Institut de la Francophonie pour l'Informatique

Travail Personnel Encadré

Module Maste1, promotion 18
année 2013/2014

Liste de sujets

1. ALLOCATION DE FRÉQUENCES DANS LES RÉSEAUX RADIO MULTI-FRÉQUENCE

Encadrement : Anthony Busson (anthony.busson@inria.fr), UCBL1
Victor Moraru (victor.moraru@auf.org), MSI-IFI

Contexte
Les réseaux modernes sont de plus en plus orientés vers la mobilité et l'autonomie des nœuds et, comme conséquence, vers l'utilisation des technologies de communication sans fil tels que wifi, bluetooth, ZegBee, wimax, etc. Le problème qu'on observe dans de tels réseaux qui partagent le même médium de communication (ils utilisent tous la même fréquence pour communiquer) est que, avec l'augmentation du nombre d'utilisateurs l'accès au canal devient de plus en plus difficile et que la capacité disponible pour chaque utilisateur diminue progressivement. L'utilisation de plusieurs fréquences pourrait être une solution à ce problème permettant d'augmenter les performances du réseau (en terme de débit, délais, etc.) mais, en même temps, imposer une algorithmique plus complexe pour la gestion de fréquences. En plus, des communications en parallèle devraient renforcer la robustesse du réseau, optimiser l'utilisation de la bande passante, économiser de l'énergie, etc..

Travail théorique
Dans un premier temps on vous propose de faire une recherche bibliographique sur les technologies de communication sans fil multi-fréquence dans les réseaux informatiques, étudier leurs limites et leur points forts. Montrer que cette approche permet d'augmenter les performances du réseau et faire le point sur les techniques spécifiques nécessaires dont elle a besoin. Dans un deuxième temps vous devez faire des recherches sur les algorithmes d'allocation de fréquence dans les réseaux multi-fréquence et les comparer.

Travail pratique
La partie pratique de ce sujet est d'implémenter en langage C trois protocoles d'assignation de fréquences dans un contexte de réseaux ad hoc où les nœuds sont équipés de plusieurs cartes. Ces trois algorithmes sont MCAIR, Tabu et Greedy. Cette implémentation rentre dans le cadre d'un code C qui existe déjà et qui doit être complété. Le reste du code permet un interfaçage avec le simulateur de réseau NS-3. A partir de l'assignation, il génère des scripts NS-3 (configuration des nœuds et routage statique). L'autre partie de ce projet consistera donc à exécuter les scripts et à mesurer la capacité offerte par ces différents algorithmes. Ce travail pourra donner lieu à une soumission dans une publication (conférence ou revue).

Pré-requis
- Connaissances en réseaux ;
- Programmation en C/C++.

Références

1. Husnain Mansoor Ali, Anthony Busson, Véronique Vèque. *Channel Assignment Algorithms: A Comparison of Graph Based Heuristics*. 4th ACM Workshop on Performance monitoring, Measurement and Evaluation of Heterogeneous Wireless and Wired Networks. October 26, 2009. Tenerife, Spain.
2. A. P. Subramanian, H. Gupta, and S. R. Das. *Minimum interference channel assignment in multi-radio wireless mesh networks*. In Sensor, Mesh and Ad Hoc Communications and Networks, 2007. SECON '07, pages 481–490, San Diego, CA June 2007.

2. APPLICATION DE LA MÉTHODE COSMIC POUR ESTIMATION DANS UN PROJET AGILE DE DÉVELOPPEMENT DE LOGICIEL

Encadrement: Ho Tuong Vinh (ho.tuong.vinh@auf.org), MSI-IFI

Contexte

L'estimation (effort, vélocity, productivité, etc) dans un projet (adopté l'approche d'Agile) de développement de logiciel est subjective et dépend fortement de l'expérience de l'équipe de développement. Il existe plusieurs méthodes (Use Case Points, IFPUG Function Points, WBS, etc) pour estimer la taille, l'effort et la durée d'un projet de logiciel mais la plupart de ces méthode ne conviennent pas pour les projets adoptés l'approche d'Agile. La méthode COSMIC Function Points est une méthode reconnue par l'ISO et elle permet l'estimation de la taille de logiciels dans différentes étapes d'un processus de développement. Récemment, quelques travaux ont été réalisés permettant affirmer que la méthode COSMIC peut être appliquée pour l'estimation dans les projets adoptés l'approche d'Agile.

Ce TPE a pour but d'étudier la méthode COSMIC et ensuite de l'expérimenter dans un projet de logiciel.

Travaux théoriques

- Étudier l'estimation dans un projet de logiciel et un projet adopté d'Agile
- Étudier la méthode COSMIC
- Étudier comment la méthode COSMIC peut appliquer dans un projet d'Agile
-

Travaux pratiques

- Appliquer la méthode COSMIC pour l'estimation de 2 projets d'Agile (avec des exigences de logiciel réelles)
- Faire une évaluation de l'efficacité de l'application de COSMIC dans un projet d'Agile

Références

1. Documents sur la méthode COSMIC : http://cosmicon.com/
2. Document sur l'approche d'Agile : http://www.agilealliance.org/

3. CALCUL SUR CARTE GRAPHIQUE POUR SIMULATIONS MULTI-AGENTS

Encadrant: Nicolas Marilleau (nmarilleau@gmail.com), IRD
Laurent Philippe (laurent.philippe@lifc.univ-fcomte.fr)

Collaborateur local: Ho Tuong Vinh (ho.tuong.vinh@auf.org), MSI-IFI

Contexte

L'étude des phénomènes réels tels que les dynamiques urbaines (déplacement de véhicules et/ou de personnes), les phénomènes écologiques (évolution de la biologie des sols) ou encore les réseaux sociaux (dynamique des graphes de relations) est indispensable pour bien connaître les enjeux qui dirigent les systèmes complexes. Les systèmes multi-agents, ou SMA, permettent de simuler efficacement ces systèmes complexes en représentant chaque entité du système par un agent ayant un comportement indépendant et en observant les phénomènes globaux qui émergent de l'exécution simultanée des agents. Or la précision de l'étude dépend bien souvent de la taille du système considéré, ce qui engendre des coûts de calcul important.

Au cours des dernières années les cartes graphiques ont évolué d'une architecture uniquement dédiée à l'affichage vers des fonctionnalités de plus en plus généralistes, sans pour autant perdre en capacité de traitement. De ce fait elles sont maintenant fréquemment détournées de leur utilisation initiale pour réaliser des calculs intensifs. Symétriquement le processeur unique laisse place à des architectures multi-processeurs (plusieurs processeurs sur une même carte mère), multi-cœurs (plusieurs cœurs dans un même processeur), voire many-coeurs (grand nombre de cœurs dans un même processeur). L'exploitation efficace de ces architectures reste pourtant un défis et rares sont les plates-formes de simulation orientées agents qui tirent partie de ces nouvelles architectures.

La bibliothèque MCMAS (Many Cores for Multi-Agent Systems) a été conçue pour déporter certains calculs d'un système multi-agents du processeur principal vers le processeur de la carte graphique, ce dernier étant souvent sous utilisé pendant la simulation d'un système.

Sujet

Une première maquette de la bibliothèque a été validée pour différentes simulations. Elle propose une interface de haut niveau en JAVA qui permet d'exécuter, sur la carte graphique, des algorithmes développés dans un langage dédié aux cartes graphiques: OpenCL. Aujourd'hui, MCMAS dénombre quelques algorithmes (diffusion, démographie, compétition), d'autres sont disponibles et peuvent être rapidement intégrés (Dikjstra par exemple).

L'objectif de ce projet est d'intégrer la bibliothèque MCMAS dans la plate-forme multi-agents GAMA: http://gama-platform.googlecode.com, plate-forme de simulation agent développée en JAVA conformément à l'architecture Eclipse/OSGI. Ainsi, ce travail se déroulera en plusieurs étapes:

1. refactoring de MCMAS: concevoir et développer une interface générique et adaptée pour le monde de la simulation orientée agent;
2. création d'un module GAMA: concevoir et développer un plugin eclipse permettant d'interconnecter la plate-forme de simulation
3. GAMA avec la bibliothèque MCMAS;
4. tester les performances: comparer, au travers d'un modèle GAMA, les temps d'exécution des différentes versions de l'algorithme de diffusion, algorithme CPU proposé nativement par GAMA *vs* algorithme GPU proposé par MCMAS.

4. DÉVELOPPEMENT D'OUTILS DE VISUALISATION 3D AU SEIN D'UNE PLATEFORME DE MODÉLISATION ET SIMULATION À BASE D'AGENTS (GAMA)

Encadrement : Alexis Drogoul (alexis.drogoul@gmail.com), Directeur de Recherches IRD,
Arnaud Grignard, Doctorant UPMC/IRD

Contexte

GAMA[1] est une plateforme open-source de modélisation et de simulation à base d'agents développée depuis 2007 au sein de l'équipe MSI-UMMISCO [2]. Conçue comme une plate-forme multi-agent généraliste, elle sert de support à de nombreuses applications thématiques développées en partenariat avec différents instituts de recherche notamment dans le cadre de l'aide à la décision en matière environnementale au Vietnam (rôle de l'environnement dans la propagation du virus H5N1 au Nord-Vietnam, modèles d'aide à l'organisation des secours en cas de catastrophe naturelle en zone urbaine, aide à la lutte contre les invasions de sauterelles dans le Delta du Mékong). La complexification croissante des besoins exprimés par les thématiciens a fait apparaître la nécessité de disposer d'outils puissants permettant de représenter et de visualiser les phénomènes modélisés selon différents points de vue, à différentes échelles spatio-temporelles et sur différents types d'architectures (ordinateurs, système embarqués, navigateur internet...). Dans ce contexte l'objectif de ce sujet TPE 2 sera (1) de participer au développement d'un ensemble d'outils déjà existants basés sur la librairie graphique OpenGL [3]; (2) d'intégrer ces outils à l'environnement GAMA et à son langage de modélisation GAML, (3) d'explorer de nouveaux modes de visualisation interactives utilisant les dernières technologies disponibles.

