

Chers collègues

Le 25 novembre 2010 a eu lieu la seconde journée de réflexion sur la prospective scientifique du laboratoire: l'objectif était de faire émerger des axes transverses et porteurs. Le matin a été consacré à la présentation courte des challenges de recherche identifiés par chaque équipe. L'après-midi, des axes de recherches transverses et ouverts ont été présentés afin d'initier une première discussion. Ce travail va se poursuivre jusqu'aux vacances en impliquant le groupe prospective, le conseil scientifique et les groupes de travail thématiques ouverts à tous. Une journée LIP6 permettra de faire le point tous ensemble.

Par ailleurs, nous avons appris que l'évaluation du laboratoire aurait lieu entre octobre 2012 et janvier 2013.

Patrick Gallinari
Directeur du laboratoire



Sommaire

- - Edito page 1
- - Serious gaming, apprendre en jouant : un nouveau paradigme éducatif ? / Equipe MOCAH page 2
- - Ethique de la recherche en TIC / Equipe ACASA..... page 3
- - Et si la vidéosurveillance nous faisait du bien ? / Equipe PEQUAN page 5
- - Courte histoire du calcul sur carte graphique / Equipe PEQUAN page 6
- - Archivage du Web / Equipe Bases de Données page 7

Nous dédions ce numéro d'INFOS LIP6 à Nicolas Bataille qui a décidé de nous quitter le 4 janvier. Tous ceux qui l'ont côtoyé au laboratoire sont frappés par un immense chagrin.

Nous tenons à faire part de notre désarroi, à exprimer toute notre tristesse et notre compassion à ses parents, sa soeur Cécile et sa compagne Natacha.

Chantal Perrichon
Directrice de la Communication

Contact : Chantal.PERRICHON@lip6.fr

Serious gaming, apprendre en jouant : un nouveau paradigme éducatif ?

Jean-Marc Labat – Équipe MOCAH

Sur le plan de la formation, nos sociétés contemporaines sont confrontées à un double défi : (i) fournir une formation tout au long de la vie et (ii) transformer le modèle traditionnel de l'enseignement (de type magistral) en un enseignement plus interactif et plus individualisé. L'évolution des méthodes de formation est donc un enjeu majeur de société sur le plan macro-économique mais aussi sur le plan individuel.

Mais quelles sont ces nouvelles méthodes de formation ? Bien entendu, les évolutions résultent de l'apport des progrès technologiques qui permettent aujourd'hui de penser un apprentissage basé très largement sur l'utilisation des TICE¹ : Apprentissage collaboratif, grâce à des environnements spécialisés mais aussi grâce aux réseaux sociaux, apprentissage mobile grâce aux pda, Iphone et autres smartphones, apprentissage pervasif grâce à la communication possible entre l'environnement et ces mêmes machines portables, apprentissage par la simulation grâce à la réalité virtuelle et la réalité augmentée et enfin le serious gaming, autrement dit l'apprentissage par le jeu.

Si l'apprentissage par le jeu n'est pas nouveau en soi, il existe depuis longtemps à l'école maternelle, l'idée que les jeux vidéo puissent



Starbank-thegame (réalisation Ktm-Advance)

servir de support à l'apprentissage dans de nombreux domaines est relativement nouvelle, particulièrement en France. Aujourd'hui, tout le monde veut son serious game ! Beaucoup de grandes entreprises, mais aussi des entités comme universciences (ex cité des sciences et de l'industrie) ou même des universités, comme Villetaneuse qui vient de lancer un appel (marché public) pour développer un serious game sur la découverte des métiers de l'Internet.

Quant à elle, l'équipe MOCAH s'est positionnée sur ce thème dès 2007 en annonçant ce thème de recherche dans les perspectives lors du document pour l'AERES. Cela fait donc 3 ans que l'équipe s'investit dans le domaine des « serious games » avec l'objectif d'apporter de nouvelles réponses au désintérêt croissant des jeunes pour la pédagogie traditionnelle. Il s'agit de s'appuyer sur les mécanismes cognitifs à

