the Gazette of Leyde in image mode. The second part of the thesis consists in a feasibility study for the development of a method of computer-aided transcription of the text. We are developping this method as an alternative to the commercial OCR software which is inefficient for old documents more or less well preserved. The transcript consists in grouping characters into classes of similar

shapes. These classes are then labeled by a specialist of that document. The present method leads to good quality results.

## **Publications**

### **Conférences internationales avec comité de lecture et actes**

Particular Words Mining and Article Spotting in Old French Gazettes. L. Eynard, Y. Leydier, H. Emptoz. Dans MLDM 2009, Springer Verlag ed. Leipzig, Allemagne. 2009.

Italic or Roman : Word Style Recognition Without A Priori Knowledge for Old Printed Documents. L. Eynard, H. Emptoz. Dans ICDAR 2009, IEEE ed. Barcelone, Espagne. 2009.

Segmentation Of Broken Characters Using Pattern Matching. L. Eynard, F Lebourgeois, H. Emptoz. Dans Ninth International Conference on Pattern Recognition and Information Processing, PRIP'2007, may 22-24, 2007, Minsk, Belarus. pp. 101-107. ISBN 978-985-6744-29-. 2007.

Virtual arm for the phantom pain therapy. L. Eynard, A Meyer, S. Bouakaz. Dans International Conference on Machine Intelligence (ACIDCA-ICMI), Tozeur, Tunisia.. 2005.

# Contributions au tri automatique de documents et de courrier d'entreprises

Djamel Gaceb

Thèse

Soutenue le 19/10/2009

à l'Institut National de Sciences Appliquées de Lyon

## Jury

Pr OGIER Jean-Marc , L3i, La Rochelle	Président
Pr CHERIET Mohamed , LIVIA, Montréal, Québec	Rapporteur
Pr VIARD-GAUDIN Christian, IRCCyN, Nantes	Rapporteur
Pr EMPTOZ Hubert, LIRIS-INSA , Lyon	Directeur
Mle EGLIN Véronique, LIRIS-INSA , Lyon	Co-directeur
M MAISONNEUVE Bruno , CESA-VINCI, Lyon	Examineur

Contact : djamel.gaceb1@insa-lyon.fr

## Résumé

Ce travail de thèse s'inscrit dans le cadre du développement de systèmes de vision industrielle pour le tri automatique de documents et de courriers d'entreprises. Ces systèmes sont par nature très exigeants en temps de traitement mais aussi en justesse et précision des résultats. Les systèmes actuels sont composés, pour la plupart, de modules séquentiels exigeant des algorithmes efficaces et rapides tout au long de la chaîne des traitements, depuis les étapes de bas niveau jusqu'aux étapes de niveau supérieur d'analyse fine et de reconnaissance des contenus. Les architectures existantes, dont nous avons balayé les spécificités dans les trois premiers chapitres de la thèse, présentent des faiblesses qui se traduisent par des erreurs de lecture et des rejets que l'on impute encore trop souvent aux OCR. Or, les étapes responsables de ces rejets et de ces erreurs de lecture sont les premières à intervenir dans le processus, à savoir celles de segmentation et de localisation de zones d'intérêts ; ces deux étapes qui s'impliquent mutuellement conditionnent les performances des systèmes et le rendement des chaînes de tri automatique. Nous avons ainsi choisi porter notre contribution sur les aspects inhérents à la segmentation des images de courriers et la localisation de leurs régions d'intérêt (comme la zone d'adresse) en investissant une nouvelle approche pyramidale de modélisation par coloration hiérarchique de graphes ; à ce jour, la coloration de graphes n'a jamais été exploitée dans un tel contexte. Elle intervient dans notre contribution à toutes les étapes d'analyse de la structure des documents ainsi que dans la prise de décision pour la reconnaissance (reconnaissance de la nature du document à traiter et reconnaissance du bloc adresse). La partie de reconnaissance a été conçue autour d'un apprentissage traité à l'aide d'un modèle unique portant sur la b-coloration de graphe. Notre architecture a été conçue pour réaliser essentiellement les étapes d'analyse de structures et de reconnaissance en garantissant une réelle coopération entre les différents modules d'analyse et de décision. Elle s'articule autour de trois grandes parties : une partie de segmentation bas niveau (binarisation et recherche de connexités), une partie d'extraction de la structure physique par coloration hiérarchique de graphe et une partie de localisation de blocs adresse et de classification de documents. Les algorithmes impliqués dans le système ont été conçus pour leur rapidité d'exécution (en adéquation avec les contraintes de temps réels), leur robustesse, et leur compatibilité. Les expérimentations réalisées dans ce contexte sont très encourageantes et offrent également de nouvelles perspectives à une plus grande diversité d'images de documents.

## Abstract

This thesis deals with the development of industrial vision systems for automatic business documents and mail sorting. These systems need very high processing time, accuracy and precision of results. The current systems are most of time made of sequential modules needing fast and efficient algorithms throughout the processing line: from low to high level stages of analysis and content recognition. The existing architectures

that we have described in the three first chapters of the thesis have shown their weaknesses that are expressed by reading errors and OCR rejections. The modules that are responsible of these rejections and reading errors are mostly the first to occur in the processes of image segmentation and interest regions location. Indeed, these two processes, involving each other, are fundamental for the system performances and the efficiency of the automatic sorting lines. In this thesis, we have chosen to focus on different sides of mail images segmentation and of relevant zones (as address block) location. We have chosen to develop a model based on a new pyramidal approach using a hierarchical graph coloring. As for now, graph coloring has never been exploited in such context. It has been introduced in our contribution at every stage of document layout analysis for the recognition and decision tasks (kind of document or address block recognition). The recognition stage is made about a training process with a unique model of graph b-coloring. Our architecture is basically designed to guarantee a good cooperation between the different modules of decision and analysis for the layout analysis and the recognition stages. It is composed of three main sections: the low-level segmentation (binarisation and connected component labeling), the physical layout extraction by hierarchical graph coloring and the address block location and document sorting. The algorithms involved in the system have been designed for their execution speed (matching with real time constraints), their robustness, and their compatibility. The experimentations made in this context are very encouraging and lead to investigate a wider diversity of document images.

## **Publications**

### **Revue internationale avec comité de lecture**

Robust Approach of Address Block Localization in Business Mail by Graph Coloring. DJ. Gaceb, V. Eglin, F Lebourgeois, H. Emptoz. International Journal of Information Technology (IAJIT) 6(3):221-229, ISSN 1683-3198. 2009.

Improvement of postal mail sorting system. DJ. Gaceb, V. Eglin, F Lebourgeois, H. Emptoz. International Journal on Document Analysis and Recognition (IJAR) 11(2):67-80, Springer, Computer Science, ISSN 1433-2833. 2008.

### **Conférences internationales avec comité de lecture et actes**

Application of graph coloring in physical layout segmentation. DJ. Gaceb, V. Eglin, F Lebourgeois, H. Emptoz. Dans International Conference on Pattern Recognition (ICPR 2008), IEEE ed. Tampa, Floride, USA. pp. 1-4. ISBN 978-1-4244-2175-. ISSN 1051-4651. 2008.

Physical Layout Segmentation of Mail Application Dedicated to Automatic Postal Sorting System. DJ. Gaceb, V. Eglin, F Lebourgeois, H. Emptoz. Dans The Eighth IAPR International Workshop on Document Analysis Systems (DAS), IEEE Computer Society ed. Nara Prefectural New Public Hall, Nara, Japan. pp. 408-414. ISBN 978-0-7695-3337-. 2008.

Address block localization based on graph theory. DJ. Gaceb, V. Eglin, F Lebourgeois, H. Emptoz. Dans Document Recognition and Retrieval XIV, SPIE Int. Soc. Opt. Eng ed. San Jose (USA, Californie). p. 12. DRR\_XIV 6815. SPIE USA. ISBN 9780819469878. 2008.

A New Pyramidal Approach for the Address Block Location Based on Hierarchical Graph Coloring. DJ. Gaceb, V. Eglin, F Lebourgeois, H. Emptoz. Dans The International Conference on Image Analysis and Recognition (ICIAR 2007), Springer Berlin / Heidelberg ed. Montreal, Canada. pp. 1276-1288. Lecture Notes in Computer Science 4633. Springer Berlin / Heidelberg Germany. ISBN 978-3-540-74258-. ISSN 0302-9743 (Pri. 2007).

Contribution to the Automatic Recognition of Business Documents. DJ. Gaceb, F Lebourgeois, V. Eglin, H. Emptoz. Dans Tenth International Workshop on Frontiers in Handwriting Recognition (IWFHR), IRISA Rennes ed. La Baule, France, 6p, (lang: en) 2006.