Dans de nombreux cas de figure, la visualisation d'un modèle complexe joue un rôle très important, voire indispensable, dans la compréhension des phénomènes représentés. Il est donc étonnant de constater que peu d'outils existent pour permettre de visualiser les simulations de façon standardisée, interactive et immersive quelque soit le support utilisé (ordinateur, systèmes embarqués, etc.). En particulier, il n'existe aucune norme permettant de sauvegarder une simulation complète et de la « rejouer » en changeant de point de vue de visualisation, la seule solution étant, dans bien des cas, de relancer l'exécution du simulateur, ce qui n'est pas toujours possible (soit pour des raisons de temps de calcul, soit que le matériel de visualisation ne s'y prête pas). Nous proposons de développer lors de ce stage de Master 2 un premier ensemble d'outils intégrés à GAMA permettant de stocker et de manipuler les informations visuelles produites par des simulations afin de pouvoir les rejouer, indépendamment de la plateforme, tout en autorisant l'utilisateur à changer de point de vue ou de mode d'affichage. Ce projet se basera principalement sur les formats actuels de stockages d'informations 3D et sur les dernières technologies disponibles en terme de visualisation et d'interaction offerte par la nouvelle norme HTML 5 (webGL [4]) ainsi que par les systèmes embarqués de type iOS ou Android pour lesquels OpenGL ES [5] est utilisé.

Ce travail donnera lieu au développement de trois points distincts liés au système de fenêtrage, aux formats de données et aux différentes architectures de visualisation. La gestion de l'affichage (fenêtrage) actuellement utilisé dans GAMA repose sur la librairie standard AWT, qui se trouve être mal intégrée au système proposé par Eclipse, nommé SWT. Une première étape consistera donc à implémenter l'affichage OpenGL comme une surcouche directe de SWT, éliminant ainsi une source non négligeable d'inconfort pour l'utilisateur. La deuxième partie du travail portera sur le développement d'une librairie intégrée à GAMA permettant l'importation et l'exportation de formats de fichier 3D. Plusieurs type de fichiers seront à évaluer (.3ds, .blend , .x3d , .vrml, .dae) mais une attention particulière sera portée au format Collada[6](.dae) déjà utilisé dans de nombreuses applications (Blender, Maya, SketchUp, GoogleEarth) et proposant à la fois une structure pour

stocker les informations liées aux géométries des agents mais aussi la prise en compte des interactions et autres dynamiques intrinsèquement liées aux thématiques de modélisation multi-agent. En s'inspirant du travail très complet déjà effectué sur la gestion des fichiers .shp (2D) au sein de Gama [7], les fichiers 3D serviront à la fois d'entrée aux différents modèles pour la création d'agents (agentification) mais aussi de sortie pour permettre le stockage et l'échange des données produites par une simulation. Dans un troisième temps, et selon l'avancement du stage, il pourra être envisagé d'intégrer la visualisation de ces fichiers sur d'autres architectures comme les systèmes embarqués ou navigateurs web. Ces modes de visualisation feront intervenir des extensions de la libraire OpenGL comme OpenGL ES (iOS, Android) et webGL (html5/ javascript).

Travail théorique

L'étudiant devra tout d'abord faire une revue des outils de visualisation utilisés dans différentes plateformes de simulation à base d'agents (Mason, Repast, NetLogo, Breve, etc.) [8] [9] [10]. Il devra ensuite étudier les différents types de formats de fichier servant à stocker de l'information produite par les applications 3D ainsi que les différentes architectures existantes permettant de faire du rendu 3D (ordinateur, systèmes embarqués, navigateur web).

Travail pratique

Depuis la version 1.5, GAMA intègre un plug-in dédié à l'affichage OpenGL. Ce plug-in servira de base de développement et devrait permettre au candidat de se familiariser très vite avec le développement à réaliser. Ce travail sera réalisé en liaison avec les développeurs de GAMA, principalement Alexis Drogoul (Directeur de Recherches à l'IRD) et Arnaud Grignard (Doctorant UPMC/IRD) au sein du laboratoire MSI à l'Institut Francophone de l'Informatique à Hanoi. L'intérêt de ce travail pratique, pour l'étudiant, sera d'acquérir une solide expérience dans le développement d'applications de visualisation 3D ainsi que dans les thématiques de représentation, visualisation et interaction avec des modèles complexes.

Prérequis

Programmation Java (Eclipse), OpenGL (OpenGL ES et webGL), Javascript/html5, intérêt pour la visualisation et les interfaces utilisateur.

Références

1. GAMA : http://code.google.com/p/gama-platform/
2. UMMISCO (http://www.ummisco.ird.fr)
3. OpenGL : http://www.opengl.org/
4. webGL : http://www.khronos.org/webgl/
5. OpenGL ES : http://www.khronos.org/opengles/
6. Collada : http://www.khronos.org/collada/
7. GAMA: a simulation platform that integrates geographical information data, agent-based modeling and multi-scale control. P Taillandier, DA Vo, E Amouroux, A Drogoul - Principles and Practice of Multi-Agent Systems, 2012
8. Survey of Agent Based Modelling and Simulation Tools. Rob Allan . 2009
9. How to Do Agent-Based Simulations in the Future: From Modeling Social Mechanisms to Emergent Phenomena and Interactive Systems Design. D.Helbing, S.Balietti . 2011
10. Visualization tools for agent-based modeling in NetLogo. D.Kornhauser. 2007

5. DESCRIPTION D'UNE DISTRIBUTION GÉNÉRALE EN DISTRIBUTION DE TYPE PHASE

Encadrement : Thomas Begin (Thomas.Begin@ens-lyon.fr), UCBL1,
Alexandre Brandwajn, UCSC

Collaborateur local : Victor Moraru (victor.moraru@auf.org), MSI-IFI

Contexte

Lorsqu'on souhaite représenter la loi du temps entre des arrivées ou celle d'un temps de service dans un modèle théorique, on choisit souvent de le faire à l'aide d'une distribution de type Phase ou Cox. Ainsi, les nombreuses méthodes numériques pour résoudre des modèles de type file d'attente multi-serveurs supposent une distribution de type Phase ou Cox. Plusieurs arguments peuvent justifier ce choix :

1) Ces distributions sont construites comme des combinaisons d'étages « exponentiels » dont la propriété sans mémoire peux faciliter la résolution du modèle,
2) Toute distribution peut être aussi finement reproduite que voulue par une distribution en Phase ou en Cox.

Des chercheurs ont proposé des algorithmes afin de reproduire systématiquement dans une distribution en Phase les 2 premiers moments (la moyenne et la variance) ou bien même les 3 premiers moments d'une distribution donnée [2, 3]. D'autres chercheurs ont proposé des méthodes heuristiques afin de reproduire l'allure générale d'une distribution donnée par une distribution de type Phase [1]. Toutefois, aucune de ces solutions ne semble convenir dans un certain nombre de domaines, notamment lorsqu'elles sont appliquées à la résolution de modèles de type files d'attente multiserveurs [4]. En particulier, les distributions bi-modales (ou multi-modales) ainsi que celles à queue lourde font certainement partie des distributions pour lesquelles les résultats des méthodes existants seront les plus mitigés.

Travail théorique

- Lecture d'articles scientifiques autour de cette problématique
- Distinction des méthodes selon leurs objectifs, leur complexité de mise en œuvre, les caractéristiques de la distribution en Phase recherchée, …

Travail pratique

- Implémentation de certains algorithmes existants
- Tests sur certains exemples de distributions
- Détermination des facteurs favorisant les mauvais résultats de ces algorithmes et par conséquent des ensembles de distributions
- Applications de ces résultats à certains modèles classiques (par exemple les files multiserveurs)

Pré-requis

- Quelques notions de programmation C ou matlab ou langage de Script
- Connaissance en probabilités ou statistiques (distribution de probabilité, moments) sont un plus

Références

1. Asmussen, S., Nerman, O., and Olsson, M. 1996. *Fitting Phase-Type Distributions via the EM Algorithm*. Scandinavian Journal of Statistics. 23, 4, 419-441.
2. Bobbio, A., Horváth, A., and Telek, M. 2005. *Matching three moments with minimal acyclic phase type distributions*. Stoch. Models, 21, 2-3, 303-326.
3. Osogami, T. and Harchol-Balter, M. 2006. *Closed form solutions for mapping general distributions to quasi-minimal PH distributions*. Performance Evaluation. 63, 6, 524-552.
4. Begin, T. and Brandwajn, M. 2009. *A Note on Aspects of Workload Characterization in Parallel Access Volumes*. CMG 2009 - Dallas (Texas, US).

6. DÉVELOPPEMENT D'UNE BASE DE CONNAISSANCE SUR LES GÈNES RÉGULATEURS DE LA RAMIFICATION CHEZ LE RIZ (*O. SATIVA*)

Encadrement : Pierre LARMANDE (pierre.larmande@ird.fr) , IRD
Collaborateur local : Ho Tuong Vinh (ho.tuong.vinh@auf.org), MSI-IFI

Mots Clefs: Linked Open Data, Semantic Web, Knowledge representation, Data mining, Bioinformatics

Contexte

Le terme de web sémantique a été proposé par Tim Berners Lee en 2001 [1] pour désigner une évolution du web qui permettrait aux données disponibles (contenus, liens) d'être plus facilement utilisables et interprétables automatiquement, par des agents logiciels [2]. Pour permettre cette évolution, un certain nombre de standards tels que le RDF (Resource Description Framework) [3] et des méthodologies [4] ont été développés par le W3C, avec pour objectif de sortir les données des silos fermés que constituent les bases de données en ligne.

Le web de données ouvertes (Linked Open Data - LOD) est une initiative visant à favoriser la publication de données structurées sur le web, non pas sous la forme de silos de données isolés les uns des autres, mais en les reliant entre elles pour constituer un réseau global d'informations.

Aujourd'hui, le web de données a ainsi pu atteindre une masse critique de données disponibles et librement réutilisables, ouvrant la porte à de nombreuses expérimentations.

Le fait que les données soient structurées en RDF présente l'avantage d'utiliser des mécanismes de requêtes (avec le langage d'interrogation SPARQL) pour exploiter au mieux les liens explicites existants entre les données. De plus ces liens peuvent être enrichis avec de nouvelles données [5]. Enfin, la compatibilité entre RDF et le langage de développement des ontologies pour le web (OWL) permet de structurer les données de manière à inférer des connaissances implicites [6,7].

Avec l'explosion des technologies 'omiques' de nombreuses données sont disponibles en RDF à travers des portails dédiés. Le portail Bio2RDF [8] recense un grande majorité des ressources bioinformatique (NCBI, GOA, InterproDB, Kegg et Pubmed). D'autres ressources sont disponibles à travers http://biolod.org ou encore Uniprot [9].

Travaux théoriques et pratiques

L'objectif du projet est (1) d'effectuer un inventaire des ressources LOD nécessaires pour répondre aux questions biologiques, (2) structurer les données dans une base de connaissance dédié en y associant des ontologies du domaine, (3) établir des requêtes sur les données structurées correspondant aux questions biologiques initiales.

