

De H. Glotin, à Embrun, 9-13 août 2012,
à Artieres, Sueur, Adam, Murisasco, Bruno, Chamroukhi, Razik, Paris.

Cher(e)s collègues,

j'espère que vous êtes regonflés à bloc. La cellule de direction de la Mission Interdisciplinaire (MI) MASTODONS CNRS s'est réunie il y a un mois, et a renvoyé ce compte rendu (pièce jointe). Notre projet SABIOD est en fait dans le top 10 par sa 'qualité scientifique et fédératrice' (16 projets sélectionnés sur 37 soumis par 73 laboratoires).

Et j'ai reçu la convocation pour présentation des avancées SABIOD au 5 décembre à la MI.

Vous lirez dans ce rapport l'objectif de MASTODONS vers les traitements en ligne de flux de données, et leur sémantisation suivants ces défis prioritaires :

Défi 1) Tirer parti des bénéfices de GRID 5000 / Cloud

- Pas d'infrastructure à acquérir ni à gérer
- Stockage massif de données à moindre coût
- Accès anytime – anywhere via Internet
- Elasticité : absorbe facilement les charges lourdes ou soudaines
- Calcul haute performance parallèle
- Cohérence et qualité des données

Défi 2) Analyse complexe à grande échelle / sémantisation

- Analyse en temps réel de flots continus de données hétérogènes
- Découvrir et comprendre les patterns caractéristiques
- Réaction en temps réel à des événements d'alerte
- Requêtes multidimensionnelles sur des grands ensembles de données
- Découvrir et comprendre des patterns
- Découvrir des corrélations entre phénomènes

Défi 3) Expertise sur flots d'événements

- Capture d'événements
- Politique de détection et de composition (requêtes sur les événements)
- Introduction d'incertitude sur l'arrivée des événements et sur le contenu de leurs messages
- Détection d'événements rares
- Politique de déclenchement, consommation, exécution (optimisation)
- Bufferisation - Analyse multiéchelle temporelle ou fréquentielle ou spatiale
- Indexation
- Analyse complexe de séquences, de motifs fréquents

Défi 4) Manipulation / visualisation des masses de données

- Navigation intuitive / contextuelle
- Visualisation de phénomènes non perceptibles (faible rapport signal sur bruit)
- Découverte de connaissances par manipulation des experts
- Invention de nouvelles métaphores graphiques
- Algorithmes de graphes performants
- Adaptation aux terminaux / équipements : ANDROID et GRID5000

Synthèse)

- Fin de l'ère « One Size Fits All » : il faut offrir des architectures de données flexibles, avec services de gestion de données adaptables à chaque type d'application / type de données : les SGBD ne sont plus visibles en tant que systèmes intégrés et cohérents
- Comment préserver les données ?
- Coût de conversion des données => formats libres
- Coût pour la migration des technologies => MiddleWare
- Passage à l'échelle et calcul haute performance => GRID 5000
- Virtualisation et indexation sémantique par apprentissage non sup., semi sup. et supervisé
- Requêtage, extraction des connaissances et visualisation
- Produits culturels et environnementaux (in / out ANDROID / diffusion WEB sur Parc Naturel)

==>

Nos Tâches dans SABIOD

=====

Plusieurs tâches ont été définies en mai dans notre projet SABIOD. Les revoici, à préciser d'ici fin août, mais surtout à instancier au moins partiellement pour fin novembre (en construire le cadre, sinon le projet serait stopé). Je les ordonne du niveau 'A:signal', 'B:étiquette' et 'C:sémantique', en citant leurs acteurs a priori (à critiquer!) :

A) Analyse bas niveau (extraction caractéristiques signal):

A1) Clustering en ligne / Modèles de Markov dit non-homogènes [Chamroukhi + Paris]

A2) Segmentation / décomposition atomique Temps-Fréquence (scattering) ou Fourier en ligne [Glotin + Razik + Adam + Paris + Sueur]

A3) Sparse coding [Paris + Razik + Glotin]

A4) Calcul en ligne des indices de diversité types richesse et dissimilarité (<http://rug.mnhn.fr/seewave>) [Sueur + Adam + Glotin]

A5) Mise en évidence non sup. d'éléments constitutifs ou des patrons séquentiels [Artières + Adam + Razik + Chamroukhi]

A6) Mise en place d'un outil en ligne d'aide à l'étiquetage des décompositions, de type "www.labelme", avec prélabels ou clustering (sur 1 et 2) [Bruno + Murisasco]

B) Analyse médiane (labellisation auto) :

B1) Apprentissage non sup / semi sup / sup de classes (multiéchelles) ou de structures de classes (hiérarchiques) [Artières + Chamroukhi + Paris]