Handwriting similarities as features for the characterization of writer style invariants and image compression. DJ. Gaceb, V. Eglin, S. Bres. Dans International Conference on Image Analysis and Recognition, Universitié de porto ed. Porto (Portugal). pp. 776-789. Lecture Notes in Computer Science Image Analysis and Recognition . Springer USA. ISBN 978-3-540-44894-. ISSN 0302-9743. 2006.

Graph b-Coloring for Automatic Recognition of Documents. DJ. Gaceb, V. Eglin, F Lebourgeois, H. Emptoz. Dans ICDAR, IAPR ed. Barcelona. pp. 261-265. IEEE . ISBN 978-0-7695-3725-2. 2009.

### **Conférences nationales avec comité de lecture et actes**

Application de la coloration de graphes dans la segmentation de courrier postal. DJ. Gaceb, V. Eglin, F Lebourgeois, H. Emptoz. Dans Manifestation des Jeunes Chercheurs en Sciences et Technologies de l'Information et de la Communication (MajecSTIC2008), Marseille, France. p. 08. 2008.

Implication de la b-coloration de graphes pour la reconnaissance automatique du type de document. DJ. Gaceb, V. Eglin, F Lebourgeois, H. Emptoz. Dans Colloque International Francophone sur l'Ecrit et le Document(CIFED), France. pp. 31-36. 2008.

Segmentation de la structure physique de courrier application dédiée à un système automatique de tri postal. DJ. Gaceb, V. Eglin, F Lebourgeois, H. Emptoz. Dans The 10th Maghrebian Conference on Software Engineering and Artificial Intelligence (MCSEAI), Oran, Algeria. pp. 548-553. 2008.

Localisation du bloc-adresse sur des objets postaux basée sur la coloration hiérarchique de graphe. DJ. Gaceb, V. Eglin, F Lebourgeois, H. Emptoz. Dans 12èmes journées d'étude et d'échange COmpression et REprésentation des Signaux Audiovisuels, CORESA., Montpellier. pp. 176-180. 2007.

Approche Coopérative d'Extraction des Zones d'Intérêt, Appliquée au Tri Automatique de Courriers d'Entreprises. DJ. Gaceb, V. Eglin, F Lebourgeois, H. Emptoz. Dans The 4th International Conference on the Sciences of Electronics, Technologies of Information and Telecommunications (SETIT 2007), IEEE ed. Hammamet, Tunisia. ISBN 978-9973-61-4742. 2007.

Contribution à la reconnaissance automatique de documents d'entreprises. DJ. Gaceb, F Lebourgeois, V. Eglin, H. Emptoz. Dans CORESA, Compression et représentation des Signaux Audiovisuels, Caen. 2006.

Extraction de similarités dans les manuscrits du patrimoine pour la compression des images et la caractérisation des styles. DJ. Gaceb, V. Eglin, S. Bres. Dans CIFED, Colloque International Francophone sur l'Ecrit et le Document, Jouve - Paris ed. Fribourg. ISBN 2-9522067-1-6. 2006.



# Le document numérique : la complexité des formes et les formes de la complexité

Joël Gardes

Thèse

Soutenue le 24/07/2009

à l'INSA Lyon

## Jury

Pr INGOLD Rolf, DIUF, Fribourg Suisse	Rapporteur
Pr BACHIMONT Bruno, Heudiasyc UTC, Compiègne France	Rapporteur
Pr JUTTEN Christian, GIPSA-lab, Grenoble France	Président
Pr NAGY George, ESCE Rensselaer Institute, Troy United States	Examineur
Pr EMPTOZ Hubert, LIRIS INSA Lyon, Lyon France	Directeur
M GARCIA Christophe, Orange Labs, Rennes France	Co-directeur
M CHENE Denis, Orange Labs, Grenoble France	Examineur

Contact : joel.gardes@orange-ftgroup.com

## Résumé

Les réseaux de télécommunication actuels offrent une interconnexion généralisée de terminaux dont les capacités de calculs et de mémoire évoluent sans cesse. De nouveaux comportements émergent face à une information devenue pervasive. L'utilisateur n'est plus passif, ni unique, c'est un acteur local du traitement de l'information. Une interface doit prendre en considération la variété des modes d'accès et d'usages qui conditionnent la présentation de tout contenu. La présentation de l'information se situe au carrefour de trois grands domaines indissociables : l'acquisition et le traitement du signal, l'ingénierie des données et des connaissances, l'interaction homme machine. Ce thème pluridisciplinaire, très proche du document numérique nous permet de comprendre et prévoir nos manières d'appréhender l'information numérique. Nous y retrouvons des sujets tels que la qualité de la numérisation et son évaluation. Cette thèse présente un socle théorique identifiant les problèmes posés par la communication numérique dans le contexte évoqué. Les éléments de solutions sont présentés par des réalisations techniques innovantes sur la reconnaissance de mots et la distance entre informations. Ces applications, issues des théories de la complexité et de la cybernétique, ont pour caractéristiques remarquables leur simplicité algorithmique face à la complexité des problèmes traités, et leur facilité d'appropriation par un utilisateur totalement inclus dans la boucle de traitement.

## Abstract

The current telecommunications networks offer a generalized interconnection of terminals whose computing and memory capacities are changing constantly. New behaviours emerge address information which has become pervasive. The user is no more passive, or unique, he is a local body of information processing. An interface must take into consideration the variety of access methods and practices that shape the presentation of content. Information presentation is at the crossroads of three major interrelated areas: acquisition and signal processing, engineering data and knowledge, human computer interaction. This multidisciplinary subject, very close to the digital document allows us to understand and predict the way we understand digital information. We find such topics as quality of scanning and its assessment. This thesis presents a theoretical foundation identifying the problems of digital communication in the context mentioned. The elements of solutions are presented through innovative technical achievements on the characters and words recognition, and distance information. These applications, from theories of complexity and cybernetics, are remarkable features with their algorithmic simplicity given the complexity of problems addressed, and ease of ownership by a user fully included in the loop.



# Caractérisation des écritures médiévales par des méthodes statistiques basées sur les cooccurrences

Koubaa Ikram Moalla

Thèse

Soutenue le 28/12/2009

à l'INSA de Lyon

## Jury

Najoua BEN AMARA , Université de Sousse,  
Eric ANQUETIL, INSA Rennes,  
Moncef CHARFI, ENIS de Sfax,  
M. Adel ALIMY, ENIS de Sfax,  
Hubert EMPTOZ, INSA de Lyon,  
Frank LE BOURGEOIS, INSA de Lyon,

Rapporteur  
Rapporteur  
Examineur  
Directeur de thèse  
Directeur de thèse  
Invité

## Résumé

Dans le domaine du traitement des images de documents, l'analyse des formes d'écritures manuscrites est une problématique encore peu abordée et plus particulièrement pour les manuscrits du Moyen Âge.

Cette thèse a pour objet l'élaboration de méthodologies d'analyse permettant de décrire et de comparer les écritures manuscrites anciennes, méthodologies d'analyse globale ne nécessitant pas segmentation.

Elle propose de nouveaux descripteurs robustes basés sur des statistiques d'ordre 2, la contribution essentielle reposant sur la notion de cooccurrence généralisée qui mesure la loi de probabilité conjointe d'informations extraites des images ; c'est une extension de la cooccurrence des niveaux de gris, utilisée jusqu'à présent pour caractériser les textures qui nous a permis d'élaborer diverses cooccurrences, spatiales relatives aux orientations et aux courbures locales des formes, paramétriques qui mesurent l'évolution d'une image subissant des transformations successives.

Le nombre de descripteurs obtenu étant très (trop) élevé, nous proposons des méthodes conçues à partir des plus récentes méthodes d'analyse statistique multidimensionnelle de réduction de ce nombre. Ces démarches nous ont conduit à introduire la notion de matrices de cooccurrences propres qui contiennent l'information essentielle permettant de décrire finement les images avec un nombre réduit de descripteurs.

Dans la partie applicative nous proposons des méthodes de classification non supervisées d'écritures médiévales.

Le nombre de groupes et leurs contenus dépendent des paramètres utilisés et des méthodes appliquées. Nous avons aussi développé un moteur de recherche d'écritures similaires.

Dans le cadre du projet ANR-MCD GRAPHEM, nous avons élaboré des méthodes permettant d'analyser et de suivre l'évolution des écritures du Moyen Age.

