

Titre :

Expérimentation du calcul hétérogène dans l'approche VIEWPOINTS

Information :

Encadrants : Fabien Michel (LIRMM, UM2) – fmichel@lirmm.fr
Philippe Lemoisson (Tetis, Cirad) – philippe.lemoisson@cirad.fr
Emmanuel Hermellin (LIRMM, UM2) – Emmanuel.Hermellin@lirmm.fr
Guillaume Surroca (LIRMM, UM2) – Guillaume.Surroca@lirmm.fr

Spécialités : DECOL, MOCA

Nombre d'étudiants : 1

Contexte: [Projet SIFR](#) (Semantic Indexing of French Biomedical Data Resources)

Ou: LIRMM

Quand: 2nd semestre 2013-2014

Mots clés :

GPU, parallélisation, algorithmes de graphe, représentation des connaissances

Technologies :

jCuda ; jOpenCL

Résumé :

Depuis l'émergence du web 2.0, les utilisateurs du web sont énormément sollicités et participent massivement à la production de contenus. Ces sollicitations se transforment en émissions de point de vues divers et variés, explicites ou non. Par exemple, un « I like » sur Facebook, ou un tag de photo sur Flickr ou tout simplement la rédaction d'un Tweet sur un sujet donné. De manière similaire, les données (scientifiques, multimédias, publiques, etc.) sont elles aussi annotées manuellement ou automatiquement pour les enrichir ou les indexer, ce qui représente autant de points de vue émis par les annotateurs. Le LIRMM et le CIRAD travaillent sur un formalisme, nommé **VIEWPOINTS**, de représentation et d'exploitation de points de vue pour aborder de façon homogène les questions liées à la création, à l'extraction et à l'évolution par interaction de connaissances au sein de communautés.

Les graphes de connaissances manipulés peuvent être très gros ; l'algorithme de parcours doit donc être performant. Ce stage a pour objet de proposer une ou des solutions tirant astucieusement partie des deux processeurs des ordinateurs : le CPU et le GPU.

Présentation du contexte

L'approche **VIEWPOINTS** [1], met en œuvre un formalisme simple et puissant qui unifie ontologies et folksonomies

Les *viewpoints* sont émis par les agents (humains ou artificiels); ils expriment la croyance de l'agent relativement à la proximité sémantique entre deux « objets » (agents, documents, topics, goals, plans ...). Chaque *viewpoint* connecte donc trois objets : l'émetteur (nécessairement un agent) et deux objets quelconques. Un *viewpoint* a un *signe* (positif/négatif), un *type* (lié à la sémantique souhaitée) et une *force* ; de plus il est *daté*. Le graphe bi-parti construit sur les « objets » et les *viewpoints* est dénommé *knowledge graph*. Le *knowledge graph* réalise la métaphore neurones-synapses : deux objets seront d'autant plus proches qu'un grand nombre de *viewpoints* positifs les relient. Ceci fonde le calcul d'une distance sur l'ensemble des objets. Les distances dans le graphe évoluent

continuellement au fil des apparitions de nouveaux *viewpoints* positifs et négatifs ; ceci réalise la métaphore de l'évolution de la connaissance collective soumise à la pression des points vue individuels.

L'approche **VIEWPOINTS** est actuellement déclinée dans plusieurs projets avec notamment deux « domaines d'application » : i) la gestion des connaissances sur la base d'un modèle agents-documents-topics et ii) le débat politique sur la base d'un modèle agents-documents-topics-goals-plans. Ces deux modèles sont élaborés sur le formalisme exposé ci-dessus, en définissant à chaque fois spécifiquement les classes d'objets et les types de *viewpoints*.

Description du travail :

L'étudiant prendra part intégrante au travail d'équipe, en particulier dans l'amélioration de l'algorithme de calcul de voisinage sémantique. Il sera attendu de l'étudiant un état de l'art du calcul hétérogène. Qu'est-ce que le calcul hétérogène ? Il s'agit d'une simple idée : pour obtenir la meilleure efficacité, les applications devraient utiliser les deux processeurs des ordinateurs : le CPU et le GPU. Le CPU tend à être le meilleur pour les opérations sérielles, avec beaucoup de branchements et d'accès aléatoires à la mémoire. Le GPU, a contrario, excelle sur des calculs massivement parallèles avec des flottants. L'étudiant proposera et décrira précisément une ou plusieurs solutions visant à faire profiter l'algorithme de calcul de voisinage du potentiel parallèle du GPU.

Il sera demandé une implémentation et une comparaison des solutions décrites.

L'autonomie et la prise d'initiative sont vivement encouragés.

Références :

- [1] Philippe Lemoisson, Guillaume Surroca, and Stefano A. Cerri. Viewpoints: An Alternative Approach toward Business Intelligence. In *eChallenges e-2013 Conference*, page 8, Dublin, Irland, October 2013.
- [2] Yangzihao Wang, John Owens. Large-Scale Graph Processing Algorithms on the GPU
- [3] J. M. Kemp, I. A. Stewart. Parallel Graph Searching Algorithms using CUDA
- [4] Buluç, A., Gilbert, J. R., & Budak, C. (2010). Solving path problems on the GPU. *Parallel Computing*, 36(5), 241-253.