Références

1. « The Semantic Web », Scientific American Magazine, May 17, (2001)
2. http://www.bnf.fr/fr/professionnels/web_semantique_donnees/s.web_semantique_intro.html
3. W3C Recommendation: RDF Primer. http://www.w3.org/TR/rdf-primer/
4. Christian Bizer, Richard Cyganiak, Tom Heath: How to Publish Linked Data on the Web. Online tutorial. http://sites.wiwiss.fu-berlin.de/suhl/bizer/pub/LinkedDataTutorial/
5. M. Scott Marshall, Richard Boyce, Helena F. Deus, Jun Zhao, Egon L. Willighagen, Matthias Samwald, Elgar Pichler, Janos Hajagos, Eric Prud'hommeaux, Susie Stephens: Emerging practices for mapping and linking life sciences data using RDF - A case series. J. Web Sem. 14: 2-13 (2012).
6. Achille Zappa, Andrea Splendiani, Paolo Romano: Towards linked open gene mutations data. BMC Bioinformatics 13(S-4): S7 (2012).

7. Matthias Samwald, Adrien Coulet, Iker Huerga, Robert L Powers, Joanne S Luciano, Robert R Freimuth, Frederick Whipple, Elgar Pichler, Eric Prud'hommeaux, Michel Dumontier, M Scott Marshall: Semantically enabling pharmacogenomic data for the realization of personalized medicine. Pharmacogenomics, 13(2): 201-212 (2012).
8. Alison Callahan, Jose Cruz-Toledo, Peter Ansell, Michel Dumontier. Bio2RDF Release 2: Improved coverage, interoperability and provenance of Life Science Linked Data. Extended Semantic Web Conference, Semantics and Big Data, 2013. To appear.
9. UniProt SPARQL endpoint: http://beta.sparql.uniprot.org/

7. DÉVELOPPEMENT D'UN SYSTÈME DE GESTION DE DONNÉES DE PHÉNOTYPAGE CHEZ LE RIZ (*O. SATIVA*)

Encadrement: Pierre LARMANDE (pierre.larmande@ird.fr),
Stéphane JOUANNIC (stephane.jouannic@ird.fr), Pascal GANTET (pascal.gantet@ird.fr), IRD

Collaborateur local : Ho Tuong Vinh (ho.tuong.vinh@auf.org), MSI-IFI

Contexte

Dans le cadre du laboratoire LMI RICE, des études de la diversité génotypique et phénotypique de variétés traditionnelles de riz vietnamien sont conduites dans le but d'identifier des gènes d'intérêt pour la compréhension de processus biologiques (développement et plasticité de la plante, résistance aux maladies) mais également pour des futur programmes d'amélioration. Ces études requièrent **la manipulation d'un important volume de données hétérogènes. Ces données peuvent être stockées sous la forme de fichier Excel, texte structuré, images ou bases de données relationnelles. Dans ce contexte, l'équipe du LMI RICE souhaite organiser ses propres jeux de données afin de pouvoir naviguer, partager et annoter ces dernières afin de les exploiter au mieux.**

La difficulté réside dans la définition de systèmes « souples », c'est à dire supportant une évolution des besoins utilisateurs avec un minimum de développement. L'importance des données médias (images dans ce cas) est à prendre en compte. En effet, leur association avec les jeux de données « textuelles » est évidente, mais elle nécessitent également la prise en compte de « méta-informations » d'abords basique comme l'auteur, la date, le lieu, géolocalisation, puis élaborée comme un système de « tagging » permettant de rechercher des associations entre les jeux de données.

Travaux théoriques

- Faire un inventaire des besoins en termes de gestion de données et de stockage associé
- Faire une recherche des solutions existantes dans le domaine biologique et/ou logiciel
- Définir les critères pour choisir une solution adaptée et pérenne
- Proposer une solution pour le système choisi

Travaux pratiques

- Appliquer la solution envisagée sur un jeu de données identifiées
- Faire une première version de système de gestion
- Définir et développer les requêtes correspondants à des questions biologiques
- Valider avec les biologistes la pertinence des résultats obtenus

8. ESTIMATION DE DENSITÉ AVEC DES TRANSFORMÉES EN ONDELETTES

Encadrement: Benoît Frénay (benoit.frenay@uclouvain.be), Université catholique de Louvain
Collaboration locale : Ho Tuong Vinh (ho.tuong.vinh@auf.org), MSI-IFI

Contexte

En apprentissage automatique, il est souvent important de pouvoir estimer la densité des données. Par exemple, en classification, apprendre la distribution des données à partir d'observations ayant été classées permet de construire un classifieur pour classer de nouvelles données. En reconnaissance de la parole, modéliser la densité du signal permet de reconnaître des sons. Le but de ce TPE sera d'explorer l'utilité des transformées en ondelettes pour l'estimation de densité. Généralisant la transformée de Fourier, la transformée en ondelettes (wavelet transform) permet de déterminer si une fonction varie rapidement ou non. Elle est par exemple utilisée pour analyser un signal à différentes échelles temporelles. L'étudiant fera d'abord un survol théorique des techniques existantes. Ensuite, il utilisera ces techniques sur des données réelles unidimensionnelles, bidimensionnelles ou en plus haute dimension si les résultats et les techniques le permettent. Ces résultats seront comparés à ceux d'autres techniques d'estimation de densité.

Travaux théoriques

- Faire un survol de quelques techniques simples pour estimer une densité de probabilité ;
- Comprendre les principes de bases de la transformée en ondelettes ;
- Faire un survol des techniques utilisant la transformée en ondelettes pour estimer une densité.

Travaux pratiques

- Implémenter des estimateurs de densité simples et avec transformée en ondelettes ;
- Tester les estimateurs sur des données unidimensionnelles, bidimensionnelles ou en plus haute dimension si les résultats et les techniques implémentées le permettent ;
- Proposer des pistes d'amélioration pour les techniques utilisées.

Références

1. The Wavelet Tutorial, Robi Polikar, http://users.rowan.edu/~polikar/WAVELETS/WTtutorial.html
2. A Wavelet Tour of Signal Processing, Stéphane Mallat, Academic Press, 1998

9. ESTIMATION DE LA CAPACITÉ DES RÉSEAUX AD HOC SANS FIL

Encadrement : Anthony Busson (anthony.busson@inria.fr), UCBL1
Victor Moraru (victor.moraru@auf.org), MSI-IFI

Contexte

L'idée de ce TPE est d'étudier la capacité des réseaux ad hoc sans fil. La capacité est ici le débit maximum que l'on peut obtenir dans le réseau. On s'intéressera particulièrement aux réseaux maillés sans fil. Il s'agit de réseaux sans fil ad hoc qui servent de réseau d'infrastructures lors d'événements ou de catastrophes. Leur intérêt est qu'ils sont peu onéreux et facile à déployer. Il sont donc adaptés à des besoins temporaires ou à la couverture de zones blanches (zones qu'il est trop coûteux de connecter au travers d'une solution filaire). Ces réseaux doivent offrir un maximum de débit aux utilisateurs, il est donc important de pouvoir l'évaluer pour un déploiement donnée. Ce TPE est divisé en deux phases. Le premier aspect est théorique: l'étudiant devra effectuer une bibliographie sur les méthodes permettant d'évaluer les capacités des réseaux sans fil. La deuxième partie consistera en l'étude de deux méthodes, qui seront fournit par les encadrants, permettant d'évaluer la capacité du réseau. Ces deux méthodes sont basés sur la notion de graphe de conflits, et de liens interférants. Pour un lien donné on essaye d'évaluer le partage de la bande passante avec les liens radios interférants. Afin d'évaluer la pertinence de ces deux méthodes, des simulations sur le simulateur NS3 devront être effectuer pour voir laquelle des deux méthodes donnent les résultats les plus fins.

Travail théorique:

- Bibliographie sur les méthodes d'évaluation des capacités des réseaux sans fil multisauts ;
- Compréhension des notions relatives à ces calculs: graphe des conflits, notion d'interférences, CCA (clear channel assesment), interférence à n sauts, taux d'utilisation.

Travail pratique:

- Implémentation de scripts NS-3 pour l'évaluation de la capacité réelle d'un réseau ;
- Parsing des résultats (scripting: shell, awk, etc.) ;
- Calcul théorique des capacités (matlab ou tout autre logiciel).

Pré-requis

- Programmation C/C++ ;
- Connaissances des réseaux sans fil ;
- Une connaissance de NS3 est un plus.

Références:

1. Akyildiz, I.F. Xudong Wang, A survey on wireless mesh networks. IEEE Communications Magazine,Volume:43 , Issue: 9.
2. Husnain Mansoor Ali, Anthony Busson, Véronique Vèque. Channel Assignment Algorithms: A Comparison of Graph Based Heuristics. 4th ACM Workshop on Performance monitoring, Measurement and Evaluation of Heterogeneous Wireless and Wired Networks. October 26, 2009. Tenerife, Spain.
3. Kamal Jain, Jitendra Padhye, Venkata N. Padmanabhan and Lili Qiu. Impact of Interference on Multi-Hop Wireless Network Performance. Wireless Networks 11, 471–487, 2005

10. ÉTUDE, IMPLÉMENTATION ET VISUALISATION DE MODÈLES DE DIFFUSION D'OPINIONS

Encadrant local : Ho Tuong Vinh (ho.tuong.vinh@auf.org), MSI-IFI
Encadrant extérieur : Benoit Gaudou (benoit.gaudou@gmail.com), Université de Toulouse

Contexte
Les modèles de diffusion d'opinion s'intéressent aux mécanismes influençant et permettant la diffusion d'une opinion dans une population d'individus. Les premiers modèles agents ont été proposé il y a plus de 15 ans. Ce sont en général des modèles abstraits et simples permettant de se focaliser sur un mécanisme unique et d'explorer son influence en profondeur. Ce sont également en général des modèles aspatiaux. Ils peuvent néanmoins parfois être étendu à sur des réseaux sociaux.

Le but de ce TPE sera d'identifier et de recencer les différents modèles de diffusion d'opinion existant dans la littérature et de les comparer. Dans un second temps, l'étudiant choisira un certain nombre de ces modèles et les implémentera dans la plate-forme de modélisation et simulation agent GAMA. Il devra ensuite les explorer et comparer les résultats de chacun. L'étudiant étendra ensuite ces modèles pour les appliquer sur des réseaux sociaux (théoriques ou réels). Un travail sur la visualisation des résultats de ces modèles théoriques sera également envisagé.

Travail théorique
- Découverte de la simulation multi-agents ;
- Étude de la plate-forme de simulation GAMA ;
- Découverte du domaine des réseaux sociaux
- Étude des différents modèles de dynamique d'opinion.