## Publications

### Conférences internationales avec comité de lecture et actes

Contribution to the discrimination of the medieval manuscript texts: Application in the palaeography. I. Moalla, F Lebourgeois, H. Emptoz. Dans Document Analysis Systems (DAS&#8217;06) Lecture Notes in Computer Science, Vol. 3872, Springer ed. Nelson, New Zealand. pp. 25-37. ISBN 3-540-32140-3. ISSN 0302-9743. 2006.

### Conférences nationales avec comité de lecture et actes

Discrimination des styles d'écriture des manuscrits médiévaux pour la paléographie. I. Moalla, F Lebourgeois, H. Emptoz, M. A. Alimi. Dans Colloque International Francophone sur l'Ecrit et le Document, . 2006.



# Algorithmes et structures de données compactes pour la visualisation interactive d'objets 3D volumineux

Clément Jamin

Thèse

Soutenue le 25/09/2009

à l'Université Claude Bernard Lyon 1

## Jury

DR Devillers Olivier, INRIA - Geometrica, Sophia Antipolis	Rapporteur
PR Michelucci Dominique, Université de Bourgogne, Dijon	Rapporteur
DR Lévy Bruno, INRIA - LORIA, Nancy	Examineur
PR Séqueira Jean, Université de la Méditerranée, Marseille	Examineur
PR Akkouche Samir, LIRIS, Lyon	Directeur
Dr Gandoin Pierre-Marie, LIRIS, Lyon	Co-directeur

Contact : pierre-marie.gandoin@liris.cnrs.fr

## Résumé

Les méthodes de compression progressives sont désormais arrivées à maturité (les taux de compression sont proches des taux théoriques) et la visualisation interactive de maillages volumineux est devenue une réalité depuis quelques années. Cependant, même si l'association de la compression et de la visualisation est souvent mentionnée comme perspective, très peu d'articles traitent réellement ce problème, et les fichiers créés par les algorithmes de visualisation sont souvent beaucoup plus volumineux que les originaux. En réalité, la compression favorise une taille réduite de fichier au détriment de l'accès rapide aux données, alors que les méthodes de visualisation se concentrent sur la rapidité de rendu : les deux objectifs s'opposent et se font concurrence. A partir d'une méthode de compression progressive existante incompatible avec le raffinement sélectif et interactif, et uniquement utilisable sur des maillages de taille modeste, cette thèse tente de réconcilier compression sans perte et visualisation en proposant de nouveaux algorithmes et structures de données qui réduisent la taille des objets tout en proposant une visualisation rapide et interactive. En plus de cette double capacité, la méthode proposée est out-of-core et peut traiter des maillages de plusieurs centaines de millions de points. Par ailleurs, elle présente l'avantage de traiter tout complexe simplicial de dimension  $n$ , des soupes de triangles aux maillages volumiques. Mots clés : Compression sans perte, Visualisation interactive, Maillages volumineux, Out-of-core

## Abstract

Progressive compression methods are now mature (obtained rates are close to theoretical bounds) and interactive visualization of huge meshes has been a reality for a few years. However, even if the combination of compression and visualization is often mentioned as a perspective, very few papers deal with this problem, and the files created by visualization algorithms are often much larger than the original ones. In fact, compression favors a low file size to the detriment of a fast data access, whereas visualization methods focus on rendering speed : both goals are opposing and competing. Starting from an existing progressive compression method incompatible with selective and interactive refinements and usable on small-sized meshes only, this thesis tries to reconcile lossless compression and visualization by proposing new algorithms and data structures which radically reduce the size of the objects while supporting a fast interactive navigation. In addition to this double capability, our method works out-of-core and can handle meshes containing several hundreds of millions vertices. Furthermore, it presents the advantage of dealing with any  $n$ -dimensional simplicial complex, which includes triangle soups or volumetric meshes. Keywords : Lossless compression, Interactive visualization, Large meshes, Out-of-core

## Publications

### Revues internationales avec comité de lecture

CHuMI Viewer: Compressive Huge Mesh Interactive Viewer. C Jamin, P.-M. Gandoin, S. Akkouche.

Computer & Graphics (). 2009. (à paraître)

#### **Reuves nationales avec comité de lecture**

Compression out-of-core pour la visualisation interactive de maillages volumineux. C Jamin, P.-M. Gandoïn, S. Akkouche. REFIG (Revue Electronique Francophone d'Informatique Graphique) (). 2009. (à paraître)

#### **Autres conférences**

Compression out-of-core pour la visualisation interactive de maillages volumineux. C Jamin, P.-M. Gandoïn, S. Akkouche. Dans AFIG, Toulouse. 2008.

#### **Rapports de recherche/technique**

A Compact Data Structure For The Navigation Into Arbitrary Meshes. C Jamin, P.-M. Gandoïn, S. Akkouche. Rapport de recherche RR-LIRIS-2008-014 , Soumis à Eurographics 2009. 2008.

# Analyse multirésolution des images de documents manuscrits : Application à l'analyse de l'écriture

Guillaume Joutel

Thèse

Soutenue le 26/06/2009

à l'Institut National de Sciences Appliquées de Lyon

## Jury

Pr Lorette Guy, IRISA, Rennes	Président
Pr Elmoataz Abderrahim, GREYC, Caen	Rapporteur
Pr Pellerin Denis, GIPSA-Lab, Grenoble	Rapporteur
Pr Emptoz Hubert, LIRIS, Lyon	Directeur
Mle Eglin Véronique, LIRIS, Lyon	Co-directeur
M. Muzerelle Denis, IRHT, Paris	Examineur

Contact : guillaume.joutel@liris.cnrs.fr

## Résumé

Les images de traits, et plus spécifiquement les images d'écritures manuscrites, ont des contenus souvent hétérogènes et nécessitent des méthodes spécifiques d'analyse pour être exploitées. Dans cette thèse, nous proposons la mise au point d'une approche de caractérisation des écritures manuscrites basée sur la transformée en ondelettes géométriques non-adaptatives que sont les Curvelets. Nous validons cette caractérisation dans différents cadres applicatifs de l'analyse d'images de documents. Les Curvelets ont été choisies pour leur propriété de bonne localisation des objets anisotropes et leur analyse directionnelle multi-échelle. Elles permettent l'extraction de deux primitives essentielles des écritures que sont l'orientation et la courbure à des niveaux d'échelles variables. Ces primitives sont rassemblées dans une matrice d'occurrences pour constituer la signature d'une écriture. Cette dernière est utilisée comme vecteur de caractéristiques dans une application de recherche d'images par le contenu. Nous proposons également une évaluation de la similarité locale entre formes plus petites, quelles qu'elles soient, où seule la primitive orientation est utilisée. Enfin, dans le cadre d'une spécialisation de notre méthode au corpus d'images fournies par le projet ANR Graphem, nous proposons la définition d'un indice de similarité, produite à partir d'échanges entre les partenaires de ce projet, qui tente de tirer partie des propriétés communes des écritures tout en essayant de prendre en compte leurs différences et leurs spécificités. Cette prise en compte, pondérable par l'utilisateur, permet une exploitation avancée des écritures du Moyen-Âge.

## Abstract

Images of strokes, especially handwriting documents images, are often composed of heterogeneous contents and require specific methods of analysis to be exploited. In this thesis, we propose the development of a characterization of handwritings based on a non-adaptive geometrical wavelet transform which is the Curvelets transform. We validate this characterization in different application frameworks of document images analysis. The Curvelets were chosen for their property of good localization of anisotropic objects and their directional multi-scale analysis. We can extract from those two essential primitives which are orientation and curvature at varying levels of scales. These primitives are gathered in a matrix of occurrences to form the signature of a handwriting. The latter is used as a features' vector in an content based image retrieval application. We also propose an evaluation of local similarity between the smaller forms, whatever they are, where the only primitive used is orientation. Finally, as part of a specialization of our method to the corpus of images provided by the ANR project Graphem, we propose the definition of a similarity index, generated from trade between partners in this project, which attempts to use common properties of handwritings while trying to address their differences and their characteristics. This consideration, weighted by the user, allows an advanced exploitation of medieval handwritings.

## **Publications**

### **Conférences internationales avec comité de lecture et actes**

Generic scale-space process for handwriting documents analysis. J.G Joutel, V. Eglin, H. Emptoz. Dans International Conference on Pattern Recognition (ICPR 2008), IEEE ed. Tampa, Floride. pp. 1-4. 2008.

Curvelets based queries for CBIR Application in handwriting collections. J.G Joutel, V. Eglin, S. Bres, H. Emptoz. Dans ICDAR, International conference on Document Analysis and Recognition, Curitiba, Brésil. pp. 649-653. ISBN 1520-5363. ISSN 0-7695-2822-8. 2007.