l'œuvre chez les joueurs de jeux vidéo comme la notion d'immersion ou celle de « flow » pour inciter les jeunes (et les moins jeunes) à acquérir les connaissances intégrées dans le jeu, en particulier en étant motivés pour rejouer plusieurs fois la même partie. En fait, la notion de « serious game » n'a pas encore une définition reconnue universellement mais, dans le cadre éducatif, on peut distinguer 3 acceptions : (i) utiliser les principes et les technologies (hard et soft) des jeux vidéo sans qu'il n'y ait d'aspect ludique dans le logiciel développé, (ii) envelopper un contenu éducatif dans des séquences ludiques *indépendantes* du contenu (par exemple avec des récompenses), acception appelée parfois « ludo éducatif » et (iii) concevoir des jeux vidéo dans lequel l'objet même du jeu (les actions à faire pour progresser dans le jeu) porte sur le contenu à acquérir. Cette dernière démarche, dans laquelle nous nous inscrivons, est évidemment plus ambitieuse et repose dans la phase amont sur une collaboration étroite entre le « game designer » et l'expert du domaine qui doivent collaborer dans un processus d'ingénierie des connaissances pour établir la nature et la forme des connaissances à transmettre et bien entendu la manière dont le joueur va y être confronté en jouant. Les problématiques de recherche sont nombreuses et enrichissent les problématiques sur lesquelles l'équipe travaille depuis longtemps, à savoir le diagnostic cognitif et les outils auteurs, problématiques qui restent ainsi au cœur de nos préoccupations car, dans les « serious games » comme dans les environnements traditionnels, il faut d'une part évaluer l'appropriation des savoirs et savoir-faire par l'apprenant et d'autre part proposer des outils auteurs aux enseignants/tuteurs qui veulent pouvoir ajouter/modifier/supprimer le contenu, voire même modifier la scénarisation (durée, ordre des connaissances, niveau de difficulté du jeu). Parmi les nouvelles problématiques, l'équipe s'intéresse en particulier à la question des méthodes de conception de « serious games », à la conception d'un framework spécifique permettant d'avoir un ensemble de composants ré-utilisables, aux dialogues Humain-Machine, éventuellement avec des agents conversationnels, et aux questions de scénarisation qui sont cruciales puisqu'il s'agit d'intégrer scénario de jeu et scénario pédagogique dans un modèle unique. Sur ces thèmes, 3 projets sont en cours : un projet financé par l'appel FEDER, et 2 projets obtenus lors de l'appel « Serious Games » lancé par Nathalie Kosciusko-Morizet en mai 2009. Le projet FEDER porte sur l'élaboration d'une méthodologie de conception de Serious Games à partir de Design Pattern et le développement de briques génériques. Quant aux 2 projets « NKM », ils sont très prometteurs. Le premier, Play&Cure, porte sur la conception d'un prototype de jeu pour des étudiants de médecine de 4^{ème} ou 5^{ème} année. Nos partenaires sont 2 PME, Game Consulting et Integrative Bio-Computing, le CHU de Rennes et Telecom SudParis. Il s'agit d'apprendre le raisonnement clinique à partir d'étude de cas (sur le thème de l'œdème) avec un environnement de réalité virtuelle permettant de visualiser les signes cliniques. Un groupe d'étudiants en médecine, gamers patentés, participent au projet en spécifiant leurs attentes sur les plans apprentissage et jeu. Le second, Donjons&Radon, présenté comme emblématique par Nathalie Kosciusko-Morizet lors de la conférence de presse d'annonce des résultats, a pour objet de développer le goût pour les études scientifiques chez des jeunes enfants de collège. Outre une PME spécialisée

(suite...)

RETOUR au SOMMAIRE



Suivi de l'apprenant (Pradeepa Thomas Benjamin)

sur les jeux, Ad-Invaders, on y trouve Microsoft, le groupe Compas de l'institut d'Ulm et le rectorat de Créteil. L'idée, très ambitieuse, consiste à faire découvrir à des élèves de 5ième le plaisir de la démarche et de la découverte scientifique. Avec des enseignants et un inspecteur de physique, nous avons retenu comme thème l'eau et la diffusion de la lumière. Ces 3 projets ont permis de renforcer l'équipe avec 2 post-doctorants et un ingénieur de développement et d'accueillir 2 nouveaux thésards sur ces problématiques, conception d'un framework et modèle de suivi du joueur/apprenant.