Travail pratique
- Implémentation et études des différents modèles choisis
- Extension des ces modèles sur des réseaux sociaux (small-world, scale-free…)
- Étudier de possibles façons de visualiser ces modèles

Références
1. Axelrod, R.: Dissemination of culture: A model of local convergence and global polarization. Journal of Conflict Resolution 41 (1997) 203–226.
2. Moscovici, S., Doise, W.: Dissension et consensus. PUF, Paris (1992)
3. Weisbuch, G., Boudjema, G.: Dynamical aspects in the adoption of agri-environmental measures. Advances in Complex Systems 2 (1999) 11–36.
4. Deffuant, G., Neau, D., Amblard, F.,Weisbuch, G.: Mixing beliefs among interacting agents. Advances in Complex Systems 3 (2000) 87–98.
5. Deffuant, G., Amblard, F., Weisbuch, G., Faure, T.: How can extremism prevail ? a study based on the relative agreement interaction model. JASSS 5 (2002)
6. Janssen, M.A.: Games & Gossip. An eBook. http://www.openabm.org/book (2010)
7. Galam, S. : Sociophysics : A review of Galam models. International Journal of Modern Physics C, Volume 19, Issue 03, pp. 409-440 (2008)
8. Taillandier, P., Vo, D. A., Amouroux, E. and Drogoul A.: GAMA: a simulation platform that integrates geographical information data, agent-based modeling and multi-scale control, Principles and Practice of Multi-Agent Systems, pp. 242–258, 2012.

11. ÉVALUATION DE LA BANDE PASSANTE RESTANTE DANS LES RÉSEAUX MAILLÉS SANS FIL BASÉS SUR LE STANDARD 802.11

Encadrement : Isabelle Guérin Lassous (Isabelle.Guerin-Lassous@ens-lyon.fr), UCBL1
Victor Moraru (victor.moraru@auf.org), MSI-IFI

Contexte

Les réseaux maillés sans fil (Wireless Mesh Networks) apparaissent très souvent comme des réseaux d'accès sans fil multisauts. De tels réseaux ont fait leur apparition en France dans les zones appelées zones blanches où la connexion filaire est inexistante. Pour que ces réseaux soient bien acceptés, il faut qu'ils soient en mesure de véhiculer correctement les applications classiques utilisées actuellement par les utilisateurs telles que le web, le streaming ou encore la ToIP. Certaines de ces applications ont des contraintes fortes, notamment en termes de délai, de gigue ou encore de bande passante. Or de part le caractère multisaut et le partage du médium radio, les réseaux maillés sans fil ont des performances limitées comparé au monde filaire.

La mise en place de qualité de service semble donc être une étape fondamentale dans l'amélioration des performances des réseaux maillés sans fil. La plupart des mécanismes de qualité de service reposent sur une estimation des ressources disponibles, comme par exemple la bande passante. Plusieurs solutions d'évaluation de bande passante disponible ont été proposées jusqu'ici. Le but du travail est d'améliorer une solution qui a été proposée récemment, appelée SABE, et d'effectuer une évaluation de performante complète de la solution.

Travail théorique

- Bien maîtriser la technique d'accès DCF de 802.11 ;
- Comprendre la problématique d'évaluation de bande passante disponible dans les réseaux sans fil multisauts basés sur 802.11 ;
- Étudier la solution à améliorer (solution SABE) ;
- Étudier le simulateur NS2.

Travail pratique

- Première évaluation de la solution SABE sous NS2 ;
- Proposer des améliorations de SABE ;
    - en termes de données à envoyer ;
    - dans l'introduction de la probabilité de collision ;
- Implémentation, validation et évaluation de la solution proposée ;
- S'il reste du temps, comparaison avec d'autres solutions.

Pré-requis

- Programmation C/C++ ;
- Connaissances réseau ;
- Une connaissance de NS2 est un plus.

Références

1. C. Sarr, C. Chaudet, G. Chelius, and I. Guérin-Lassous. *Bandwidth Estimation for IEEE802.11-Based AdHoc Networks*. IEEE Transactions on Mobile Computing, 7(10):1228–1241, 2008.
2. H. Zhao, E. Garcia-Palacios, J. Wei, and Y. Xi. *Accurate available bandwidth estimation in IEEE 802.11-based ad hoc networks*. Comput. Commun., 32:1050–1057, April 2009.
3. Rapport de stage de master sur SABE.

12. HADOOP SUR OPENSTACK

Encadrement : Nguyen Hong Quang (nguyen.hong.quang@auf.org), MSI-IFI

Contexte
Hadoop est un framework Java libre destiné à faciliter la création d'applications distribuées et échelonnables (scalables). Il permet aux applications de travailler avec des milliers de nœuds et des pétaoctets de données. Hadoop a été inspiré par les publications MapReduce, GoogleFS et BigTable de Google.

OpenStack est un projet informatique de service d'infrastructure (Infrastructure as a Service (IaaS)) du domaine du cloud computing, mené par la Fondation Openstack.

La fondation Openstack est une organisation non-commerciale qui a pour but de promouvoir le projet Openstack ainsi que de protéger et d'aider les développeurs et toute la communauté Openstack.

Beaucoup d'entreprises ont rejoint la fondation Openstack. Parmi celles-ci on retrouve : Canonical, Red Hat, SUSE Linux, AT&T, Cisco, Dell, HP, IBM et Yahoo!

Ce TPE adresse la collaboration entre deux logiciels libres Hadoop et OpenStack afin de profiter les avantages de tous les deux.

Travail théorique
Étudier le concept et l'architecture de Hadoop [1, 2], le nuage IaaS Openstack [3] et le projet « Hadoop on OpenStack » [4].

Travail pratique
Effectuer une installation et configuration de Hadoop sur le nuage OpenStack de l'IFI.

Pré-requis
Connaissances de système Linux

Références

1. Hadoop (Wikipedia) http://fr.wikipedia.org/wiki/Hadoop
2. Hadoop – Une introduction, http://blog.inovia-conseil.fr/?p=46
3. OpenStack, http://www.openstack.org/software/
4. Projet Hadoop on OpenStack, http://fr.slideshare.net/lukjanovsv/hong-kong-openstack-summit-savanna-hadoop-on-openstack

13. IMPLÉMENTATION D'UN THÉORÈME PROVER POUR DÉDUIRE DES PROPRIÉTÉS ÉMERGENTES DE SIMULATIONS SOCIALES MULTI-AGENTS

Encadrant local : Ho Tuong Vinh (ho.tuong.vinh@auf.org), MSI-IFI
Encadrants extérieurs : Benoit Gaudou (benoit.gaudou@gmail.com), Université de Toulouse

Contexte
Dans un travail précédent (Gaudou et al., 2012) nous avons proposé un formalisme pour étudier comment faire des simulations sociales en logique. Afin d'atteindre cet objectif, nous avons introduit une logique dynamique avec affectations et itérations bornées et non-bornées. Nous avons montré que notre logique permet de représenter et de raisonner sur un exemple paradigmatique de la simulation sociale : le modèle de ségrégation de Schelling. Nous avons établi également un lien entre simulation et planification. En particulier, nous avons montré que le problème de la vérification qu'une certaine propriété P (comme la ségrégation) va émerger après n pas de simulation peut se ramener à un problème de planification à horizon n, c'est-à-dire le problème de la vérification de l'existence d'un plan de longueur au plus n assurant qu'un certain but va être atteint, problème qui a été très étudié en IA.

Le but du TPE est d'implémenter un therom prover pour cette logique afin d'étudier la faisabilité pratique de l'approche proposée. Ce travail pourra se baser sur l'étude préliminaire faite par un étudiant de Master 2

Travail théorique
- Découverte de la simulation multi-agents et du modèle de Schelling ;
- Étude des travaux existants (Gaudou et al., 2012) et (Cultien, 2011);
- Étude du formalisme logique utilisé;
- Étude des différents theorem provers existant.

Travail pratique
- Implémentation d'un theorem prover ;
- Tests de faisabilité et limites.

Références
1. Taillandier, P. ; Drogoul A. ; Vo D.A. & Amouroux, E. (2010), GAMA : a simulation platform that integrates geographical information data, agent-based modeling and multi-scale control. In PRIMA'10, Kolkata, India, pp. 67—74.
2. Gaudou, B., Herzig, A., Lorini, E., Sibertin-Blanc, C.: How to do social simulation in logic: modeling the segregation game in a dynamic logic of assignments, In Multi-Agent-Based Simulation XII, 2012.
3. T. C. Schelling. Dynamic Models of Segregation. Journal of Mathematical Sociology, 1 : 143–186, 1971.
4. C. Cultien : Création d'un solveur pour une logique dynamique avec affectation propositionnelle à l'aide d'une traduction en QBF. Rapport de Master 2, 2011.

14. LE TEST DIRIGÉ PAR LES MODÈLES

Encadrement: Ho Tuong Vinh (ho.tuong.vinh@auf.org), MSI-IFI

Collaboration externe : CAMP Olivier (olivier.camp@eseo.fr), ESEO, Angers, France

Contexte

Le test d'une application logicielle est une activité importante dans son cycle de vie. Cette phase de développement permet de s'assurer du bon fonctionnement de l'application. Traditionnellement les tests sont écrits manuellement et ils représentent environ 50 % du coût d'un projet informatique. Il est pour cette raison important d'optimiser cette phase du développement en proposant des méthodes facilitant la production des tests.

Depuis la fin du XX^{ème} siècle l'ingénierie dirigée par les modèles (IDM) promeut les modèles en tant qu'entité de première classe dans le développement logiciel. C'est une approche par laquelle tout ou partie d'une application informatique est générée à partir de modèle. Les modèles deviennent ainsi des objets indispensable au développement de système.

Le test dirigé par les modèles (Model Based Testing – MBT) s'inscrit dans cette approche et consiste à générer les tests permettant de vérifier le bon fonctionnement d'une application à partir des modèles de l'application. La modélisation graphique atteint cependant ses limites en termes d'expressivité, et il peut être nécessaire de rajouter manuellement des annotations et/ou de s'aider d'une notation ou d'un langage (comme OCL, par exemple) pour rajouter de la sémantique au modèle.

Travaux théoriques

- Faire un état de l'art des travaux dans le domaine du Test Dirigé par les Modèle et classifier les approches en fonction des approches utilisés.
- Proposer une méthodologie pour intégrer une modélisation des tests dans le modèle de l'application.
- Proposer un processus permettant de générer les tests à partir d'un modèle. On s'intéressera en particulier à des tests pour tester des programmes Java (Junit, TestNG).

Travaux pratiques

- Mettre en œuvre la méthodologie proposée pour permettre la génération automatique de test dans l'environnement Eclipse ou NetBeans.