### **Conférences nationales avec comité de lecture et actes**

Extraction de caractéristiques dans les images par transformée multi-échelle. J.G Joutel, V. Eglin, S. Bres, H. Emptoz. Dans GRETSI, Groupe de Recherche et Etudes du Traitement du Signal, Troyes. pp. 469-472. 2007.

Une nouvelle approche pour indexer les documents manuscrits anciens. J.G Joutel, V. Eglin, H. Emptoz. Dans Colloque International Francophone sur l'Ecrit et le Document 2008, . pp. 85-90. 2008.

# Reconstruction 3D à partir de séquences vidéo pour l'acquisition du mouvement de personnages en temps réel et sans marqueur

**Brice Michoud**

**Thèse**

Soutenue le 30/09/2009

à l'Université Claude Bernard Lyon 1

## **Jury**

Pr. MOHR Roger, LJK, Grenoble	Président
Pr. DHOME Michel, LASMEA, Clermont Ferrand	Rapporteur
Pr. GUITTON Pascal, LABRI, Bordeaux	Président
Pr. BOUATOUCH Kadi, IRISA, Rennes	Examineur
Pr. DIPANDA Albert, LE2I, Dijon	Examineur
Pr. BOUAKAZ Saïda, LIRIS, Villeurbanne	Directeur
Mcf GUILLOU Erwan, LIRIS, Villeurbanne	Co-directeur

Contact : brice.michoud@liris.cnrs.fr

## **Résumé**

Dans cette thèse, nous nous intéressons à l'acquisition automatique de mouvements 3D de personnes. Cette opération doit être réalisée sans un équipement spécialisé (marqueurs ou habillage spécifique), pour rendre son utilisation générale, sous la contrainte du temps réel. Cette condition est nécessaire pour permettre des applications interactives. Pour répondre à ces questions, nous sommes amenés à traiter de la reconstruction et l'analyse de la forme 3D. Concernant le problème de reconstruction 3D en temps réel d'entités en mouvement à partir de plusieurs vues, les approches existantes font souvent appel à des calculs complexes incompatibles avec la contrainte du temps réel. Les approches du type \Nsfs{} (SFS) offrent un compromis intéressant entre efficacité algorithmique et précision. Ces dernières utilisent les silhouettes issues de chaque caméra pour proposer un volume englobant des objets. Cependant elles nécessitent un environnement particulièrement contraint, dont le placement minutieux des caméras. Les travaux présentés dans ce manuscrit généralisent l'utilisation des approches SFS à des environnements peu contrôlés. Ils s'appuient sur le domaine de visibilité de chaque caméra, et relaxent les contraintes de leur placement. L'introduction de critères géométriques réduit la quantité d'artefacts générés. Enfin une mesure de confiance sur l'existence de chaque point 3D compense les erreurs d'extraction de silhouette. Les principales méthodes d'acquisition de mouvements sans marqueur s'appuient sur une modélisation paramétrique du corps. L'acquisition du mouvement revient à déterminer les paramètres offrant la meilleure corrélation entre le modèle et la reconstruction 3D. Ceci implique la résolution d'un système de grande dimension et donc des temps de calculs prohibitifs. Notre objectif étant le suivi temps réel, nous proposons des méthodes qui offrent la précision requise et le temps réel. En premier lieu nous présentons deux approches construites sur l'extraction de la structure topologique de la forme 3D. Ces premières approches temps réel, restent sensibles aux erreurs induites par l'utilisation d'un faible nombre de caméras. Pour augmenter la robustesse, nous nous appuyons sur un marquage naturel des extrémités du corps : la peau. Couplé à un suivi temporel par filtre de Kalman, à un recalage d'objets géométriques simples (ellipsoïdes, sphères, etc.), nous proposons un système temps réel, offrant une erreur de l'ordre de 6%. Celui-ci acquiert le mouvement d'une personne en temps réel à partir de deux vues. De par sa robustesse, il permet le suivi simultané de plusieurs personnes, même lors de contacts. Les résultats obtenus ouvrent des perspectives à un transfert vers des applications grand public.

## **Publications**

### **Conférences internationales avec comité de lecture et actes**

Towards Removing Ghost-Components from Visual-Hull Estimations. B Michoud, E Guillou, S. Bouakaz, M.

Barnachon, A Meyer. Dans ICIG, Xi'An. 2009. (à paraître)

Largest Silhouette-Equivalent Volume for 3D Shapes Modeling without Ghost Object. B Michoud, S. Bouakaz, E Guillou, H. Briceno Pulido. Dans Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications, In conjunction with ECCV 2008, Marseille, France. 2008. (à paraître)

Real-time and Marker-free 3D Motion capture for Home Entertainment Oriented Applications. B Michoud, E Guillou, H. Briceno Pulido, S. Bouakaz. Dans ACCV'2007 : 8th Asian Conference on Computer Vision, Tokyo, Japan. 2007.

Real-time and Markerless 3D Human Motion Capture using Multiple Views. B Michoud, E Guillou, S. Bouakaz. Dans 2nd Human Motion Workshop, Rio de Janeiro, Brazil. 2007.

Real-Time Marker-free Motion Capture from multiple cameras. B Michoud, E Guillou, H. Briceno Pulido, S. Bouakaz. Dans ICCV'2007 : Eleventh IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil. 2007.

Shape From Silhouette: Towards a Solution for Partial Visibility Problem. B Michoud, E Guillou, S. Bouakaz. Dans Eurographics 2006, C.Hansen, D.Fellner ed. Vienna, Austria. pp. 13-16. Eurographics 2006 Short Papers Preceedings . ISSN 1017-4656. 2006.

Human model and pose Reconstruction from Multi-views. B Michoud, E Guillou, S. Bouakaz. Dans International Conference on Machine Intelligence (ACIDCA-ICMI), Tozeur, Tunisia. 2005.

#### **Conférences nationales avec comité de lecture et actes**

Reconstruction d'humanoïde à partir de plusieurs vues. B Michoud, E Guillou, S. Bouakaz. Dans Journées AFIG 2005, Strasbourg, France. pp. 171-181. 2005.

Reconstruction géométrique par estimation de posture. M. Barnachon, B Michoud, E Guillou, S. Bouakaz. Dans ORASIS, Trégastel. 2009.

Reconstruction Géométrique et Photométrie Incrémentale d'un Personnage en Mouvement à partir de Séquences Vidéo. M. Barnachon, E Guillou, B Michoud. Dans AFIG 2008, Toulouse. 2008.

Real-time and Markerless Full-Body Human Motion Capture. B Michoud, E Guillou, S. Bouakaz. Dans Groupe de Travail Animation et Simulation 2007 (GTAS'07), Lyon. 2007.

Extension de l'espace d'acquisition pour les méthodes de Shape From Silhouette. B Michoud, E Guillou, S. Bouakaz. Dans COMpression et REprésentation de Signaux Audiovisuels (CORESA) 2006, Caen, France. 2006.

#### **Autres conférences**

Extraction d'un Environnement Photométrique et Géométrique pour la Réalité Augmentée. F. Fouquet, F.-J.-P. Farrugia, B Michoud. Dans AFIG 2008 - 21èmes Journées de l'Association Française d'Informatique Graphique, Mathias Paulin, Loïc Barthe ed. Toulouse. pp. 227-236. IRIT Presse . ISBN 978-2-917490-03-. 2008.

#### **Rapports de recherche/technique**

Removing Camera Placement Constraints in Shape from Silhouette on Large Acquisition Volumes. B Michoud, E Guillou, H. Briceno Pulido, S. Bouakaz. Rapport de recherche RR-LIRIS-2008-016 2008.