Au-delà de l'effet de mode, il est probable que le serious gaming deviendra, avec l'apprentissage mobile, pervasive et collaboratif, sans oublier l'accès à toutes les ressources disponibles en ligne, qu'elles soient statiques comme la vidéo ou dynamiques comme les simulations, l'un des éléments qui permet de penser que l'on va vers

la définition d'un nouveau paradigme éducatif qui amènera les apprenants à être beaucoup plus acteurs de leur formation qu'ils ne le sont aujourd'hui. Les plus jeunes enseignants chercheurs parmi nous enseigneront très probablement bien différemment dans 20 ans, comme l'ARP « l'apprentissage demain : savoirs et outils » nous incite à en poser les fondements. C'est en tout cas le vœu que je forme afin que la France ne reste pas à la traîne des pays européens sur ce plan.

PS : Dernière minute : Le projet FUI « Serious Play » auquel participe l'équipe MOCAH a été accepté

¹ En France, on parle de TICE –Technologies de l'Information et de la Communication pour l'Enseignement , d'EIAH – Environnements Informatiques pour l'Apprentissage Humain , ou encore de FOAD Formation Ouverte A Distance , à l'international, on parle de E-Learning ou, plus récemment, de TEL – Technology Enhanced Learning .

Contact : Jean-Marc.Labat@lip6.fr

Ethique de la recherche en TIC

Jean-Gabriel – Equipe ACASA

Organisée à l'initiative de l'INRIA et du CNRS, une réunion s'est tenue le 12 janvier 2010 à la Maison de la Chimie, à Paris, pour rendre publics deux rapports portant sur l'éthique de la recherche en STIC. L'un de ces rapports émanait du Comité d'Ethique du CNRS (COMETS), l'autre, d'un petit groupe de travail réuni à l'INRIA, sous la responsabilité de Claude Kirchner. Les directeurs des ces deux organismes profitèrent de cette rencontre pour charger ALLISTENE (Alliance des Sciences et Technologies du Numérique) de réunir une commission interdisciplinaire sur ce sujet. En février prochain, au cours du congrès SPECIF, se tiendra un atelier sur l'enseignement de l'éthique et des dimensions sociétales de l'informatique. Dans la journée des doctorants qui précède cette manifestation, une table ronde se consacrera, elle aussi, à l'éthique de la recherche en STIC. Bref, longtemps ignorées dans notre pays, les questions d'éthique des STICs, et, en particulier, les questions d'éthique de la recherche en STIC, semblent aujourd'hui faire l'objet d'une certaine attention dans la communauté des chercheurs et des enseignants-chercheurs en informatique.

À n'en pas doute, cela provient du succès des STICs qui transforment, jour après jour, notre société. Travail, éducation, culture, guerre, politique, journalisme... les ordinateurs modifient toutes les activités humaines ; ils changent la vie. Les normes qui régissent les rapports inter-humains s'en trouvent bouleversées. Pendant longtemps, les informaticiens en général, et les chercheurs en informatique, en particulier, n'y prenaient pas garde ; ils se réfugiaient derrière leur innocence et la pureté de leurs intentions pour éluder ces questions. Aujourd'hui, cette attitude n'est plus admissible. Mais, dans le même temps, et pour paraphraser Georges Clemenceau , l'informatique devient une affaire trop sérieuse pour la laisser aux mains des seuls informaticiens. Les ordinateurs prennent désormais une place si grande qu'il appartient à tous d'en évaluer les bénéfices et d'en dénoncer les excès. On peut alors se demander quel rôle particulier incombe aux chercheurs en informatique dans la réflexion éthique, quelle place cette réflexion est-elle susceptible de prendre dans un laboratoire et quelles solutions les universitaires contribueront-ils à apporter. Pour comprendre les enjeux, prenons quelques exemples concrets.

(suite...)

Le premier tient à l'intimité de la vie privée. Avec les cartes bancaires, les passes Navigo, les cartes Vitale, la télédétection spatiale, les téléphones portables, la géolocalisation, les réseaux sociaux, les courriels, les blogs, les micro-blogs, etc. toutes les informations sur nos achats, nos déplacements, nos visites chez le médecin, nos médicaments, nos courriers électroniques, nos affinités, etc. sont connues. Si on les croisait toutes, on suivrait à la trace chacune des actions de chacun d'entre nous, au point de déceler si ce n'est l'ensemble des pensées individuelles, tout au moins l'état d'esprit dans lequel chacun se trouve, à chaque moment de son existence. Beaucoup trouvent cela désagréable, voire même insupportable. D'autres souhaitent en profiter pour instituer une transparence totale et pour l'étendre à l'ensemble de la société afin d'en finir avec l'ère du secret. La mise en place du gouvernement 2.0 par le président Obama au début de son mandat atteste de cette tentation, puisqu'il s'agissait de donner un accès libre à toutes données publiques aux Etats-Unis ; l'attitude ambivalente de nos contemporains vis-à-vis des révélations de Wikileaks révèle les difficultés et les limites de cette entreprise. La question reste donc ouverte.