B2) Classification sur dizaines de milliers de classes (HTTP://LSHTC.IIT.DEMOKRITOS.GR/LSHTC3_CALL) [Artières + Paris]

B3) Transfert learning (inter-espèces) [Paris + Razik]

B4) Extension hiérarchique [Artières + Glotin + Paris + Sueur]

B5) Corrélation des patterns multiéchelles [Artières + Chamroukhi + Sueur + Adam]

C) Analyse haut niveau (usage, manipulation, expertise et synthèse de données) :

C1) Association des hiérarchies à des ontologies [Bruno + Murisasco + Artières]

C2) Aide à la manipulation experte des index (Olap?), et des blocks (block web) [Bruno + Murisasco]

C3) Découvrir des corrélations entre phénomènes par manipulation des data [Adam + Sueur + Glotin + Bruno]

C4) Génération de résumé (Géographique, Statistique, Temporelle, Acoustique,...) [Bruno + Murisasco + Glotin + Adam + Sueur]

C5) Partage et requêtes sur connaissances hétérogènes [Bruno + Murisasco]

C6) Entrée sortie sur Android [Bruno + Murisasco + Glotin] (depuis et vers Parc Naturel ou Labo ou public)

Technologies

=====

Pour répondre à la stratégie MASTODONS, ces tâches sont à développer suivant trois stratégies technologiques clés :

=> Simplicité et portage des codes (interprété, Python, octave, R, matlab)

=> Portage sur GRID 5000 des modèles de clustering, classification, indexation

=> Portage sur ANDROID pour acquisition et visualisation des data / synthèses

=> Visibilité des données, de leur traitement, et de leur analyse via web (html5) *** EN ANGLAIS ***

Calendrier

=====

Nous devons présenter un premier bilan du prototype de SABIOD le 5 décembre à la MI, pour évaluation (le calendrier MASTODONS est ainsi...) !

Nous recevons à la DR12 CNRS 35 Keu pour construire le prologiciel de notre plateforme SABIOD (à

liquider avant fin 2012!).

Votre AVIS sur CDD à recruter dès le 1er septembre

=====
Nous avons donc 3 mois pour lancer et valider le projet devant le comité scient. CNRS MASTODON. Si nous sommes convainquants, le projet perdure 1, voire 5 ans. Nous sommes complémentaires et motivés, j'espère que SABIOD perdurera plusieurs années.

Nous avons lancé fin juin sur nos listes un appel pour recrutement à partir de bac+5. A ce jour nous avons reçues les 6 candidatures suivantes pour début CDD le 1er septembre, couvrant différents axes (traitement math., apprentissage, SGBD), plus 3 en internes mais qui seraient en mi temps.

==> Nous devons faire notre choix pour lundi 20 août, afin que j'annonce aux sélectionnés leur prise de fonction et que les dossiers CNRS soient rédigés dans les temps (en général 2 semaines de procédure).

J'ai un avis sur chaque candidat, mais je vous laisse renvoyer (REPLY ALL) les vôtres pour débat, espérant que vous aurez 30 minutes pour cette question d'ici vendredi. Nous devrions recruter 2 ou 3 CDDs en parallèle vu le travail annoncé ! J'allonge systématiquement chacun des CDD à 6 mois (donc + 2mois) sur un autre de mes crédits (4 mois de sept à dec. n'étaient pas assez attractifs).

Voici la liste des 6 candidats, mais visez lire dossier svp :

Nicolo Socrate Falzoi : Bac+5 Physique Math. Gène, compétence pour tâche A, bon anglais

Qasim MALIK : Bac+5 IA Paris VI, compétences pour tâches A et B, bon anglais

Sidahmed Benabderrahmane : Dr IA Bioinfo Loria + Ing., compétences pour A et B et C, bon anglais

Nouha BOUJELBEN : Bac+5 LORIA en IA, compétences pour tâches B et C, bon anglais, libre dès 20 nov. (tarde, mais motivée, pourrait plus tôt).

Chaker ABIDI NASRI : Ingé Bac+5 Tunis, compétences pour tâches C, anglais moyen.

Mamadou GUEYE : Ingé bac+5 Toulon, hors sujet.

Autres candidats intéressés qui devraient consommer quelques semaines à mi-temps :

- Un diplômé de l'X, qui souhaite travailler plusieurs mois à mi-temps avec J. Sueur sur tâches A.
- Une étudiante de l'X (Cong_Xue) qui souhaite travailler en stage avec O. Adam (cv joint)
- Un master 2 en écologie et bioinfo (O. Dufour) qui est en stage au LSIS et PNPC depuis janvier et qui commence à être efficace, il pourrait continuer quelques mois s'il ne trouve pas de thèse.

PJ :

- le rapport de la réunion de la MI.
- tous les dossiers complets tels que reçus des candidats.

Bien à vous,
Hervé