**Alain Pujol**

Soutenue le 12/06/2009

à l'Ecole Centrale de Lyon

### Jury

Pr Rombaut Micèle, Gipsa, UJF, Grenoble  
Pr Mérialdo Bernard, Eurecom, Nice  
Pr Lambert Patrick, Polytech Savoie, Annecy  
Dr Csurka Gabriella, XRCE, Grenoble  
Dr Quenot Georges, LIP, Grenoble  
Pr Chen Liming, Liris, Ecully

Président  
Rapporteur  
Rapporteur  
Examineur  
Examineur  
Directeur

Contact : [liming.chen@ec-lyon.fr](mailto:liming.chen@ec-lyon.fr)

### Résumé

La classification d'images par le contenu visuel est un domaine particulièrement actif et difficile de l'analyse d'images. En n'imposant aucune restriction sur les images traitées, on se retrouve en effet face à un contenu qui peut être composite, ambigu et qui plus est acquis dans de mauvaises conditions. Aussi difficile qu'elle puisse paraître, cette activité pose pourtant très rarement des problèmes à un être humain qui, quelle que soit la complexité de l'image d'origine, parvient toujours très rapidement à une décision. Idéalement un système d'indexation automatique devrait permettre de rechercher des concepts dans une image hétérogène et de savoir détecter leur présence comme leur absence de manière non-mutuellement-exclusive. Notre objectif a d'abord été de nous inspirer de la performance de la classification humaine pour en tirer des procédés d'analyse nous mettant dans de bonnes conditions pour nous acquitter de cette tâche. Nous avons également déterminé des caractéristiques de forme pertinentes pour nous assister dans la tâche de classification. Enfin nous avons développé une classification efficace qui puisse s'adapter à ces conditions difficiles. Les contributions de cette thèse portent sur les informations extraites de l'image, le procédé d'extraction ainsi que sur leur utilisation pour accéder à un verdict de classification. Notre première contribution concerne les informations extraites de l'image. En nous inspirant des principes de la perception humaine, nous voulons exploiter le fait que les informations les plus importantes sont des informations singulières au sein d'une image. Ainsi nous basons notre descripteur de forme sur des segments, information visuellement bien plus significative que les gradients traditionnellement utilisés. Différentes formes de descripteurs sont explorées dans l'optique d'une classification globale (type d'image) comme d'une recherche d'objets visuels. Notre seconde contribution porte sur la réduction du nombre de couleurs au sein d'une image le but est de simplifier sans l'endommager l'information de couleur afin de pouvoir l'exploiter ultérieurement de manière efficace. En particulier le but est ici de permettre une segmentation efficace de l'image par clustering qui, sans réduction préalable, porterait sur des dizaines de milliers de couleurs. Pour éviter d'endommager l'image, nous avons utilisé la théorie de l'information afin de quantifier l'information qu'apportait une couleur par rapport aux autres. Ceci nous permet ainsi de repérer et de sélectionner des couleurs perceptuellement importantes car singulières. Notre troisième contribution est la continuation logique de la précédente : il s'agit d'un algorithme de segmentation d'image en régions de couleur homogène. Il s'agit d'un procédé en trois étapes. On effectue une réduction préalable du nombre de couleurs ainsi qu'un filtrage visant à améliorer sa robustesse. On effectue ensuite une détection rapide d'un nombre idéal de couleurs quantifiées suivi par une quantification par le procédé présenté dans le précédent paragraphe. Enfin nous opérons une séparation des régions spatialement disjointes et une fusion des régions trop petites ou trop proches de leurs voisines. Cet algorithme a pour principaux atouts sa robustesse et sa capacité à produire des régions de taille importantes à partir desquelles on peut envisager d'extraire des caractéristiques visuelles. Notre quatrième contribution est une

synthèse des précédentes : nous intégrons les éléments développés dans un système de classification automatique. Nous partons du constat que l'étude de la perception humaine suggère d'une part une domination de la perception globale sur la perception locale et d'autre part met en relief l'importance des relations entre différentes parties d'une image. Nous choisissons donc constituer une plateforme de classification se basant sur des informations extraites à partir de régions déterminées par notre algorithme de segmentation. Nous utilisons nos caractéristiques basées sur des segments, complétées par des caractéristiques visuelles obtenues à partir de descripteurs existants (couleur, texture...) auxquelles nous ajoutons des informations provenant des régions voisines. Ces informations, collectées sur des images d'entraînement, sont fusionnées pour la constitution d'un "vocabulaire visuel". Toute image est ensuite exprimée à partir de ce vocabulaire et peut dès lors être classée en utilisant un procédé d'apprentissage supervisé.

## **Abstract**

Image indexing based on visual content is an especially active and challenging field in image processing. Without any restriction on processed images, we indeed face contents which may be heterogeneous, ambiguous and also acquired in poor conditions. As difficult as it may appear, most of the time, this activity poses very few problems to human beings who always reach a quick classification decision, whichever the complexity of the original image. An automated indexing system should, ideally, allow searching for concepts within a heterogeneous image and being able to detect their presence as well as their absence in a non-mutually exclusive way. Our first objective was to draw means of processing information from human perception which would put us into good conditions to make a successful classification. We also devised efficient shape features to help us in this classification task. Finally, we developed an efficient classification process that could adapt to these difficult conditions. The contributions of this thesis are about the basic features extracted from the image, the extraction process itself as well as the classification process itself. Our first contribution is related to the information extracted from the image. The principles of human perception lead us to look for singular information as it is deemed as perceptually more important. As a consequence, we based our shape descriptor on line segments which are visually more significant than the usually used gradient values. Several feature structures are proposed aiming both at global classification (type of the image) and visual object categorization. Our second contribution is about image color reduction; its purpose is to simplify color information without damaging it in order to use it more efficiently. More specifically, our purpose here is to allow efficient color clustering which would otherwise involve tens of thousands of colors. To avoid damaging the original image, we used information theory in order to quantify the amount of information provided by a color compared to the others. This allows us to identify and select singular and therefore perceptually important colors. Our third contribution is in direct continuation of the previous one: it consists in a color based image segmentation algorithm. It is a three-step process. We first reduce the number of colors and enhance robustness to noise through filtering. We then determine an ideal number of quantized colors followed by the quantization itself using the process introduced in the previous paragraph. We finally separate regions which are not spatially connected and merge regions that are either too small or too similar to their neighbors. This algorithm's assets are mainly its robustness as well as its ability to produce large coarse regions from which we can extract visual features. Our fourth contribution is a synthesis of the previous ones: we integrate the elements we devised into an automated classification system. Basing on the study of human perception, we observe the precedence of global perception over local perception as well as the importance of the relationship between neighboring parts of an image. We therefore chose to build a classification platform built from information extracted from coarse regions provided by our segmentation algorithm. We use our line segment features combined to other visual characteristics provided by existing features (such as color, texture...) to which we add information from neighboring regions. This information, collected on a training set, is gathered to build a "visual vocabulary". All images are then described through this vocabulary and can be therefore classified using a supervised learning process.

## **Publications**

### **Conférences internationales avec comité de lecture et actes**

Region based Visual Object Categorization using Segment Features and Polynomial Modeling. huanzh Fu, A

Pujol, E Dellandréa, L. Chen. Dans Joint IAPR International Workshops on Structural, Syntactic and Statistical Pattern Recognition (S+SSPR 2008), . pp. 277-286. 2008.

Coarse Adaptive Color Image Segmentation for Visual Object Classification. A Pujol, L. Chen. Dans 15th International Conference on Systems, Signals and Image Processing, Bratislava. 2008.

Color Quantization For Image Processing Using Self Information. A Pujol, L. Chen. Dans international Conference on Information Communications and Signal Processing (ICICS), Singapore. 2007.

Hough Transform Based Cityscape Classifier. A Pujol, L. Chen. Dans 6th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), Montreux. pp. 00-00. 2005.

#### **Edition scientifique d'ouvrages**

Line Segment Based Edge Feature Using Hough transform. A Pujol, L. Chen. (The 7th IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP)), , Palma de Mallorca. 6p. 2007.

#### **Rapports de recherche/technique**

Études des critères de catégorisation d'images de scènes naturelles complexes. C. Métge, A Pujol, M. Ardabilian. Rapport de recherche RR-LIRIS-2009-022 2009.

Region based visual object categorization using segment features and polynomial image modeling. huanzh Fu, A Pujol, E Dellandréa, L. Chen. Rapport de recherche RR-LIRIS-2008-013 2008.