Il n'en demeure pas moins que l'exploitation systématique et sauvage des informations personnelles donne le frisson. Sans doute, conviendrait-il non seulement d'interdire, mais aussi d'empêcher que la fusion des données individuelles s'opère à notre insu. Cependant, les systèmes de recommandation, les publicités adaptatives et, plus généralement, les techniques de fouille de données que nous mettons au point dans nos laboratoires tendent à faire leur beurre de toutes ces informations. Comment s'assurer que cela reste conforme à des normes édictées ou à des standards définis par tous ? Cette question ouverte fait déjà l'objet de recherches. C'est en particulier l'objet du . Elle nous concerne directement.

Notons, à ce propos, qu'en tant que chercheurs, il nous appartient aussi d'éprouver la sécurité des systèmes informatiques en allant jusqu'aux limites du possible, au risque, parfois, de nous mettre hors la loi. Cela nous place, individuellement et collectivement, dans des situations juridiques difficiles. Nous devons donc réclamer et obtenir un statut dérogatoire, une sorte de franchise universitaire qui nous prémunisse contre des accusations probables, à condition bien évidemment, de ne pas abuser des données récoltées lors de nos expérimentations.

À ces questions relatives à la vie privée et à la sécurité, s'en ajoutent bien d'autres. Certaines portent sur la responsabilité des concepteurs, des fabricants ou des propriétaires de robots autonomes. Pendant longtemps, on anticipait assez facilement ce qu'un automate matériel faisait. Aujourd'hui, les dispositifs qui nous entourent comportent, de plus en plus souvent, des organes de calcul électroniques. Ainsi en va-t-il, par exemple, des automobiles, des machines à laver, des agendas électroniques ou des appareils photographiques. Leur complexité fait qu'il devient impossible à leurs concepteurs, à leurs utilisateurs ou à leurs propriétaires, d'anticiper leurs comportements dans toutes les situations. Dès lors, qui doit être tenu pour responsable en cas d'accidents ? Ajoutons à cela, la dimension sociale des dispositifs matériels qui, désormais, interagissent avec d'autres dispositifs qui eux-mêmes répondent à des normes définies collectivement. Ainsi en va-t-il des navigateurs GPS qui recourent à des techniques de géolocalisation. L'artefact ne s'envisage donc plus isolément ; la responsabilité de ses concepteurs ou de ses utilisateurs ne se définit plus indépendamment de l'ensemble du système social au sein duquel il a été conçu, fabriqué et utilisé. On parle alors d' pour caractériser les objets techniques qui nous environnent et au milieu desquels il nous est donné de vivre. Comme ceux-ci ne prennent sens qu'au regard de la société, le chercheur qui contribue à rendre leur conception possible ne peut plus ignorer les conséquences sociales et éthiques de son travail.

Pour illustrer ce dernier point, évoquons la robotique militaire. Grâce aux progrès réalisés dans nos laboratoires, on fabrique des véhicules dotés d'une totale autonomie de déplacement. Le champ de bataille est donc parcouru d'aéronefs et de chars d'assaut sans pilote. Des automatismes en assurent la navigation. Parfois, ils se contentent de prendre de l'information, d'autres fois, on les équipe d'armes offensives afin qu'ils participent activement aux combats. Jusqu'à présent, la décision tactique, à savoir le déclenchement des tirs, était prise à distance, par des hommes. Or, depuis peu, on réalise des systèmes d'armes totalement autonomes. Certains d'entre eux ont été placés à la frontière entre la Corée du Nord et la Corée du Sud. En songeant à ces robots soldats, comment ne pas s'interroger sur le sens de nos travaux ?

