acyclique. Dans l'absolu, on veut un FAS minimal, mais ce problème est NP-difficile. Nous adaptons et améliorons une heuristique proposée par Eades et al. qui permet une construction en temps linéaire. Nous montrons comment cet algorithme bénéficie à la complémentation d'automates de Büchi déterministes et la traduction d'automates de Rabin en automates de Büchi. En fonction de l'automate traité on remarque une amélioration montant jusqu'à 31%. Ces résultats varient beaucoup selon le nombre de cycles et d'états acceptant.

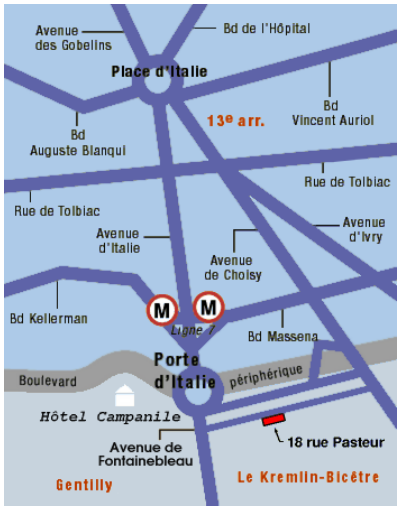
Quatre stagiaires internationaux au LRDE

Le LRDE est heureux d'accueillir trois étudiants indiens de notre partenaire l' Indian Institute of Technology Jodhpur [2] et un étudiant vietnamien de l'Institut de la Francophonie pour l'Informatique de Hanoi [3].

Divya Grover (Bachelor), Rupali et Divya Sharma (Master) travaillent avec l'équipe Image.

Duc Canh Luu (Master) a intégré l'équipe Vaucanson.

Contacter le LRDE



18, rue Pasteur
Paritalie, bâtiment X, aile Mistral
2e étage, droite droite

Tél. : 01 53 14 59 22
Fax : 01 53 14 59 13
Contact : info@lrde.epita.fr
Les permanents : lrde@lrde.epita.fr
Site Web : http://www.lrde.epita.fr

… et surtout, passez nous voir ;
vous serez toujours les bienvenus !

2. Indian Institute of Technology Jodhpur, http://www.iitj.ac.in/.
3. Institut de la Francophonie pour l'Informatique de Hanoi, http://www.ifi.vnu.edu.vn/.

L'air de rien N° 31

Séminaire CSI de juillet 2014

L'aléastriel du **L**aboratoire de **R**echerche et de **D**éveloppement de l'**EPITA** [1] Numéro 31, Juillet 2014

Édito

par *Valentin Tolmer (CSI 2016)*

Ce numéro "Séminaire CSI" a pour objectif de présenter le séminaire de fin d'année des étudiants du LRDE. Ainsi tous les ans, nous exposons notre travail de recherche du semestre sur nos projets respectifs. Cette fois, les présentations de tous les élèves de première et de deuxième année se dérouleront sur toute la journée du 10 juillet.

Pour nous, ces présentations sont importantes, car elles nous permettent non seulement de partager nos travaux et les résultats obtenus, mais aussi elles sont le point culminant d'une expérience nouvelle pour certains et enrichissante pour tous. Aussi nous vous attendons nombreux et intéressés, pour pouvoir partager ce qu'on a découvert tout au long du semestre.

Programme du séminaire CSI du 10 juillet 2014

Matin

10 juillet 10h30–13h15 — amphi 3 KB	
Vaucanson	
10h30	Composition de transducteurs dans Vaucanson 2 — *Valentin Tolmer*
11h00	Recherche de mots synchronisants dans un DFA — *Antoine Pietri*
Olena	
11h30	Détection de logotypes et autres invariants caractéristiques à l'aide de descripteurs SIFT — *Anthony Seure*
12h15	Inpainting d'images couleur — *Nicolas Allain*
12h45	Utilisation du clustering pour la reconnaissance de caractères — *Antoine Lecubin*

Après-midi

10 juillet 14h30–17h15 — amphi 3 KB	
Speaker ID	
14h30	Regroupement de locuteur à base de Self-Organizing Map — *Jean-Luc Bounthong*
15h00	Le partitionnement de Newman-Girvan pour les systèmes de reconnaissance du locuteur — *Jimmy Yeh*
15h30	L'algorithme de partitionnement de Markov pour les systèmes de reconnaissance du locuteur — *Fanny Riols*
16h15	Détection de communautés avec l'algorithme de l'Infomap — *Mickael Saada*
Spot	
16h45	Un Feedback Arc Set pour Spot — *Alexandre Lewkowicz*