# Formalisme statistique pour ensembles de structures discrètes

Thèse

Sébastien Rebecchi

Soutenue le 23/09/2009

à l'Institut National des Sciences Appliquées de Lyon

## Jury

Pr Fleury Éric, LIP, Lyon	Président
Pr Brun Luc, GREYC, Caen	Rapporteur
Pr Lecroq Thierry, LITIS, Mont-Saint-Aignan	Rapporteur
Pr Bloch Isabelle, LTCL, Paris	Examineur
Pr de la Higuera Colin, LINA, Nantes	Examineur
Pr Jolion Jean-Michel, LIRIS, Villeurbanne	Directeur

Contact : [sebastien.rebecchi@liris.cnrs.fr](mailto:sebastien.rebecchi@liris.cnrs.fr)

## Résumé

En reconnaissance de formes, le codage de l'information extraite des données est une étape décisive, et l'utilisation de structures semble être le choix le plus pertinent, leur puissance de représentation semblant illimitée. Cependant, le codage sous forme de vecteurs de caractéristiques numériques offre l'avantage de permettre, par la suite, l'utilisation de nombreux algorithmes efficaces développés spécifiquement pour la classification de vecteurs numériques dans des disciplines connexes à la reconnaissance de formes (apprentissage automatique, inférence statistique...). De ce constat est né un champ de recherche dédié à la caractérisation statistique des espaces de structures. Les travaux les plus notables dans cette catégorie sont ceux basés sur la topologie induite par la distance d'édition. Cette approche se voit cependant assujettie à des problèmes de complexité qui pourraient théoriquement s'avérer insurmontables dans le paradigme calculatoire actuel. Parallèlement, ont vu le jour un ensemble des travaux basés sur une transformation des structures sous forme de vecteurs numériques. Le problème des méthodes proposés jusqu'à présent dans cette philosophie est le manque de pouvoir caractéristique, dans l'espace structurel initial, des informations extraites dans l'espace vectoriel. Enfin, les probabilités sont un paradigme largement utilisé pour la classification de données structurées, via la modélisation de distribution de structures au moyen de machines à états, ou autres graphes aléatoires, et l'utilisation massive du classifieur par maximisation de vraisemblance. Dans cette thèse, outre le passage en revue des méthodes précitées, nous nous concentrons sur le traitement probabiliste d'ensembles de structures discrètes. Nous proposons plus particulièrement la traduction aux espaces structurels de critères permettant de définir les notions statistiques d'uniformité et de normalité de lois de probabilités. Nous proposons également une réflexion sur la définition de variables aléatoires de structures à valeurs dans un espace vectoriel, avec pour perspective la possibilité d'application, dans le domaine structurel, du théorème central limite, résultat d'importance fondamentale en théorie des probabilités et statistique. D'un point de vue applicatif, nous évaluons les apports d'une partie de nos travaux pour la résolution de problèmes typiques en reconnaissance de formes, à savoir la classification de séquences d'ADN et la classification d'images de chiffres dessinés à la main. Nous utilisons le classifieur par maximisation de vraisemblance, en estimant la distribution de chaque classe par une loi normale de chaînes, telle que définie dans la partie théorique de cette thèse. Une conclusion négative tirée des expérimentations est notre manque de compétitivité vis-à-vis des méthodes les plus performantes sur chaque problème, notamment celles profitant de l'apport de connaissance experte biologique dans le cas des séquences d'ADN. Cependant, ce point nous donne des idées de perspectives de travail visant à améliorer notre classifieur, comme par exemple le fait de se concentrer plus sérieusement sur la phase d'apprentissage de la fonction de coût, trop simpliste dans cette thèse. Pour ce qui est du point positif, nous montrons qu'il est possible d'améliorer les résultats obtenus par les classifieurs basés sur la distance d'édition, très utilisée en reconnaissance de formes structurelle.

Notre classifieur obtient des résultats honorables même lorsque la fonction de coût n'est pas très appropriée au problème, ce grâce à l'apport du cadre probabiliste qui permet de se détacher en partie de l'influence de cette fonction. Ce n'est pas le cas pour le classifieur à la plus proche médiane et le classifieur aux  $k$  plus proches voisins, qui échouent fortement lorsqu'ils sont basés sur la distance d'édition selon la même fonction de coût.

## **Abstract**

In pattern recognition, the coding of information extracted from the data is a decisive phase, and the use of structures seems to be the most pertinent choice, since their representative power seems to be unlimited. However, the coding in the form of numeric feature vectors offers the advantage of enabling, in the sequel, the use of numerous efficient algorithms specifically developed for the classification of numeric vectors in fields connected to pattern recognition (machine learning, statistical inference ...). From this observation was born a research field devoted to the statistical characterization of structure spaces. The most notable works in this category are the ones based on the topology induced by the edit distance. This approach is yet subject to complexity problems that could theoretically be unsolvable within the current computational paradigm. At the same time have been developed a set of works based on a transformation of structures into numeric vectors. The problem of the methods proposed as yet within this philosophy is the lack of characteristic power, in the initial structure space, of information extracted in the vector space. Finally, probabilities are a widely used paradigm for the classification of structured data, via the modelling of distributions of structures by the way of state machines, or other random graphs, and a massive use of the maximum likelihood classifier. In this thesis, in addition to reviewing the precited methods, we concentrate on the probabilistic processing of sets of structures. We propose more particularly the translation to structure spaces of criteria enabling to define the statistical notions of uniformity and normality of probability laws. We propose as well a reflection on the definition of structural random variables taking their values in a vector space, having in prospect the possibility of applying, in the structural domain, the central limit theorem, a fundamentally important result in probability theory and statistics. From an applicative point of view, we evaluate the contribution of a part of our work for the resolution of typical problems in pattern recognition, namely the classification of DNA sequences and the classification of images of handwritten digits. We use the maximum likelihood classifier, estimating the distribution of each class by a Gaussian distribution, as defined in the theoretical part of this thesis. A negative conclusion drawn from these experimentations is our lack of competitiveness regarding the most reliable methods on each problem, in particular the ones that take advantage of biological expert knowledge in the case of DNA sequences. However, this point gives us insights for a future work aiming at improving our classifier, such as the need to concentrate more seriously on the cost function learning phase, that is too simplistic in this thesis. As for the positive point, we show that it is possible to improve the results obtained by the classifiers based on the edit distance, widely used in structural pattern recognition. Our classifier obtains honourable results even when the cost function is not well adapted to the problem, this thanks to the contribution of the probabilistic framework that reduces the influence of this function. This is not the case for the nearest median classifier and the  $k$  nearest neighbors classifier, both of them hardly failing when being based on the edit distance with respect to the same cost function.

## **Publications**

### **Conférences internationales avec comité de lecture et actes**

On the Gaussian distribution of strings. S Rebecchi, JM Jolion. Dans ICPR, Tampa, USA. 2008.

Conférences nationales avec comité de lecture et actes

Lois uniformes et normales de chaînes discrètes. S Rebecchi, JM Jolion. Dans RFIA, Amiens, France. pp. 471-480. 2008.

### **Rapports de recherche/technique**

Distribution uniforme de structures. S Rebecchi, JM Jolion. Rapport de recherche RR-LIRIS-2008-020 2008.

General statistical framework: draft for the proposition of a possible theory. S Rebecchi. Rapport de recherche RR-LIRIS-2007-031 2007.

**Lionel Robinault**

Soutenue le 08/09/2009

à l'Université Lyon2

### Jury

Mme Rombaut Michèle, GIPSA lab, Grenoble	Rapporteur
Mme Vincent Nicole , Crip5, Paris	Rapporteur
M. Chateau Thierry , LASMEA, Clermont-Ferrand	Examineur
M. Perez Patrick , IRISA, Rennes	Président
M. Miguet Serge, LIRIS, Lyon	Directeur
M. Bres Stéphane, LIRIS, Lyon	Co-directeur
M. Ducatez Jean-Baptiste, Foxstream, Lyon	Examineur

Contact : [lionel.robinault@liris.cnrs.fr](mailto:lionel.robinault@liris.cnrs.fr)

### Résumé

Le travail de thèse que nous présentons s'articule autour de l'utilisation de caméras motorisées à trois degrés de liberté, également appelées caméras PTZ. Ces caméras peuvent être pilotées suivant deux angles. L'angle de panorama permet une rotation autour d'un axe vertical et l'angle de tangage permet une rotation autour d'un axe horizontal. Si, théoriquement, ces caméras permettent donc une vue omnidirectionnelle, elles limitent le plus souvent la rotation suivant l'angle de panorama mais surtout suivant l'angle de tangage. En plus du pilotage des rotations, ces caméras permettent également de contrôler la distance focale permettant ainsi un degré de liberté supplémentaire. Par rapport à d'autres modèles, les caméras PTZ permettent de construire un panorama - représentation étendue d'une scène construite à partir d'une collection d'images - de très grande résolution. La première étape dans la construction d'un panorama est l'acquisition des différentes prises de vue. A cet effet, nous avons réalisé une étude théorique permettant une couverture optimale de la sphère à partir de surfaces rectangulaires en limitant les zones de recouvrement. Cette étude nous permet de calculer une trajectoire optimale de la caméra et de limiter le nombre de prises de vues nécessaires à la représentation de la scène. Nous proposons également différents traitements permettant d'améliorer sensiblement le rendu et de corriger la plupart des défauts liés à l'assemblage d'une collection d'images acquises avec des paramètres de prises de vue différents. Une part importante de notre travail a été consacrée au recalage automatique d'images en temps réel, c'est à dire que chaque étape est effectuée en moins de 40ms pour permettre le traitement de 25 images par seconde. La technologie que nous avons développée permet d'obtenir un recalage particulièrement précis avec un temps d'exécution de l'ordre de 4ms (AMD1.8MHz). Enfin, nous proposons deux applications de suivi d'objets en mouvement directement issues de nos travaux de recherche. La première associe une caméra PTZ à un miroir sphérique. L'association de ces deux éléments permet de détecter tout objet en mouvement dans la scène puis de se focaliser sur l'un d'eux. Dans le cadre de cette application, nous proposons un algorithme de calibrage automatique de l'ensemble caméra et miroir. La deuxième application n'exploite que la caméra PTZ et permet la segmentation et le suivi des objets dans la scène pendant le mouvement de la caméra. Par rapport aux applications classiques de suivi de cible en mouvement avec une caméra PTZ, notre approche se différencie par le fait que nous réalisons une segmentation fine des objets permettant leur classification.