Vie privée, véhicules autonomes, robotique militaire, aucune des questions évoquées ici ne laisse insensible. Qui plus est, indépendamment de nos sentiments et de nos éventuels atermoiements, un dispositif technique dont l'emploi aurait des conséquences contraires aux normes éthiques en vigueur ternirait l'image de la société qui le fabrique. Bref, plus que jamais, il appartient au chercheur d'anticiper, au plus tôt, les conséquences néfastes des dispositifs qu'il contribue à mettre en place et, dès qu'il en a connaissance, de tout mettre en œuvre pour empêcher qu'elles ne se réalisent. À cette fin, il faut qu'il soit alerté. C'est le rôle assigné à la commission interdisciplinaire qui doit se mettre en place dans le cadre d'ALLISTENE. Mais, en même temps, il doit infléchir ses orientations scientifiques, voire contribuer, par ses propres travaux, à trouver des solutions nouvelles, à des problèmes nouveaux. Enfin, il doit éclairer la société sur les conséquences de certains choix techniques, comme par exemple, ceux qui portent sur le vote électronique. Bref, si, comme nous l'avons dit, l'informatique devient une affaire trop sérieuse pour la laisser aux mains des seuls informaticiens, c'est aussi une affaire trop complexe pour en traiter sans les informaticiens.

Contact : Jean-Gabriel.Ganascia@lip6.fr

Et si la vidéosurveillance nous faisait du bien ?

Séverine Dubuisson – Équipe PEQUAN



De nos jours, la vidéosurveillance est souvent associée au « flicage » et à la perte des libertés individuelles telle qu'a pu nous le décrire, au début du siècle dernier, George Orwell. Pourtant, il existe bien des situations dans lesquelles elle pourrait nous être d'une grande utilité, en particulier quand on cherche, non plus à l'utiliser à des fins d'identification, mais plutôt pour l'analyse de comportements. Cette analyse peut prendre plusieurs formes.

Tout d'abord, la possibilité d'identifier un comportement anormal, voire dangereux, parmi un ensemble de comportements. Dans ce cas, ce sont essentiellement les lieux publics qui sont visés. Tous les ans, une compétition entre chercheurs (PETS: Performance Evaluation of Tracking and Surveillance) a lieu conjointement avec deux des meilleures conférences internationales de vision (ICCV, CVPR), pour comparer les dernières méthodologies issues de la recherche académique autour de l'analyse de situations. Le thème est défini chaque année parmi lesquelles l'estimation de la densité d'une foule, la détection d'abandon de bagage dans un hall de gare, de bagarre dans une foule, d'une trajectoire anormale (d'une voiture sur une route ou d'un piéton dans un lieu public), autant de situations qui pourraient, si les outils sont bien utilisés, renforcer notre sécurité.

Mais la vidéosurveillance peut également être utilisée à des fins thérapeutiques. Aujourd'hui, et fort heureusement, les êtres humains vivent de plus en plus âgés et en meilleure santé à leur propre domicile. Equiper de caméras les logements de personnes âgées ou affectées d'un trouble mental a un intérêt double : tout en gardant leur indépendance, elles pourraient aujourd'hui bénéficier du fait qu'on leur vienne rapidement en aide grâce à la vidéosurveillance. Une thématique du LIP6, intégrée à l'équipe PEQUAN, s'intéresse à la modélisation de systèmes dynamiques et travaille actuellement sur l'analyse du comportement. Une collaboration avec Jean Meunier, professeur à l'Université de Montréal, laboratoire DIRO, travaille d'une part sur la détection de chute par vidéosurveillance multi caméra, et l'analyse de démarche d'autre part. Un système de d'analyse du mouvement de la personne et de détection automatique d'événements anormaux, telles que les chutes est en cours de développement et permettrait d'envoyer une alarme vers les services de secours en cas de nécessité. Un deuxième travail dans le cadre de cette même collaboration concerne l'analyse vidéo de la démarche de patients de manière à déterminer les soins à administrer en fonction de la trace temporelle qu'on en aurait. Equipés d'indices visuels, les algorithmes permettraient une analyse bien plus fine qu'actuellement de manière à déterminer quelles articulations sont atteintes et la gravité de la pathologie par rapport à une démarche « normale ». Enfin, un projet en collaboration avec l'équipe Décision du LIP6 ainsi que l'équipe TII du LTCI (Télécom ParisTech) concerne plus précisément les personnes atteintes d'un trouble mental. Un système de vidéosurveillance est en cours de développement afin de vérifier que ces personnes effectuent les tâches nécessaires au quotidien, en particulier la prise de médicaments indispensables pour le patient. Là encore, il est nécessaire de modéliser le comportement du patient à l'aide d'une suite d'analyses de mouvements élémentaires (remplir un verre d'eau, prendre le médicament, l'amener à la bouche, boire, reposer le verre), le tout pour vérifier la cohérence de cet ensemble.