1. L'air de rien, http://publis.lrde.epita.fr/LrdeBulletin.

Vaucanson

Composition de transducteurs dans Vaucanson 2

par *Valentin Tolmer* *10 juillet, 10h30, Amphi 3*

Les transducteurs sont utilisés dans beaucoup de domaines, comme par exemple en linguistique pour modéliser des règles phonologiques, pour les expressions régulières, pour des languages de spécification, pour de la reconnaissance vocale... Quand on les manipule, un outil pour le moins indispensable est la composition. En tant que tel, il est essentiel de l'implémenter dans Vaucanson, de manière efficace. Ce rapport va présenter les fondations sur lesquelles s'appuie la composition de transducteurs, puis son implémentation et son optimisation. La composition est considérée ici comme un cas particulier du produit d'automates à transitions spontanées, donc trois algorithmes de produit sont présentés ici, suivi de concepts d'implémentation essentiels.

Recherche de mots synchronisants dans un DFA

par *Antoine Pietri* *10 juillet, 11h00, Amphi 3*

Le problème de recherche de mots synchronisants les plus courts possibles est un problème important qui a beaucoup d'applications (orienteurs mécaniques, problème de coloration de route, vérification de modèles, bioinformatique, protocoles réseaux, etc.) Un **mot synchronisant** (ou une **séquence de réinitialisation**) pour un automate fini déterministe est une séquence d'étiquettes qui envoie n'importe quel état de l'automate d'entrée à un seul et même état. Trouver le plus court mot synchronisant dans un automate général est NP-complet, c'est pourquoi la plupart des algorithmes sont des heuristiques qui cherchent à trouver des mots les plus courts possibles en temps polynomial. Dans cet exposé, nous comparerons les différentes approches utilisées par les algorithmes principaux les plus connus (Glouton et Cycle, SynchroP et SynchroPL), en termes de complexité spatiale et temporelle, et les résultats effectifs (longueur moyenne des mots trouvés, temps utilisé par l'algorithme en moyenne). Nous discuterons aussi des différentes représentations intermédiaires utilisées par ces algorithmes, et comment utiliser les informations qu'elles contiennent.

Olena

Détection de logotypes et autres invariants caractéristiques à l'aide de descripteurs SIFT

par *Anthony Seure* *10 juillet, 11h30, Amphi 3*

La détection d'éléments discriminants dans une image est un sujet très actif de vision par ordinateur. Aujourd'hui, les applications sont très diverses, allant de la robotique à la photographie numérique assistée. Notre exposé se concentrera sur la détection de logotypes dans des images naturelles. Pour ce faire, nous nous basons sur Olena, une plateforme libre, générique et performante de traitement d'images afin d'implémenter un détecteur de points-clés : les descripteurs SIFT.

Inpainting d'images couleur

par *Nicolas Allain* *10 juillet, 12h15, Amphi 3*

L'inpainting est une technique de traitement d'images. Son but est de reconstruire une zone d'une image sans connaître l'aspect originel de ladite image. Il existe deux grandes familles de méthodes d'inpainting : l'une est basée sur la prolongation des zones à fort contraste, l'autre sur la synthèse de textures. Ce procédé est utile dans le cas de sous-titres incrustés dans une vidéo ou de texte dans une image. Un inpainting est également utilisé sur des images dégradées, abîmées, ou présentant un objet non désiré. Notre but est de reconstruire le fond caché par du texte précédemment segmenté. Nous étudions la méthode de Khodadadi basée sur l'ordre dans lequel les pixels dégradés de l'image sont reconstruits via une synthèse de texture. Nous utilisons le critère de l'erreur quadratique moyenne pour évaluer nos résultats. Nous analysons les résultats obtenus, les améliorations effectuées, et les progrès obtenus.

Utilisation du clustering pour la reconnaissance de caractères

par *Antoine Lecubin* *10 juillet, 12h45, Amphi 3*

Nous présenterons une méthode d'amélioration de la classification dans l'application de reconnaissance de caractères du laboratoire basée sur le groupement des caractères en classes. Nous verrons d'abord comment déterminer et imbriquer les groupes de caractères les plus intéressants à classer ensemble grâce à un algorithme de clustering appliqué aux données fournies par le descripteur à base d'ondelettes. Puis nous nous pencherons sur l'adaptation de la phase de classification pour qu'elle s'effectue en plusieurs temps et prenne en compte ces groupements. Enfin nous nous intéresserons à l'évolution du taux de reconnaissance obtenue grâce à cette méthode.