Références

1. M. Utting, B. Legeard, **Practical Model-Based Testing : A Tools Approach,** Elsevier, 2010
2. Samba Diaw, Rédouane Lbath, and Bernard Coulette, État de l'art sur le développement logiciel **basé sur les transformations de modèles,** *Technique et Science Informatiques 29(4-5):505-536 2010*
3. Utting, M., Pretschner, A. and Legeard, B. , **A taxonomy of model-based testing approaches**. Software Testing Verification and Reliability, 22: 297–312, Wiley, 2012
4. SCHIEFERDECKER Ina, **Model Based Testing,** IEEE Software, vol. 29, n^{O}1, pp. 14-18, 2012
5. Mohamed Mussa, Ferhat Khendek, T**owards a Model Based Approach for Integration Testing,** SDL 2011: Integrating System and Software Modeling, pp 106-121, Springer, 2012
6. Tissot R., Julliand J., Masson P-A.. **Contribution à la génération automatique de tests à partir de modèles et de schémas de test comme critères de sélection dynamiques.** Thèse de doctorat, Laboratoire d'Informatique de Université Franche-Comté, 2009.
7. Helaine Sousa, Denivalda Lopes, Zair Abdelouahab, Daniela Barreiro Claro, Slimane Hammoudi, **An approach for model driven testing Framework, metamodels and Tools**, International Journal of Computer Systems Science and Engineering - IJCSSE (Vol 26 No 4 July 2011) - CRL Publishing. 2011.

15. L'EXPLORATION AUTOMATIQUE DES OPINIONS EN LIGNE

Encadrant : Dr Nguyen Manh Hung (nmhufng@yahoo.com)
chercheur associé

Contexte

On est dans la cadre de la commerce électronique dans la quelle un client souhaite acheter un produit mais il n'a pas assez de connaissance sur le produit: sa qualité, sa utilisabilité, etc. Donc il veut référer des opinions des gens qui ont déjà achetés et utilisés le produit.

Le problème est que ces opinions sur le produit sont situés partout en ligne, dans plusieurs pages webs différents. Donc, comment peut-on collecter des opinions sur un produit? Et comment peut-on reconnaître automatiquement que chaque opinion soit positive, négative ou bien neutre quand elle est écrite en langage naturelle?

L'objectif du sujet est de proposer un cadre pour un système de collecter et reconnaître automatiquement des opinions en ligne sur un ou des produits.

Travail théorique

- Proposer un mécanisme permettant de collecter les textes qui représentent des opinions en ligne
- Proposer un mécanisme permettant de reconnaître des opinions (positive, négative, neutre) qui sont représentés en textes dans un langage naturelle (Vietnamien)

Travail pratique

- Développer un système de collecter et de reconnaître automatiquement des opinions en ligne qui peut donner des recommandations (sur un produit) aux utilisateurs (clients).

Références :

1. Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social Web, *Journal of the American Society for Information Science and Technology*, 63(1), 163-173.
2. Garas, A., Garcia, D., Skowron, M., & Schweitzer, F. (2012). Emotional persistence in online chatting communities. *Scientific Reports*, 2, article 402.
3. Thelwall, M., Buckley, K., Paltoglou, G. Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12), 2544–2558.
4. Liu B., "Tutorial on sentiment analysis" based on Chapter 11 of the book "Web Data Mining – Exploring Hyperlinks, Contents and Usage Data". (http://www.cs.uic.edu/~liub/), 2007.
5. Sriram Raghavan, Hector Garcia-Molina*, Crawling the HiddenWeb,* Computer Science Department, Stanford University, Stanford, CA 94305, USA.
6. Stephen W. Liddle, Sai Ho Yau, and David W. Embley*, On the Automatic Extraction of Data from the Hidden Web.* In Proceedings of the International Workshop on Data Semantics in Web Information Systems (DASWIS-2001).

16. MÉTAGÉNOMIQUE SUR LA GRILLE DE CALCUL

Encadrement : Nguyen Hong Quang (nguyen.hong.quang@auf.org), MSI-IFI
Doan Trung Tung (dttung@gmail.org), IFI-MSI

Contexte
Une grille informatique ou « grid » est une infrastructure virtuelle constituée d'un ensemble de ressources informatiques potentiellement partagées, distribuées, hétérogènes, délocalisées et autonomes. Depuis 10 ans, la technologie des grilles informatiques a permis le développement de véritables infrastructures distribuées fournissant de vastes ressources de calculs et de stockage aux communautés scientifiques.

La métagénomique est confrontée par ses objets d'études et ses procédés aux défis de l'analyse des données de séquençage à haut débit liés aux volumes énormes de données et aux temps d´exécution des algorithmes. Pour la biologie et la génomique à grande échelle, on peut distinguer deux besoins principaux:

- la production des données brutes (séquences), qui réclame des moyens importants de stockage et d'archivage et des moyens relativement modérés de calculs.
- le traitement/service/support bio-informatique permettant de produire des données 'filtrées' à plus forte valeur ajoutée et exploitables par les communautés intéressées (biologie/médecine/agronomie).

L'utilisation de la grille a été étudiée pour ces deux besoins.

Ce TPE adresse en particulier le portage d'un pipeline d'analyse métagénomique sur la grille. Le pipeline PANAM est un pipeline dédié à l'analyse de séquences environnementales et il est déjà déployé sur la grille avec la plateforme DIRAC. Maintenant, il faut adapter / optimiser PANAM pour qu'il puisse analyser avec l'enjeu de données plus grandes.

Travail théorique
Etudier l'architecture de la grille générale et l'infrastructure de la grille EGI [1] dont l'IFI possède un nœud. Etudier l'intergiciel de la grille en générale (gLite par exemple) et la plateforme de la grille DIRAC [1] et ses outils. Étudier les méthodes de porter une application sur la grille.

Travail pratique
Collaboration avec Dr. Doan Trung Tung, ancien thésard du MSI qui travaille actuellement sur la grille EGE, l'étudiant devrai re-déployer PANAM sur le nœud de l'IFI (PANAM est actuellement déployé sur un nœud française) et puis l'adaptera et l'optimisera pour l'enjeu de données plus grandes.

Pré-requis
Connaissances de système Linux, le langage C/C++/Java.

Références
1. La grille EGI: http://www.egi.eu
2. gLite : http://www.cern.ch/glite
3. DIRAC : http://diracgrid.org/

17. MÉTHODES ET OUTILS D'ANALYSE DES CODES DE LOGICIEL

Encadrant: Ho Tuong Vinh (ho.tuong.vinh@auf.org), MSI-IFI

Contexte

La révision de code source des programmes logiciels est un moyen efficace pour assurer la qualité de logiciels. Pourtant, la révision manuelle coûte très chère en temps de temps et de ressource humaine. Une révision automatique à l'aide des outils d'analyse de code source est une solution prometteuse. Au fil des années, plusieurs travaux de recherche ainsi que les outils logiciels ont été réalisés.

Ce TPE a pour but de faire un bilan de l'état de l'art des travaux de recherche récents dans le domaine de l'analyse automatique de code de logiciel. En termes de travail pratique, certains logiciels existants seront choisis et expérimentés afin d'avoir une vue sur l'efficacité de ces outils.

Travaux théoriques

- Étudier les techniques pour la révision de codes
- Faire un bilan des travaux de recherche en analyse de code

Travaux pratiques

- Établir une liste des outils d'analyse de code populaire
- Faire une évaluation de l'efficacité de ces outils

Références

1. http://en.wikipedia.org/wiki/Static_program_analysis
2. SCRUB: a tool for code reviews, http://spinroot.com/gerard/pdf/ScrubPaper_rev.pdf

18. MODÈLE DE GÉNÉRALISATION DE DONNÉES GÉOGRAPHIQUES

Encadrant : Patrick Taillandier (patrick.taillandier@gmail.com), UMR IDEES – Univ. de Rouen
Collaborateur local : Ho Tuong Vinh (ho.tuong.vinh@auf.org), MSI-IFI

La généralisation de données géographiques est un problème crucial en cartographie. L'ACI (Association Cartographique Internationale) a défini en 1973 la généralisation cartographique comme « la sélection et la représentation simplifiée de détails en fonction de l'échelle et des objectifs de la carte ». La figure 1 donne un exemple de généralisation cartographique : comme le montre la figure, elle ne se limite pas à une simple réduction de la taille de la carte mais nécessite l'application de nombreuses opérations telles que des grossissements, des déplacements, ou des éliminations afin de garantir une bonne lisibilité de la carte tout en gardant l'information essentielle de la carte d'origine.

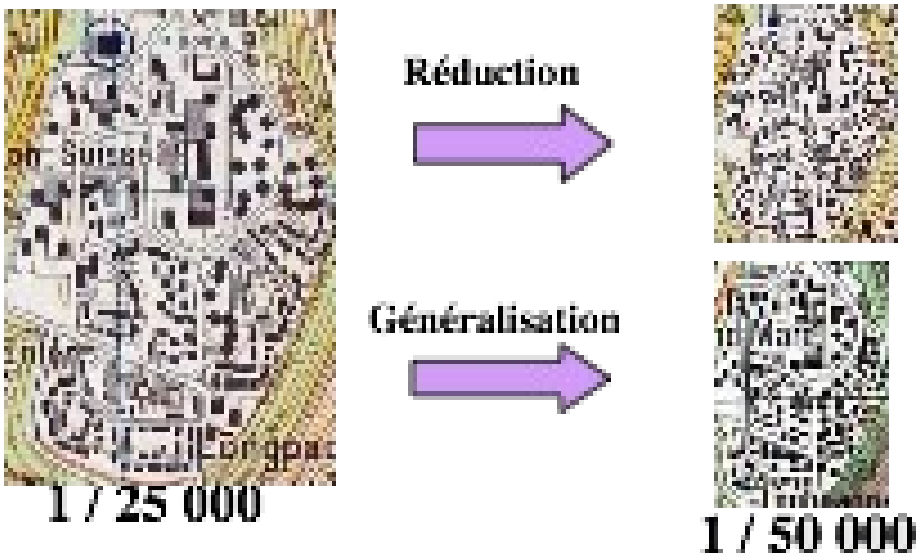

FIG. 1 – Généralisation cartographique

Différentes approches ont été étudiées pour résoudre le problème de l'automatisation de la généralisation dans son ensemble. Une première approche consiste à voir ce problème comme un problème d'optimisation global où les transformations sur les objets géographiques se déroulent de façon continue (Sester, 2000; Harrie & Sarjakoski, 2002). Ces méthodes, bien adaptées pour les faibles changements d'échelle, ne permettent pas l'application de transformations brusques (grossissement, simplification, etc.) nécessaires à des changements d'échelle plus importants. Pour faire face à cela une autre approche consiste à traiter ce problème d'automatisation de la généralisation par l'application d'opérations locales. Dans ce cadre, la mise en place d'un processus global de généralisation automatique revient à déterminer le choix et le séquencement d'algorithmes de généralisation (que nous désignerons dans la suite de cet article par le terme «action») à appliquer sur les divers objets géographiques. Plusieurs travaux ont visé à répondre à ce problème en adoptant une approche agents. Parmi eux, l'un des modèles les plus utilisés est le modèle AGENT (Ruas, 1999,), qui propose de modéliser les objets géographiques (routes, bâtiments, ...) sous la forme d'agents responsables de leur propre généralisation.