### Abstract

The thesis considers the use of motorized cameras with 3 degrees of freedom which are commonly called PTZ cameras. The orientation of such cameras is controlled according to two angles: the panorama angle describes the degree of rotation around the vertical axis and the tilt angle refers to rotation along a

meridian line. Theoretically, these cameras can cover an omnidirectional field of vision of 4psr. Generally, the panorama angle and especially the tilt angle are limited for such cameras. In addition to control of the orientation of the camera, it is also possible to control focal distance, thus allowing an additional degree of freedom. Compared to other material, PTZ cameras thus allow one to build a panorama of very high resolution. A panorama is a wide representation of a scene built starting from a collection of images. The first stage in the construction of a panorama is the acquisition of the various images. To this end, we made a theoretical study to determine the optimal paving of the sphere with rectangular surfaces to minimize the number of zones of recovery. This study enables us to calculate an optimal trajectory of the camera and to limit the number of images necessary to the representation of the scene. We also propose various processing techniques which appreciably improve the rendering of the mosaic image and correct the majority of the defaults related to the assembly of a collection of images which were acquired with differing image capture parameters. A significant part of our work was used to the automatic image registration in real time, i.e. lower than 40ms. The technology that we developed makes it possible to obtain a particularly precise image registration with an computation time about 4ms (AMD1.8MHz). Our research leads directly to two proposed applications for the tracking of moving objects. The first involves the use of a PTZ camera and a spherical mirror. The combination of these two elements makes it possible to detect any motion object in the scene and to then to focus itself on one of them. Within the framework of this application, we propose an automatic algorithm of calibration of the system. The second application exploits only PTZ camera and allows the segmentation and the tracking of the objects in the scene during the movement of the camera. Compared to the traditional applications of motion detection with a PTZ camera, our approach is different by the fact that it compute a precise segmentation of the objects allowing their classification.

## **Publications**

### **Conférences internationales avec comité de lecture et actes**

Self-calibration and control of a PTZ camera based on a spherical mirror. R.L Robinault, I. Pop, S. Miguet. Dans 6th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS 2009), IEEE, Signal Processing Society ed. Genova, Italy. 2009. (à paraître)

Real time foreground object detection using PTZ camera. R.L Robinault, S. Bres, S. Miguet. Dans International Conference on Computer Vision, Theory and Applications (VISAPP'09), Lisboa, Portugal. pp. 609-614. 2009.

Panoramic mosaicing optimization. R.L Robinault, S. Bres, S. Miguet. Dans International Conference on Image Analysis and Processing, Modena, Italy, 10-14 september 2007. pp. 548-553. 2007.

# Algorithmes d'extraction de modèles géométriques discrets pour la représentation robuste des formes

Tristan Roussillon

Thèse

Soutenue le 19/11/2009

à l'Université Lumière Lyon 2

## Jury

M. Eckhardt Ulrich, Université de Hambourg, Hambourg, Allemagne	Rapporteur
M. Lachaud Jacques-Olivier, LAMA, Chambéry	Rapporteur
M. Blanc-Talon Jacques, DGA, Bagneux	Examineur
M. Chassery Jean-Marc, Gipsa-Lab, Grenoble	Président
M. Olivier Devillers, INRIA - Sophia-Antipolis, Sophia-Antipolis	Examineur
Mlle Sivignon Isabelle, LIRIS, Lyon	Co-directeur
Mme Tougne Laure, LIRIS, Lyon	Directeur

Contact : [tristan.roussillon@liris.cnrs.fr](mailto:tristan.roussillon@liris.cnrs.fr)

## Résumé

Cette thèse se situe à l'interface entre l'analyse d'images, dont l'objectif est la description automatique du contenu visuel, et la géométrie discrète, qui est l'un des domaines dédiés au traitement des images numériques. Pour être stocké et manipulé sur un ordinateur, un signal observé est régulièrement échantillonné. L'image numérique, qui est le résultat de ce processus d'acquisition, est donc constituée d'un ensemble fini d'éléments distincts. La géométrie discrète se propose d'étudier les propriétés géométriques d'un tel espace dépourvu de continuité. Dans ce cadre, nous avons considéré les régions homogènes et porteuses de sens d'une image, avec l'objectif de représenter leur contour au moyen de modèles géométriques ou de les décrire à l'aide de mesures. L'étendue des applications de ce travail en analyse d'images est vaste, que ce soit au cours du processus de segmentation, ou en vue de la reconnaissance d'un objet. Nous nous sommes concentrés sur trois modèles géométriques discrets définis par la discrétisation de Gauss : la partie convexe ou concave, l'arc de cercle discret et le segment de droite discrète. Nous avons élaboré des algorithmes dynamiques (mise à jour à la volée de la décision et du paramétrage), exacts (calculs en nombres entiers sans erreur d'approximation) et rapides (calculs simplifiés par l'exploitation de propriétés arithmétiques et complexité en temps linéaire) qui détectent ces modèles sur un contour. L'exécution de ces algorithmes le long d'un contour aboutit à des décompositions ou à des polygonalisations réversibles. De plus, nous avons défini des mesures de convexité, linéarité et circularité, qui vérifient un ensemble de propriétés fondamentales : elles sont robustes aux transformations rigides, elles s'appliquent à des parties de contour et leur valeur maximale est atteinte pour le modèle de forme qui sert de comparaison et uniquement sur celui-ci. Ces mesures servent à l'introduction de nouveaux modèles dotés d'un paramètre variant entre 0 et 1. Le paramètre est fixé à 1 quand on est sûr de la position du contour, mais fixé à une valeur inférieure quand le contour est susceptible d'avoir été déplacé par un bruit d'acquisition. Cette approche pragmatique permet de décomposer de manière robuste un contour en segments de droite ou en parties convexes et concaves.

## Abstract

The work presented in this thesis concerns the fields of image analysis and discrete geometry. Image analysis aims at automatically describing the visual content of a digital image and discrete geometry provides tools devoted to digital image processing. A two-dimensional analog signal is regularly sampled in order to be handled on computers. This acquisition process results in a digital image, which is made up of a finite set of discrete elements. The topic of discrete geometry is to study the geometric properties of such kind of discrete spaces. In this work, we consider homogeneous regions of an image having a meaning for a user. The objective is to represent their digital contour by means of geometric patterns and compute

measures. The scope of applications is wide in image analysis. For instance, our results would be of great interest for segmentation or object recognition. We focus on three discrete geometric patterns defined by Gauss digitization: the convex or concave part, the digital straight segment and the digital circular arc. We present several algorithms that detect or recognize these patterns on a digital contour. These algorithms are on-line, exact (integer-only computations without any approximation error) and fast (simplified computations thanks to arithmetic properties and linear-time complexity). They provide a way for segmenting a digital contour or for representing a digital contour by a reversible polygon. Moreover, we define a measure of convexity, a measure of straightness and a measure of circularity. These measures fulfil the following important properties: they are robust to rigid transformations, they may be applied on any part of a digital contour, they reach their maximal value for the template with which the data are compared to. From these measures, we introduce new patterns having a parameter that ranges from 0 to 1. The parameter is set to 1 when the localisation of the digital contour is reliable, but is set to a lower value when the digital contour is expected to have been shifted because of some acquisition noise. This measure-based approach provides a way for robustly decomposing a digital contour into convex, concave or straight parts.