Speaker ID

Regroupement de locuteur à base de Self-Organizing Map

par *Jean-Luc Bounthong* *10 juillet, 14h30, Amphi 3*

Les i-vectors représentent actuellement l'état de l'art dans le domaine de la vérification du locuteur. Des résultats intéressants sont obtenus à partir de classifieurs tel que la distance cosinus (CD). Cependant, un tel classifieur nécessite un apprentissage supervisé et devient donc inutilisable avec une base de données sans étiquette. Dans cette étude, nous explorerons une méthode à base de Self-Organizing Map (SOM) pour étiqueter une base de données quelconque. L'objectif étant de fournir une méthode pour entraîner les classifieurs supervisés sur une base de données non étiquetée. Nous allons aussi comparer l'efficacité de notre méthode avec d'autres méthodes telles que Infomap, Markov Clustering et Girvan-Newman.

Le partitionnement de Newman-Girvan pour les systèmes de reconnaissance du locuteur

par *Jimmy Yeh* *10 juillet, 15h00, Amphi 3*

La distance cosinus est la méthode de décision la plus utilisée, elle nécessite une base d'apprentissage étiquetée afin d'appliquer des algorithmes supervisés. Pour augmenter la taille de la base de développement et dans le but de réduire les coûts de constitution de ces bases, l'utilisation de données non étiquetées pour l'entraînement du système devient nécessaire. Le but de ce travail est de tester le partitionnement de Newman-Girvan afin d'étiqueter les i-vectors inconnus.

L'algorithme de partitionnement de Markov pour les systèmes de reconnaissance du locuteur

par *Fanny Riols* *10 juillet, 15h30, Amphi 3*

Grâce à des méthodes d'apprentissage supervisé (la Distance Cosinus avec l'Analyse Discriminante Linéaire et la méthode Within Class Covariance), des progrès significatifs ont été réalisés dans ce domaine. Cependant, de récentes recherches proposent d'utiliser une base de données non étiquetées d'i-vectors, afin d'augmenter la taille de l'ensemble des données d'entraînement et de réduire le coût de constitution de cette base. C'est pourquoi nous basons notre étude sur l'espace des i-vectors, et travaillons ainsi avec des méthodes d'apprentissage non supervisé. Dans cette étude, nous utilisons une méthode de partitionnement, le processus de Markov Clustering (MCL), afin de regrouper de façon naturelle les i-vectors qui représentent un même locuteur dans un ensemble d'entités. L'algorithme MCL est un algorithme de partitionnement non supervisé rapide et extensible, basé sur la simulation de flux stochastiques dans les graphes. Le résultat du partitionnement est utilisé dans le système supervisé standard de vérification du locuteur pour évaluer les performances. Nous allons aussi comparer celles-ci avec d'autres méthodes de regroupement, comme l'Infomap, le Self-Organizing Map et Girvan Newmann.

Détection de communautés avec l'algorithme de l'Infomap

par *Mickaël Saada* *10 juillet, 16h15, Amphi 3*

Les récentes recherches sur les i-vectors proposent d'utiliser des données non étiquetées pour augmenter la taille des données d'entraînement et par la même occasion réduire les coûts de la collecte de ces données. C'est pourquoi nous basons notre étude dans l'espace des i-vectors et travaillons sur des méthodes non supervisées. Dans cette étude, nous explorerons la méthode de l'Infomap qui permet de trouver les structures communautaires dans des réseaux afin d'étiqueter des données non étiquetées. Le but de cette méthode est de maximiser la modularité d'un graphe en associant les paires de sommets qui maximisent la modularité. Cet algorithme est divisé en 3 parties : un algorithme "glouton", le recuit simulé et l'algorithme du heat-bath. Nous allons aussi comparer l'efficacité de notre méthode avec d'autres méthodes telles que Markov Clustering, Grivan-Newman et Self-Organizing Map.

Spot

Un Feedback Arc Set pour Spot

par *Alexandre Lewkowicz* *10 juillet, 16h45, Amphi 3*

Spot est une bibliothèque extensible pour le model checking qui utilise les automates de Büchi généralisés à transitions acceptantes. Il contient de nombreux algorithmes de pointes. Dans cet exposé, nous nous concentrons sur deux de ses algorithmes qui construisent des automates avec plus de transitions que nécessaire. En pratique ces constructions utiliseraient moins de transitions si elles pouvaient calculer un feedback arc set (FAS), c'est-à-dire un ensemble de transitions à retirer du graphe pour le rendre