Ce modèle comprend deux types d'agents, les agents micro qui représentent les objets géographiques élémentaires (bâtiment, tronçon de route, ...) et les agents meso qui représentent des

groupements d'agents. Un agent meso peut aussi bien être composé d'agents micro (un lotissement composé de bâtiments) que d'autres agents meso (un quartier composé de lotissements).

Les interactions entre agents s'effectuent de manière hiérarchique. Ce sont les agents de niveau supérieur qui vont déclencher la généralisation de leurs sous-agents (un agent lotissement va déclencher la généralisation des bâtiments le composant).

La généralisation des agents est guidée par un ensemble de contraintes qui traduisent les spécifications du produit cartographique souhaité. Un exemple de contrainte est, pour un agent bâtiment, d'être suffisamment gros pour être lisible. Les agents disposent d'une liste d'actions de généralisation. Le choix de l'application de ces actions dépend des satisfactions des contraintes qui sont elles-mêmes calculées à partir des caractéristiques géométriques de l'agent (son état). Pour satisfaire au mieux ses contraintes, un agent géographique réalise un cycle d'actions durant lequel il peut tester les différentes actions afin d'atteindre un état parfait (où toutes ses contraintes sont parfaitement satisfaites) ou au moins un état satisfaisant. Le cycle d'actions se traduit par un parcours en profondeur de type préfixe d'un arbre d'états. La figure 3 donne un exemple d'arbre obtenu pour la généralisation au 1:25 000 d'un bâtiment. Le passage d'un état à un autre correspond à l'application par l'agent de l'une des actions.

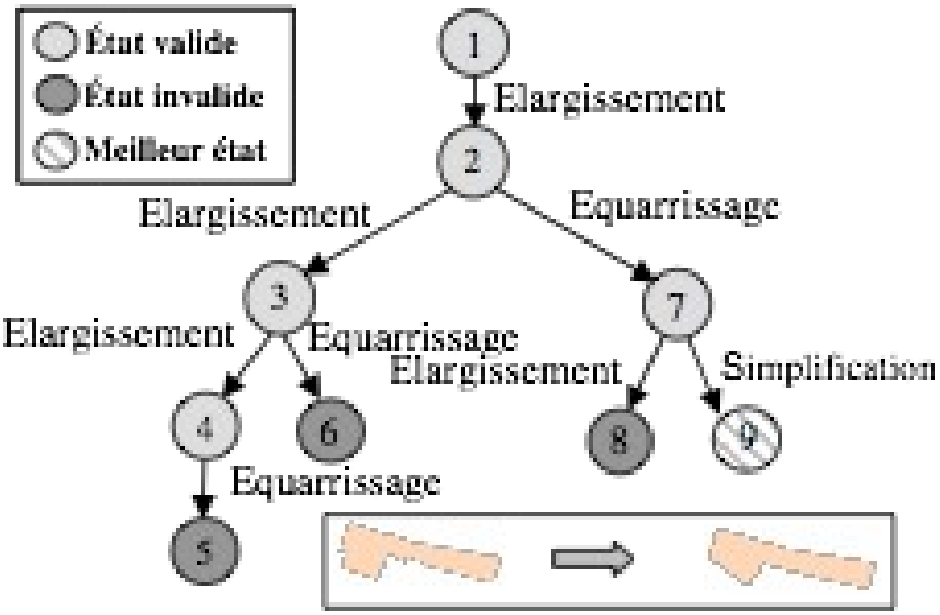

FIG. 2 – Exemple d'arbre d'états pour un bâtiment

L'objectif de ce TPE sera de recoder le modèle AGENT dans GAMA afin de permettre la généralisation de données géographiques. L'étudiant devra en particulier proposer une application pour la généralisation du bâti (bâtiments) dans une ville. Une étude sera nécessaire afin d'établir une liste d'algorithmes de généralisation (les actions des agents) et de contraintes facilement recodables dans GAMA.

19. PORTAGE DE L'APPLICATION SCIENTIFIQUE DANS UNE FÉDÉRATION DE CLOUDS IAAS

Encadrement : Nguyen Hong Quang (nguyen.hong.quang@auf.org), MSI-IFI

Contexte
Le portage de l'application consiste à créer une image disque de base (instantané d'un système d'exploitation) sur laquelle tous les logiciels nécessaires au fonctionnement de l'application auront été installés (customisation). Comme la création des images disques de base peut être fastidieuse (utilisation d'un hyperviseur, démarrage d'un OS, installation des paquets systèmes et contextualisation (voir ci-dessous)), il est intéressant de pouvoir utiliser un instantané d'OS issu du MarketPlace, catalogue d'image développé dans le cadre du projet StratusLab. De plus, l'image disque customisée (contenant le « pipeline ») a été référencée sur ce MarketPlace afin d'automatiser le portage de l'application. Un utilisateur peut donc lancer une machine virtuelle qui sera entièrement configurée pour une exécution immédiate de l'application scientifique. Et toujours grâce à l'interface du MarketPlace, les utilisateurs peuvent avoir une utilisation collective de l'image disque (lancement par plusieurs utilisateurs de machines virtuelles avec la même image disque et/ou utilisation par plusieurs utilisateurs de la même machine virtuelle).

Lors de la création d'image disque, il est nécessaire d'effectuer une contextualisation afin de permettre à l'utilisateur de pouvoir se connecter aux machines virtuelles instanciées sur le Cloud. Cette contextualisation a été réalisée via un procédé générique (CloudInit [5], paquet système supportant un grand nombre d'intergiciel) afin de pouvoir accéder à différentes plateformes de Cloud. Ainsi, nous avons pu utiliser l'image disque customisée aussi bien sur StratusLab que sur une autre infrastructure de Cloud (OpenStack par exemple). L'utilisateur a donc accès à différentes plateforme de Cloud de manière quasi transparente.

Ce TPE adresse le problème de portage d'application entre différents Clouds IaaS.

Travail théorique
Étudier le concept et l'architecture de deux Clouds IaaS [1, 2], le concept Marketplace [3] de StratusLab et le problèmatique du portage de l'image d'un cloud à un autre [4]

Travail pratique
Basé sur les connaissances acquises du travail théorique, chercher une solution pour porter une image du MarketPlace du Cloud StratusLab à celui OpenStack.

Pré-requis
Connaissances de système Linux

Références
1. Stratuslab : http://stratuslab.eu
2. OpenStack, Open Source Cloud Computing Software: http://www.openstack.org
3. StratusLab:Marketplace, https://indico.in2p3.fr/conferenceDisplay.py?confId=6006
4. Vers une fédération de Clouds, http://succes2013.sciencesconf.org/24958/document
5. CloudInit : https://help.ubuntu.com/community/CloudInit

20. PROPOSITION D'UN MODÈLE POUR MODÉLISER LA VARIATION DE L'INONDATION

Encadrant : Nguyen Manh Hung (nmhufng@yahoo.com),
Ho Tuong Vinh (ho.tuong.vinh@auf.org)

Le typhon et leur conséquence comme l'inondation affecte fortement le centre du Vietnam tous les ans. Les dommages peuvent être réduites si on peut prévoir les zones inondés et proposer des scénarios à évacuer les gens (et leurs propriétés) vers les lieux de haut attitude avant l'arrivée du typhon et de l'inondation.

L'objectif de ce sujet est de proposer un modèle à base de système multi-agent pour modéliser l'évolution dynamique de l'inondation dans une région, ou bien sur la longueur d'un fleuve en prenant en compte plusieurs informations: la géographie de la région (le fleuve), le niveau de l'eau à base de pluie et typhon, des zones inondés, des zones non inondés, etc...

Travail théorique

- Proposer un modèle permettant de modéliser: une région ou un fleuve et leur bords, le flux et la taille de l'inondation, la variation de l'inondation en fonction de temps.

Travail pratique

- Développer un modèle de simulation de l'inondation sur un plateforme convient, 2D ou 3D

Références :

1. Hong You (2013). A strategic modelling framework for victim estimation in floods by linking flood and evacuation modelling, A case study: Land Van Maas en Waals . Master thesis at University of Twente, 2013.
2. *Stochastic Modeling of Extreme Floods on the American River* at. *Folsom Dam. US Army Corps of Engineers, RD-48, 2005.*
3. Y. Tramblay, R. Bouaicha, L. Brocca, W. Dorigo, C. Bouvier, S. Camici, and E. Servat (2012). Estimation of antecedent wetness conditions for flood modelling in northern Morocco. Hydrol. Earth Syst. Sci., 16, 4375–4386, 2012.
4. Jeroen C. J. H. Aerts, Ning Lin, Wouter Botzen, Kerry Emanuel, Hans de Moel (2013). Low-Probability Flood Risk Modeling for New York City *Risk Analysis*, Vol. 33, No. 5. (1 May 2013), pp. 772-788.
5. Albert Chen, Slobodan Djordjevic, Jorge Leandro, Barry Evans, Dragan Savic(2008). Simulation of the building blockage effect in urban flood modelling. 11th International Conference on Urban Drainage, Edinburgh, Scotland, UK, 2008.

21. REPRÉSENTER, MODÉLISER ET ANALYSER LE PROCESSUS DE GESTION DE PRÉVENTION ET DE SECOURS EN CAS DE TYPHON

Encadrement : Ho Tuong Vinh (ho.tuong.vinh@auf.org), IFI-MSI

Chaque année le Vietnam est touché par plusieurs typhons. Pour réduire l'impact des typhons, chaque province a des plans de gestion pour la prévention et le secours en cas de typhon. Ces plans sont souvent sous forme textuelle (un document texte). L'objectif de ce TPE est de représenter, modéliser et simuler les activités décrites dans le texte sous forme des processus à l'aide des outils de BPM (Business Process Modeling). Grâce à ce travail, on peut visualiser et observer les interactions entre les acteurs dans le processus de gestion de secours, permettant de faire une évaluation sur l'efficacité de la coordination.

Travaux théoriques

- Étudier les techniques pour la modélisation de processus

Travaux pratiques

- Choisir et appliquer des outils de modélisation de processus pour représenter et modéliser les activités de gestion de secours décrites dans un document texte.