La vidéosurveillance permettra sûrement d'améliorer notre sécurité et notre confort au quotidien. En prenant soin de ne pas viser l'identification de la personne filmée, c'est son attitude dans un contexte précis, en relation avec ce qui l'entoure qui sera prépondérant.

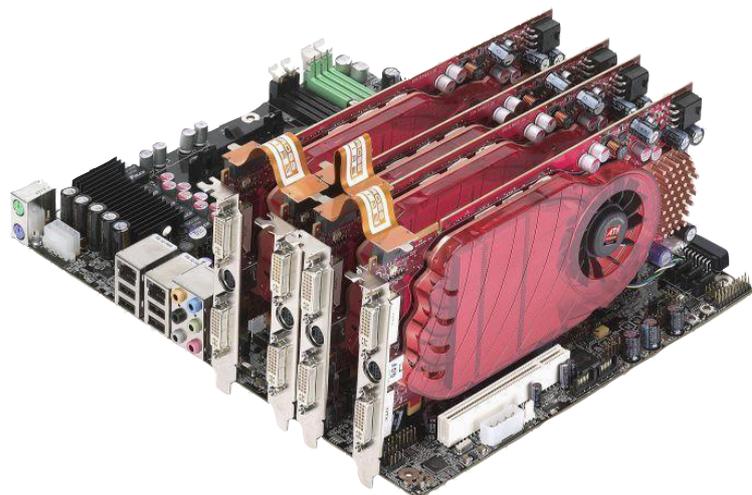
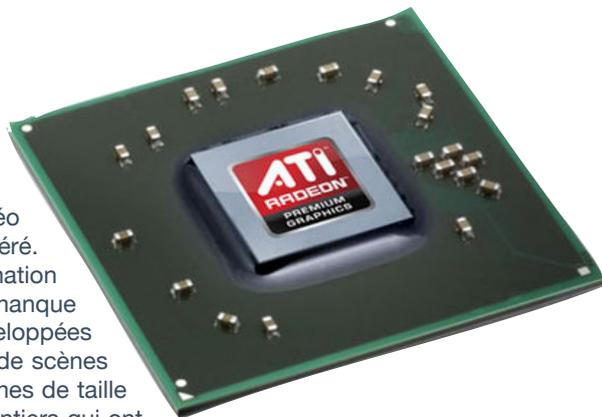
Séverine Dubuisson
Enseignant-Chercheur
Laboratoire d'Informatique de Paris 6 (LIP6)
104 avenue du Président Kennedy 75016 Paris, Bureau 702
Tel.: +33 (0)1 44 27 71 29 Fax: +33 (0)1 44 27 74 95
<http://lip6.fr/Severine.Dubuisson>

Contact : Séverine Dubuisson

Courte histoire du calcul sur carte graphique

Jean-Luc Lamotte - Equipe PEQUAN

Dans l'informatique grand public, le multimédia et plus particulièrement les jeux vidéo d'action occupe une place très particulière par sa prospérité et le chiffres d'affaires généré. La génération des sacro-saintes 25 images par seconde nécessaires pour une animation fluide et le besoin de fournir des images toujours plus réalistes se sont heurtés au manque de puissance des processeurs généralistes. Des cartes spécialisées ont alors été développées : les GPU (Graphic Processing Unit). Elles sont spécialisées dans la représentation de scènes 3D et le plaquage de texture qui constitue les opérations de base. Pour des problèmes de taille mémoire, l'ensemble des éléments des scènes était représenté avec des nombres entiers qui ont vu leur taille augmenter au fur et à mesure de leur évolution : 8, 12, 16 bits. Les développeurs de jeu ont répondu à la demande du grand public en créant des images de plus en plus réalistes. Le marché était en plein essor et à la fin des années 1990, le nombre de constructeur de cartes graphiques était important : 3Dfx, ATI, Cirrus Logic, Hercules, Matrox, Nvidia, Tseng Labs, Trident, ... Il y avait cependant deux ombres au tableau : d'une part, les calculs effectués avec des entiers nécessitaient d'utiliser un grand nombre d'astuces de calcul qui compliquaient l'écriture des programmes et d'autre part, la recherche d'un rendu réaliste obligeait à prendre en compte les phénomènes physiques : absorption, diffusion, réflexion diffraction, ... Les fabricants de GPU se sont alors lancés dans le développement de processeurs beaucoup plus complexes. Début 2000, les nombres flottants 16 bits apparaissent sur les cartes graphiques NVIDIA et un an plus tard les shaders. Ces derniers permettaient aux programmeurs d'écrire des petits morceaux de codes (des shaders) pour améliorer le rendu des images. Au niveau de l'architecture, les processeurs graphiques intégraient des capacités de calcul vectoriel et des caches. En 2002, les cartes graphiques et le processeur Intel Pentium 4 avaient des puissances de calcul équivalentes de l'ordre de quelques GFLOPS. Il ne restait plus que deux principaux fournisseurs : AMD et NVIDIA. En 2010, les cartes graphiques les plus puissantes ont deux processeurs graphiques et atteignent 4 TFLOPS en arithmétique virgule flottante 32 bits IEEE 754 alors que les processeurs multi-core ne dépassent pas actuellement les 300 GFLOPS.