## **Publications**

### **Revues internationales avec comité de lecture**

Automatic computation of pebble roundness using digital imagery and discrete geometry . T. Roussillon, H. Piégay, I. Sivignon, L Tougne, F. Lavigne. *Computers and Geosciences* (). 2009. (à paraître)

Measure of Circularity for Digital Curves and its Fast Computation. T. Roussillon, I. Sivignon, L Tougne. *Pattern Recognition* (). 2009. (à paraître)

### **Conférences internationales avec comité de lecture et actes**

What Does Digital Straightness Tell About Digital Convexity. T. Roussillon, L Tougne, I. Sivignon. Dans 13th International Workshop on Combinatorial Image Analysis (IWCI 2009), Cancun, Mexique. LNCS . Springer . 2009. (à paraître)

On-line Recognition of Digital Circular Arcs. T. Roussillon, I. Sivignon, L Tougne. Dans 15-th IAPR International Conference on Discrete Geometry for Computer Imagery (DGCI), Montreal (CANADA). LNCS . Springer . 2009. (à paraître)

Robust Decomposition of a Digital Curve into Convex and Concave Parts. T. Roussillon, I. Sivignon, L Tougne. Dans International Conference on Pattern Recognition (ICPR 2008), Tampa (Etats-Unis). 2008.

Computation of Binary Objects Sides Number using Discrete Geometry, Application to Automatic Pebbles Shape Analysis . T. Roussillon, I. Sivignon, L Tougne. Dans 14th International Conference on Image Analysis and Processing (ICIAP 2007), IEEE Computer Society ed. Modena (Italy). pp. 763-768. 2007.

### **Rapports de recherche/technique**

Discrete Circularity Measure. T. Roussillon, I. Sivignon, L Tougne. Rapport de recherche RR-LIRIS-2007-029 2007.

On-Line Recognition of Digital Arcs. T. Roussillon, I. Sivignon, L Tougne. Rapport de recherche RR-LIRIS-2009-008 , Soumis à DGCI 2009. 2009.

Measure of Circularity for Digital Curves and its Fast Computation. T. Roussillon, I. Sivignon, L Tougne. Rapport de recherche RR-LIRIS-2008-019 , Soumis à Pattern Recognition. 2008.

# Quantization-Based Blind Watermarking of Three-Dimensional Meshes

Kai Wang

Thèse

Soutenue le 06/11/2009

à l'INSA de Lyon

## Jury

M. Barni Mauro, Università di Siena, Siena, Italy	Rapporteur
M. Lévy Bruno, INRIA Nancy Grand Est, Nancy	Rapporteur
M. Macq Benoît, Université Catholique de Louvain, Louvain, Belgium	Rapporteur
M. Chassery Jean-Marc, GIPSA-lab, Grenoble	Président
Mme Fontaine Caroline, IRISA Rennes, Rennes	Examineur
M. Baskurt Atilla, LIRIS, Villeurbanne	Directeur
Mme Denis Florence, LIRIS, Villeurbanne	Co-directeur
M. Lavoué Guillaume, LIRIS, Villeurbanne	Co-directeur

Contact : kai.wang@insa-lyon.fr

## Résumé

Avec l'accroissement de l'utilisation des objets tridimensionnels (3D) dans diverses applications, une attention de plus en plus forte a été portée sur la protection de ce contenu 3D par des techniques de tatouage numérique. Dans ce travail de thèse, nous avons réalisé d'abord un état de l'art complet sur le domaine du tatouage de maillages, avec un point de vue original centré sur les attaques. Puis, nous avons proposé plusieurs méthodes de tatouage aveugles basées sur des techniques de quantification par le schéma de Costa scalaire (SCS). Nous avons choisi différents domaines appropriés d'un maillage 3D pour l'insertion de tatouages aveugles : le domaine ondelettes d'un maillage semi-régulier, et les domaines spatial et spectral d'un maillage général. Les primitives de tatouage, qui sont soumises à la quantification scalaire, sont respectivement les normes et les orientations des vecteurs de coefficients d'ondelettes, les moments volumiques analytiques et les amplitudes spectrales en harmoniques variétés. Enfin, nous avons conçu et implémenté un système de benchmark pour les techniques de tatouage robustes de maillages. Ce benchmark est accessible librement sur Internet et contient une collection de modèles 3D, un outil logiciel et deux protocoles d'évaluation orientés sur différentes applications. Les méthodes de tatouage robustes proposées dans ce manuscrit ainsi qu'une méthode récente de l'état de l'art sont comparées grâce à ce benchmark. Les résultats obtenus montrent l'efficacité des méthodes proposées ainsi que la pertinence du système de benchmark.

## Abstract

With the increasing use of three-dimensional (3-D) models in various practical applications, more and more attention has been paid on the research of digital watermarking techniques for 3-D polygonal meshes. In this thesis, we first provide a comprehensive survey on the state of the art in 3-D mesh watermarking, with an original attack-centric investigation. Then, we make use of the scalar Costa quantization scheme to construct a number of effective blind mesh watermarking schemes. We successfully embed multi-bit quantization-based blind watermarks in three different mesh domains: the wavelet domain of a semi-regular mesh, and the spatial and spectral domains of a general mesh. The watermarking primitives, which are subject to scalar Costa quantization, are respectively the norms and orientations of the wavelet coefficient vectors, the analytic volume moments and the manifold harmonics spectral amplitudes. Finally, we detail the design and implementation of a robust mesh watermark benchmarking system, which has been made publicly available on-line. This benchmarking system comprises a standard mesh data set, a software tool and two application-oriented evaluation protocols. The robust mesh watermarking schemes proposed in this thesis and a state-of-the-art method are compared within this benchmarking framework.

The comparison results demonstrate both the effectiveness of our blind watermarking schemes and the relevance of our benchmarking system.

## **Publications**

### **Revue internationale avec comité de lecture**

Hierarchical watermarking of semiregular meshes based on wavelet transform. K. Wang, G Lavoué, F Denis, A. Baskurt. IEEE Transactions on Information Forensics and Security 3(4):620-634, IEEE. 2008.

A comprehensive survey on three-dimensional mesh watermarking. K. Wang, G Lavoué, F Denis, A. Baskurt. IEEE Transactions on Multimedia 10(8):1513-1527, IEEE. 2008.

### **Conférences internationales avec comité de lecture et actes**

A fragile watermarking scheme for authentication of semi-regular meshes. K. Wang, G Lavoué, F Denis, A. Baskurt. Dans Eurographics Short Papers, Crete, Greece. pp. 5-8. Eurographics Association . 2008.

Hierarchical blind watermarking of 3D triangular meshes. K. Wang, G Lavoué, F Denis, A. Baskurt. Dans IEEE International Conference on Multimedia & Expo, Beijing, China. pp. 1235-1238. IEEE . 2007.

Three-dimensional meshes watermarking: Review and attack-centric investigation. K. Wang, G Lavoué, F Denis, A. Baskurt. Dans International Workshop on Information Hiding, Saint-Malo, France. pp. 50-64. Lecture Notes in Computer Science . Springer-Verlag . 2007.

Blind and robust mesh watermarking using manifold harmonics. K. Wang, M. Luo, A. Bors, F Denis. Dans IEEE International Conference on Image Processing, Cairo, Egypt. pp. 3657-3660. IEEE . 2009.

Local patch blind spectral watermarking method for 3D graphics. M. Luo, K. Wang, A. Bors, G Lavoué. Dans International Workshop on Digital Watermarking, Surrey, UK. pp. 211-226. Lecture Notes in Computer Science . Springer-Verlag . 2009.

### **Conférences nationales avec comité de lecture et actes**

Tatouage hiérarchique et aveugle de maillages tridimensionnels. K. Wang, G Lavoué, F Denis, A. Baskurt. Dans CORESA'07 : COMpression et REprésentation des Signaux Audiovisuels, Montpellier, France. pp. 139-143. 2007.

Tatouage robuste et aveugle de maillages 3D basé sur les moments volumiques. K. Wang, G Lavoué, F Denis, A. Baskurt. Dans CORESA'09 : COMpression et REprésentation des Signaux Audiovisuels, Toulouse, France. pp. 162-167. 2009.

### **Contributions à un ouvrage**

Blind watermarking of three-dimensional meshes: Review, recent advances and future opportunities. K. Wang, G Lavoué, F Denis, A. Baskurt. Advanced Techniques in Multimedia Watermarking: Image, Video and Audio Applications 2010. (à paraître)

### **Rapports de recherche/technique**

A Benchmark for 3-D Mesh Watermarking. K. Wang, G Lavoué, F Denis, A. Baskurt. Rapport de recherche RR-LIRIS-2009-018 2009.

A comprehensive survey on three-dimensional mesh watermarking. K. Wang, G Lavoué, F Denis, A. Baskurt. Rapport de recherche RR-LIRIS-2008-005 2008.

Three-Dimensional Meshes Watermarking: Review and Attack-Centric Investigation. K. Wang, G Lavoué, F Denis, A. Baskurt. Rapport de recherche RR-LIRIS-2007-008 , Soumis à Conférence Information Hiding 2007 (IH'07). 2007.