Références

1. Le Nguyen Tuan Thanh, Chihab Hanachi, Serge Stinckwich and Ho Tuong Vinh. **Representing, Simulating and Analysing Ho Chi Minh City Tsunami Plan by Means of Process Models,** ISCRAM Vietnam 2013, http://www.doesnotunderstand.org/public/ISCRAM-VN2013.html

22. RÉSEAUX SENSITIFS ADAPTATIFS

Encadrement : Victor Moraru (victor.moraru@auf.org), IFI
Olivier Camp (olivier.camp@eseo.fr), ESEO

Contexte
La vente de terminaux mobiles (TM) (téléphones intelligents, tablettes, etc.) a explosé pendant les dernières années. Ces terminaux sont en réalité des vrais petits ordinateurs disposant des recoures de calcul et d'affichage assez importantes, des divers capteurs (GPS, accéléromètre, lumière, vidéo, audio, etc.) qui permettent d'élargir considérablement leur fonctionnalités et, ce qui est assez important, ils sont toujours avec leurs propriétaires et ont presque toujours accès au réseau en utilisant des diverses technologies (wifi, 3G, etc.). En se déplaçant en continu, en captant des informations à partir de l'environnement et transmettant cette information sur le réseau, les TM et leurs propriétaires peuvent devenir des acteurs importants pour constituer des réseaux de capteurs urbains (*urban sensing*, en anglais). Le réseau, à son tour, peut bénéficier des informations sur les TM présents dans son entourage (position géographique, valeurs des certaines paramètres mesurés, etc.) pour s'adapter au mieux à des certains besoins spécifiques qui concernent .

Le sujet de ce TPE est liée plutôt a cette deuxième approche : il s'agit d'adapter le contenu diffusé par des terminaux fixes (une sorte d'écrans publicitaires) en fonction du public présent devant eux. Dans un premier temps, on vous propose de faire connaissance avec tout type de réseaux qui pourraient être intéressant dans notre contexte (réseaux adaptatifs, cognitifs, etc.), avec certaines études de cas pour mieux comprendre les méthodes et les approches utilisés. Dans un deuxième cas il s'agit, pour un cas concret qui sera a définir avec les encadrants, de proposer une méthode pour gérer la gestion adaptative du contenu basé sur des connaissances extraites à partir du réseau. Vos propositions seront modélisés et simulés en utilisant la plateforme GAMA.

Travail théorique:
- Bibliographie sur les réseaux urbains sensitifs (urban sensing) et sur les études de cas pour les réseaux adaptatifs, volet sur les réseaux cognitifs ;
- Proposition d'une méthode pour extraire des informations utilises a partir de l'observation du réseau ;
- Proposition d'une méthode pour l'affichage adaptatif en fonction du type du public présent ;
- Conception d'un scénario pour implémenter et tester votre méthode ;
- Étudier l'approche de modélisation à base d'agents.

Travail pratique:
- La prise en main de la plateforme GAMA orientée sur la modélisation à base d'agents
- Concevoir le modèle en utilisant les moyens disponibles dans la plateforme GAMA ;
- Valider votre méthode en faisant des expérimentations avec le modèle GAMA.

Pré-requis
- Connaissances des réseaux sans fil ;
- Une connaissance sur la modélisation à base d'agents est un plus.

Références:
1. Andrew T. Campbell et all., *The Rise of People-Centric Sensing*, IEEE Internet Computing, v.12 n.4, p.12-21, July 2008
2. N.D. Lane et al., *Urban Sensing Systems: Opportunistic or Participatory?* Proc. 9th Workshop Mobile Computing Systems and Applications (WMCSA 08), ACM Press, 2008
3. GAMA: http://gama-platform.googlecode.com

23. SIMULATION D'ÉVACUATION DANS LE CAS D'INONDATION DANS UNE ZONE URBAINE

Encadrants*:* Ho Tuong Vinh (ho.tuong.vinh@auf.org), MSI-IFI
Le Van Minh (lvminh@ifi.edu.vn), MSI-IFI

Contexte

Aujourd'hui, à cause du réchauffement climatique et de la destruction de forêts, il y a souvent les inondation dans les pays tropicaux qui reçoivent une grande quantité de pluie chaque année. Au lieu d'invertir sur solution pour chercher et sauver les victimes (qui semble plus chère mais moins efficace pour les pays du sud), on pense à démarrer l'évacuation pour sauver les gens avant l'inondation arrive. Pour évaluer l'évacuation, on construit le modèle d'évacuation et puis lance la simulation. Le problème est qu'il y a des routes qui sont bloqués pas l'eau et aussi qu'il y a de problème avec les moyennes de transport si le niveau de l'eau est assez haut.

Le but de ce TPE est de créer la simulation d'évacuation de piétons dans le cas d'inondation. Dans cette simulation les comportements de gens prennent en compte sur les problèmes causés par l'augmentation de l'eau (quelques routes sont bloqués, quelques zones sont isolés).

Travail théorique

- Comment créer simulation d'évacuation
    - Comment simuler le phénomène qu'il y a des routes bloqués par l'augmentation de l'eau
    - Comment simuler les comportements de gens dans l'évacuation
- Quels sont les problèmes causés par l'augmentation de l'eau
    - Quelques routes sont bloqués
    - Quelques zones sont isolés
- Construire les scénarios: L'étudiant choisit une ville et puis fait les scénarios d'évacuation

Travail pratique

- Créer la simulation avec GAMA (plate-forme à base d'agent)
    - Simulation d'évacuation de piétons
    - Simulation d'évacuation avec des moyennes de transport (niveau avencé)
    - Simulation de l'augmentation de l'eau (optionnel)
- Évaluer les résultats.

Références

1. Dawson, R. J., Peppe, R. G. and Wang, M. (2011) An agent based model for risk-based flood incident management, Natural Hazards , (doi: 10.1007/s11069-011-9745-4).
2. A. Mordvintsev; V.V. Krzhizhanovskaya; M.H. Lees and P.M.A. Sloot: Simulation of City Evacuation Coupled to Flood Dynamics, in Proceedings of the 6th International Conference on Pedestrian and Evacuation Dynamics. Book of abstracts., 6th International Conference on Pedestrian and Evacuation Dynamics, pp. 156-158. ETH Zurich, Zurich, 2012.
3. Y. LIU, N. OKADA, D. SHEN, S. LI: Agent-based Flood Evacuation Simulation of Life-threatening Conditions Using Vitae System Model, Journal of Natural Disaster Science, 2010

24. SIMULATION POUR L'ORGANISATION DES MOYENS DE SECOURS DANS LE CAS DE TREMBLEMENT DE TERRE EN ZONE URBAINE

Encadrement : Ho Tuong Vinh (ho.tuong.vinh@auf.org), MSI-IFI
Nguyen Manh Hung (nmhufng@yahoo.com)

Collaboration externe : NGUYEN Hong Phuong (IGP, VAST)

Contexte

Dans le cadre d'une collaboration entre l'équipe MSI et la l'institut de Géophysique de la VAST plusieurs travaux se sont développés autour du modèle et de la simulation pour la gestion de secours en cas de tremblement de terre en zone urbain. Grâce au simulateur ArcRisk, développé par l'IGP, avec les données SIG nous avons pu construire des scénarios de tremblement de terre et proposer des outils pour estimer les dommages faits aux infrastructures et les blessés et victimes de ces scénarios. En utilisant ces informations, nous avons pu développer un modèle à base de système multi-agents permettant simuler l'organisation des actions de secours. Des travaux préliminaires ont déjà été conduits permettant expérimenter quelques stratégies de secours, mais beaucoup reste à faire pour que l'on puisse utiliser les résultats des simulations de manière effective.
L'objectif de ce TPE est d'évaluer le modèle existant et proposer des améliorations pour le rendre plus efficace.

Travail théorique

- Faire un survol sur les modèles de gestion de secours au cas de tremblement de terre en zone urbain
- Évaluer le modèle existant et proposer des améliorations
- Proposer quelques scénarios pour évaluer le modèle amélioré

Travail pratique

- Implémenter les améliorations proposées
- Évaluer le nouveau modèle en expérimentant quelques scénarios de gestion de secours

Références

1. Thanh-Quang Chu, Alexis Drogoul, Alain Boucher & Jean-Daniel Zucker. Interactive Learning of Independent Experts' Criteria for Rescue Simulations. Journal of Universal Computer Science, 15(13), pp. 2701-2725, 2009
2. Le Xuan Sang, ArcRisk2GAMA - Interfacer le simulateur de tremblements de terre ArcRisk et la plateforme GAMA, Rapport de TPE, IFI, 2012

25. SOFTWARE DEFINED NETWORKING AVEC OPENDAYLIGHT ET OPENSTACK

Encadrement : Nguyen Hong Quang (nguyen.hong.quang@auf.org), MSI-IFI

Contexte
Le Software-Defined Networking (SDN) est un nouveau paradigme d'architecture réseau où le plan de contrôle est totalement découplé du plan de données. Ce découplage permet de déployer le plan de contrôle sur des plateformes dont les capacités sont plus grandes que celles des commutateurs réseau classiques. Enfin, cette abstraction à travers une API réseau standard permet un développement de services réseau à forte valeur ajoutée affranchi des spécificités des équipementiers.

Le protocole de communication le plus avancé entre un plan de contrôle logiquement centralisé (un ou plusieurs contrôleurs) et le plan de données (des commutateurs réseau) est OpenFlow. Il est standardisé par l'Open Networking Foundation (ONF) et implémenté par de nombreux équipementiers, dont Cisco, IBM, Juniper, HP, NEC et Ericsson. La version 1.0 de la spécification date de février 2011. Dans OpenFlow, les décisions de routage sont prises par le contrôleur pour chaque flux de données et poussées dans les switches sous forme de simples instructions de commutation. Le Software-Defined Networking est un concept-clef pour faire le pont entre la gestion dynamique des ressources réseau d'un côté et la demande en connectivité et en Qualité de service (QoS) des applications de type cloud computing

OpenDaylight est un framework ouvert, dirigé par la communauté, appuyée par l'industrie, pour accélérer l'adoption, en favorisant de nouvelles innovations, réduire les risques et la création d'une approche plus transparente au SDN.

Comme un projet de collaboration sous Linux Foundation, OpenDaylight est structuré selon les meilleures pratiques de développement open source, et est composé des principales organisations de l'industrie de la technologie.

Ce TPE adresse le développement du SDN et en particulier l'approche propose par OpenDaylight et son implémentation sur l'IaaS OpenStack.

Travail théorique
Étudier le concept et l'architecture du SDN [1, 2], le protocole OpenFlow[3] et l'approche OpenDaylight [4].

Travail pratique
Effectuer une installation et configuration de OpenDaylight sur le nuage OpenStack de l'IFI. A définir ultérieurement.