Le GPGPU (General-Purpose computation on Graphics Processing Units) est apparu avec les shaders. Rapidement, un certain nombre de spécialistes en calcul scientifique se sont rendus compte que les GPU pouvaient être détournés pour faire du calcul intensif haute performance pour un faible coût. En utilisant OpenGL ou DirectX, ils ont décrit des opérations matricielles sur le CPU, transféré les données sur le GPU, exécuté le code sur le GPU et récupéré les résultats sur le CPU. Les bases du GPGPU étaient posées. Un nouveau paradigme de programmation est né : le stream processing. Conceptuellement, il s'agit d'exécuter sur chaque donnée d'un programme les mêmes instructions sans expliciter l'allocation aux unités de calculs, les communications et les synchronisations. Brook a été le premier langage accessible dédié au GPGPU. C'est une surcouche de C. Il a été développé à partir de 2004 par le Stanford University Graphics group. En 2007, NVIDIA a diffusé son propre langage : CUDA. Les gros moyens en développement et ses capacités à atteindre facilement de très hauts niveaux de performance ont permis à NVIDIA de s'imposer dans le monde du calcul scientifique. Insuffisamment soutenu, le développement Brook a été arrêté.

La convergence du monde du multimédia et du calcul intensif a été réalisée autour d'un même support. Les cartes GPU sont capables d'atteindre de très hauts niveaux de performance dans les deux domaines. En décembre 2010, la première machine au TOP500 est basée sur une architecture hybride CPU-GPU. Elle a atteint 2,57 PFLOPS avec le benchmark LINPACK.

Les perspectives pour le GPGPU sont larges. Au niveau de l'architecture, les deux constructeurs de cartes graphiques annoncent des cartes toujours plus puissantes. Mais AMD développe un projet innovant : le projet FUSION. Il a pour objectif d'intégrer un processeur graphique dans un processeur multi-cœur permettant ainsi d'augmenter considérablement la puissance de calcul des CPU et de supprimer les coûteux transferts mémoire CPU-GPU. Dans le domaine du logiciel, la communauté industrielle et scientifique a compris l'importance du GPGPU car les cartes graphiques se retrouvent dans beaucoup de systèmes informatiques : des téléphones portables aux grands calculateurs. En 2009, une norme ouverte a été définie par le Khronos Group : OpenCL (Open Computing Language). C'est la combinaison d'une API et d'un langage de programmation dérivé du langage C. Elle a pour vocation d'offrir un langage de programmation commun à tous les systèmes comprenant un CPU et un GPU. Une question reste : est-il possible de programmer les téléphones portables et les super-ordinateurs avec les mêmes outils ?



Contact : Jean-Luc.Lamotte@lip6.fr

[RETOUR au SOMMAIRE](#)

De très grandes quantités d'informations sont actuellement disponibles sur le Web. Au même titre que toute information, il est important d'en conserver une trace. Avec le développement du Web, son archivage est devenu un enjeu majeur pour la préservation du patrimoine.