Pré-requis
Connaissances de système Linux

Références
1. SDN (Wikipedia) http://en.wikipedia.org/wiki/Software-defined_networking
2. Software-Defined Networking: The New Norm for Networks, WP, http://www.opennetworking.org/images/stories/downloads/sdn-resources/white-papers/wp-sdn-newnorm.pdf
3. OpenFlow, https://www.opennetworking.org/sdn-resources/onf-specifications/openflow/
4. OpenDaylight Project, http://www.opendaylight.org/

26. SOFTWARE DEFINED STORAGE AVEC OPENSTACK SWIFT

Encadrement : Nguyen Hong Quang (nguyen.hong.quang@auf.org), MSI-IFI

Contexte
Stockage définie par logiciel (Software-Defined Storage - SDS) est un terme de marketing pour promouvoir les technologies de stockage de données informatiques. Basé sur des concepts similaires que les réseaux définie par logiciel (SDN), l'intérêt SDS a augmenté après la société utilisant ce terme a été acquis pour plus d'un milliard de dollars en 2012.

Swift est un système de stockage d'objets multi-locataire, hautement évolutive et durable qui a été conçu pour stocker de grandes quantités de données non structurées à faible coût via une API HTTP RESTful. En tant qu'un des deux composantes initiales du projet OpenStack, Swift est utilisé pour répondre à une variété de besoins. Les plages d'utilisation de Swift sont de petits déploiements pour "juste" le stockage d'images de VM, à la mission des groupes de stockage critiques pour les sites à fort volume, le développement d'applications mobiles, les applications de partage de fichiers privés, l'analyse de données et le stockage privé infrastructure-as-a-service.

Ce TPE adresse le concept et l'architecture du SDS en particulier l'exemple d'un SDS fonctionnel Openstack Swift

Travail théorique
Étudier l'architecture du SDS [1,2], software-defined datacentre (SDDC) [3], Openstack Swift [4].

Travail pratique
Installer et configurer une instance de Openstack Swift sur le nuage Openstack de l'IFI. Travail précis a définir ultérieurement.

Pré-requis
Connaissances de système Linux.

Références

1. SDS, http://en.wikipedia.org/wiki/Software_defined_storage
2. The Fundamentals of Software-Defined Storage, http://san.coraid.com/rs/coraid/images/SB-Coraid_SoftwareDefinedStorage.pdf
3. Software-defined datacentres demystified, http://www.computerweekly.com/feature/Software-defined-datacen
4. Openstack Swift, http://swiftstack.com/openstack-swift/

27. TRAITEMENT DE REQUÊTE TOP-K SUR LES DONNÉES LIÉES *(TOP-K LINKED DATA QUERY PROCESSING)*

Encadrement: VU Tuyet Trinh (trinhvt@soict.hut.edu.vn, vttrinh@gmail.com) ,
Institut Polytechnique de Hanoi

Contexte

Les requêtes top-k sont pour l'objectif de retourner les k meilleurs résultats aux utilisateurs. Par conséquence,les résultats doivent être ordonnes selon des critères. Traitement des requêtes Top-k devient une fonctionnalité importante (voire indispensable) dans un grand nombre de systèmes émergents tels que la surveillance du réseau, de la fabrication, réseau de capteurs, recherche sur le Web, etc. Dans ces systèmes, les données sont distribuées a travers de réseau (client-server, P2P,…) et ont de représentation varie et complexe (structure vs non-structure, graphe, etc.). Cela fait que le traitement des requêtes top-k pourrait être complique en terme de calcul et de communication.

Ce projet a pour but d'étudier les techniques de traitement de requête top-k et en particulier, les requêtes top-k sur les données liées et distribuées.

Travaux théoriques

- Faire un survol des techniques de traitement de requêtes Top-k dans les bases de données relationnelles
- Faire une étude des besoins, des difficultés dans le traitement de requêtes Top-k sur les données liées et l'état actuel
- Identifier des grilles pour l'analyse et la synthèse des techniques de traitement de requêtes Top-k
- Si possible, proposer des améliorations

Travaux pratiques

- Implémenter et comparer quelques techniques afin de compléter la synthèse
- Réaliser une démonstration de requête top-k avec les jeux de données réelles telles que DBLP, google+ circle, etc.

Références

1. Ilyas, I.F., Beskales, G., Soliman, M.A.: A survey of top-*k* query processing techniques in relational database systems. ACM Comput. Surv. **40**(4) (2008)
2. Schlobach, S.: Top-k reasoning for the semantic web. In: ISWC. (2011) 55-59
3. Mouratidis, K., Bakiras, S., Papadias, D.: Continuous monitoring of top-k queries over sliding windows. In: SIGMOD Conference. (2006) 635-646
4. Le-Phuoc, D., Dao-Tran, M., Xavier Parreira, J., Hauswirth, M.: A native and adaptive approach for unified processing of linked streams and linked data. In: ISWC. (2011) 370-388
5. Magliacane, S., Bozzon, A., Della Valle, E.: Efficient execution of top-k sparql queries. In: ISWC. (2012) 344-360
6. A. Wagner, T. Tran, G. Ladwig, A. Harth, and R. Studer: Top-k Linked Data Query Processing. In ESWC 2012.

28. VERS UN MOTEUR DE RECHERCHE SÉMANTIQUE

Encadrant : Dr Nguyen Manh Hung (nmhufng@yahoo.com),
chercheur associé

Contexte

La majorité des moteurs de recherche curant est basé sur le comparaison des mots clés à base de syntaxe. Ce technique donne des résultats de recherche qui ne sont pas bons: beaucoup des résultats non concernés les mots clés, trop de résultats retournés donc difficile à trouver ceux qui sont bons ou non.

Dans ce contexte, le besoin d'un moteur de recherche à base de sémantique est posé naturellement. L'objectif du sujet est de proposer un cadre pour un moteur de recherche sémantique qui se compose de trois composants principaux:

Premièrement, le composant qui génère les mots clés à partir des mots clé originaux. Le résultat est un ensemble des mots clés qui ont relation sémantique aux mots clés originaux. Pour ce faire, on peut utiliser WordNet ou une ontologie.

Deuxièmement, le composant de recherche basique à base de syntaxe. Il prend les mots clés générés comme les entrés, puis cherche les documents qui les contiennent, et puis donne les résultats comme les moteurs de recherche normal qui sont classifié par syntaxe. Pour ce faire, on peut utiliser un moteur de recherche existant, par exemple Google's search engine, etc. ou bien proposer un nouveau moteur de recherche à base de syntaxe.

Troisièmement, le composant de classement les résultats. Il prend les résultats du 2nd composant comme les entrés, puis les classifie en regardant leur classement à base de syntaxe (2nd sortie) et la relation sémantique du mot clé correspondant au mot clé original (1ème partie). Le résultat est un ensemble des documents qui sont classés. Il faut proposer un nouveau algorithme pour le classement.

Travail théorique

- Proposer un cadre générique pour un moteur de recherche sémantique
- Proposer des solutions pour trois composants dans ce cadre:
- une solution pour générer les mots clés relatifs à partir des mots clés originaux
- une solution pour l'utilité d'un moteur de recherche à base de syntaxe: soit utilisation d'un moteur existant, soit proposition d'un nouveau
- une solution pour classer les résultats finaux regardant leur classement à base de syntaxe (2nd sortie) et la relation sémantique du mot clé correspondant au mot clé original (1ème partie)

Travail pratique

- Développer une page Web pour appliquer le moteur de recherche sémantique proposé.

Références :

1. Manh Hung Nguyen and Tan Hiep Nguyen. Towards a semantic search engine based on normal search engines. National Conference on Information Technology and Telecommunications, Hanoi, 4-5 December 2012.
2. Dinh Que Tran and Manh Hung Nguyen. A mathematical model for semantic similarity measures. *South-East Asian Journal of Sciences*, 1(1):32–45, 2012.
3. Fabrizio Lamberti, Andrea Sanna, and Claudio Demartini. A relation-based page rank algorithm for semantic web search engines. *IEEE Trans. on Knowl. and Data Eng.*, 21(1):123–136, January 2009.
4. Renaud Delbru, Nickolai Toupikov, Michele Catasta, Giovanni Tummarello, and Stefan Decker. Hierarchical link analysis for ranking web data. In *Proceedings of the 7th international conference on The Semantic Web: research and Applications - Volume Part II*, ESWC'10, pages 225–239, Berlin, Heidelberg, 2010. Springer-Verlag.

29. VISUALISATION AGILE DE LA SIMULATION MULTI-AGENTS. APPLICATION À L'ÉPIDÉMIOLOGIE

Encadrement: Bui Thi Mai Anh (maianht2@gmail.com), IFI/MSI
Serge Stinckwich (serge.stinckwich@gmail.com), IRD/UPMC

Contexte

Basé sur une approche dite de live coding [1], Roassal est un langage basé sur Smalltalk permettant de scripter rapidement des visualisations de données en fonction des besoins des utilisateurs.

Roassal a été utilisé jusqu'à présent sur des informations statiques (hiérarchie de classes, ...) qui d'une part n'évoluent pas au cours du temps et d'autre part ne sont pas spatialisées. Si on veut pouvoir visualiser des systèmes multi-agents comprenant de nombreuses entités en interaction (simulation multi-robots, nués d'oiseaux (swarms) par exemple), il est nécessaire de pouvoir représenter/manipuler explicitement le temps et l'espace au niveau de l'outil de visualisation.

Calder est un prototype de langage qui a été conçu pour répondre à ces nouvelles exigences. Extension de Roassal, il est proche de celui du langage PROTO [3] développé par le MIT. Un prototype de ce langage de visualisation est disponible ici [4].

L'objectif de ce projet est de d'utiliser Roassal pour développer des visualisations de systèmes multi-agents en utilisant des principes cognitifs de visualisations tels que ceux définis ici [5] comme l'utilisation de couleurs pour afficher les agents qui ont la même direction ou bien de l'utilisation du flou.

De tels principes permettent à l'utilisateur de telles simulations de mieux comprendre les phénomènes.

On utilisera comme exemples de systèmes multi-agents, des applications liés à l'épidémiologie dans le cadre de la thèse de Bui Thi Mai Anh.

Travail théorique

Se familiariser avec Roassal et Calder [2,4], étudier les principes de visualisation multi-agents développés dans l'article [5].

T**ravail pratique**

développer des exemples de visualisation multi-agents au moyen de Calder, faire des propositions de modification du langage Calder.

Références

1. http://flowingdata.com/2012/03/19/live-coding-implemented/
2. Roassal, un outil de visualisation agile : http://objectprofile.com/roassal-home.html
3. PROTO: The langage of Space/Time : http://proto.bbn.com/Proto/Proto.html
4. http://smalltalkhub.com/#!/~UMMISCO/Calder
5. Design Guidelines for Agent Based Model http://jasss.soc.surrey.ac.uk/12/2/1.html