En général, l'archivage du Web est effectué de façon automatique, à l'aide de robots (ou crawlers), qui parcourent le Web à intervalles réguliers et aspirent les pages à archiver. Pour maintenir les archives à jour et préserver leur qualité, ces robots doivent revisiter périodiquement les pages et répercuter les modifications. Cependant, à cause de la très grande quantité de pages à archiver, de leur forte évolution, et des limites des capacités de stockage et de bande passante, il n'est pas possible de revisiter toutes les pages d'un site et de capturer toutes les modifications de manière exhaustive. Les archives seront donc forcément incomplètes.

L'objectif d'un système d'archivage est de préserver l'information à des fins de réutilisation. Il faut donc fournir des archives de la meilleure qualité possible et un langage d'interrogation permettant de retrouver les archives en fonction de critères temporels et de critères sur le contenu. L'équipe Bases de Données s'intéresse, en partenariat avec l'INA, à ces deux aspects dans le cadre du projet ANR CARTEC (www.projet-cartec.org) dont l'objectif est d'archiver les sites de l'audiovisuel français.

La qualité d'une archive se définit par sa complétude et par la cohérence de ses versions de pages. La complétude s'évalue en fonction de la proportion de versions utiles contenues dans l'archive, par rapport à celles ayant existé dans le Web. La cohérence assure que si un utilisateur accède à une version d'une page d'un site, il peut aussi accéder à d'autres versions de pages du même site correspondant au même point dans le temps. En d'autres termes, des versions de pages (d'un même site) sont cohérentes si elles ont existé en même temps sur le Web. Comme il n'est pas possible d'archiver toutes les pages d'un site de façon instantanée, le risque que certaines pages aient été modifiées entre le début et la fin de l'archivage n'est pas négligeable.



Pour garantir une bonne qualité des archives, nous proposons une approche basée sur l'importance des modifications dans les pages. Certaines modifications, comme la publicité ou la décoration, qui ne concernent pas le contenu sémantique de la page sont considérées comme peu importantes et peuvent être ignorées. Cette approche nous permet d'améliorer la complétude de l'archive en ne conservant que les versions les plus importantes (celles dont les modifications sont significatives, au sens où elles concernent des informations importantes), et d'assurer une bonne cohérence en garantissant que les versions cohérentes qu'on a obtenues sont les plus importantes. L'idée est que certaines incohérences sont considérées comme moins importantes que d'autres, car elles concernent des pages et des modifications moins importantes.

Comme le système d'archivage ne contrôle pas les modifications sur les sites, il est nécessaire de prévoir le comportement des sites en observant leur comportement passé. Pour cela, le système doit comparer les versions d'une même page qu'il a préalablement capturées. Les méthodes classiques pour détecter les modifications dans les pages reposent sur des algorithmes d'analyse de texte. Une contrainte forte dans ce projet étant de conserver l'aspect visuel des pages (puisque'il s'agit de sites audiovisuels), nous avons choisi de nous appuyer sur une analyse visuelle des pages pour déterminer les modifications importantes. L'algorithme découpe la page en blocs, et permet d'attribuer une importance à chaque bloc de la page en fonction de son emplacement, de sa taille, de son contenu. Typiquement, une information importante se situe en milieu de page, une publicité se trouvera en haut de la page, un copyright en bas à droite. L'importance des modifications d'une page se calcule en fonction de trois paramètres qui sont l'importance du bloc modifié, l'importance de l'opération effectuée (ajout, suppression, modification) et de la nature de l'élément modifié (texte, image, lien...), et du nombre de modifications par bloc.

Cette méthode permet des gains substantiels en termes de complétude de l'archive. Elle sera étendue afin d'améliorer la cohérence. D'autres gains proviennent de l'utilisation de patterns, qui mémorisent l'importance des modifications faites sur une page à différents moments de la journée. Les autres approches se contentent d'un taux de changement moyen, moins précis.

Un autre aspect de nos travaux porte sur l'utilisation de l'archive : recherche d'information, requêtes et navigation. Le problème est de tenir compte à la fois de la dimension temporelle de l'archive, de son incomplétude ainsi que du comportement propre aux utilisateurs d'archives (chercheurs en sciences humaines pour une grande part).

Contact : Anne.Doucet@lip6.fr - Stephane.Gancarski@lip6.fr

- Directeur de la publication : Patrick Gallinari
- Comité de rédaction : Chantal Perrichon, Pierre Sens
- Maquette : Frédéric Delvalle

[RETOUR au SOMMAIRE](#) ▣