CENTRE DE RECHERCHE COMMUN
Inria - Microsoft Research

BIENVENUE DANS UN MONDE OÙ L'INNOVATION NAÎT DE LA COLLABORATION

© INRIA / C. DUPONT

MICHEL COSNARD
Président-Directeur général d'Inria

© ALEX MAHIEU

ALAIN CROZIER
Président de Microsoft France

ÉDITORIAL

Microsoft Research et Inria ont fait le pari d'ouvrir un laboratoire de recherche fondamentale commun au cœur de la future université Paris-Saclay. Avec un objectif ambitieux : créer, en France, un centre de recherche public-privé de rang mondial autour de problématiques-clés pour l'industrie numérique et le progrès scientifique.

Ce partenariat est venu sceller une tradition de relations étroites entre les chercheurs des deux institutions, fondées sur la confiance et le respect mutuels, et une même vision stratégique.

Aujourd'hui, le Centre de recherche commun constitue le plus important des partenariats industriels d'Inria. Il a permis de faire venir des chercheurs de très haut niveau en région parisienne et d'obtenir des résultats de premier plan. En témoigne le prestigieux prix Turing attribué à Leslie Lamport en mars 2014.

Cette démarche démontre l'engagement à long terme de Microsoft au sein de la communauté scientifique, en France, et sa contribution à un effort de recherche ouvert entre laboratoires publics et laboratoires industriels. Inria et Microsoft ont prouvé par là-même qu'une collaboration fructueuse est possible entre un institut de recherche public français et une entreprise privée américaine, et qu'elle peut bénéficier à la société tout entière.

Fruit d'un investissement partagé dont les résultats sont publics et mis à la disposition de la communauté scientifique internationale, le Centre constitue un pôle d'excellence scientifique et d'attraction mondiale.

Le laboratoire s'ouvre aujourd'hui à de nouveaux champs de recherche, tels que la sécurité des services Internet, les applications du Big Data à la médecine et à la biologie, ou la protection de la vie privée en ligne. Nous vous invitons à partir à sa découverte en plongeant au cœur du laboratoire, aux côtés de chercheurs qui façonnent, aujourd'hui, le monde de demain.

MICHEL COSNARD, ALAIN CROZIER

CENTRE DE RECHERCHE COMMUN
Inria - Microsoft Research

Centre de recherche commun Inria-Microsoft Research
Campus de l'École polytechnique, bâtiment Alan Turing
1, rue Honoré-d'Estienne-d'Orves - 91120 Palaiseau, France
Tél. : 01 65 35 69 69 – Fax : 01 69 35 69 69
www.msr-inria.fr

Conception et réalisation graphique : Virginie Kahn – Rédaction : Spintank/Centre de recherche commun Inria-Microsoft Research - Secrétariat de rédaction : Pauline Feuillâtre - Impression : Impression GRAPH 2000 – Tél. : 02 32 43 43 43 – Document imprimé sur du papier issu de forêts gérées durablement, avec des encres végétales. Avril 2014

SOMMAIRE

INVENTER AUJOURD'HUI LE MONDE DE DEMAIN 2

INTERVIEW **de Laurent Massoulié** 4
« Sans cette culture partagée de l'excellence scientifique, le partenariat n'existerait pas »

UN CENTRE DE RECHERCHE DE RANG MONDIAL 8

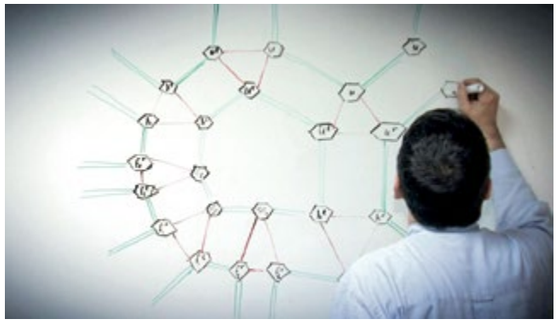
UN IMPACT SCIENTIFIQUE MAJEUR 10

QUATRE GRANDS AXES DE RECHERCHE 15

LES TROIS CLÉS D'UN LABORATOIRE OUVERT SUR LE MONDE 27

4

Laurent Massoulié, directeur du Centre de recherche commun, nous ouvre les portes d'un laboratoire unique en son genre et nous plonge au cœur de ses problématiques.



10

Utiliser un ordinateur pour faire la démonstration irréfutable d'un théorème mathématique, explorer le fonctionnement du cerveau grâce au cloud, ou encore prévenir des maladies grâce à un modèle numérique de cœur humain : partez à la découverte des projets les plus marquants du laboratoire.

27

Le laboratoire commun s'inscrit dans une logique d'ouverture et de partage avec la communauté scientifique mais également académique et entrepreneuriale.

15

Big Data, apprentissage statistique, imagerie médicale, méthodes formelles… : les chercheurs du Centre de recherche commun inventent au quotidien les technologies qui révolutionneront demain notre vie quotidienne.

Comment **mieux ajuster les thérapies** aux particularités de chaque malade ?

© COLLÈGE DE FRANCE / PATRICK IMBERT

NICHOLAS AYACHE

L'imagerie médicale est un domaine qui connaît des évolutions sans précédent avec l'avènement de la 3D et de la modélisation. Les progrès sont tels qu'il est désormais possible, à partir des images d'un individu, de modéliser les organes et leurs pathologies afin de créer un patient virtuel suffisamment réaliste pour guider le diagnostic, le choix, la planification et l'application d'une thérapie. On peut ainsi représenter de manière précise des tumeurs et surtout surveiller et simuler des évolutions en 4D, c'est-à-dire dans l'espace et dans le temps : cela pourrait devenir un élément-clé d'appréciation pour déterminer et ajuster les thérapies.

Comment **mieux protéger notre vie privée** en ligne ?

© BERNARD LACHAUD

LAURENT MASSOULIÉ

Nous accédons à des services en ligne de plus en plus personnalisés, car fondés sur une connaissance toujours plus fine de nos usages : nos déplacements, achats, lectures, contacts, conversations sont mis à contribution pour mieux anticiper nos besoins. Pour bénéficier de cette personnalisation, devons-nous renoncer à toute protection de notre vie privée ? Certainement pas ! Nos données privées peuvent être « anonymisées », cryptées, ou encore délibérément distordues avant d'être soumises aux fournisseurs de services en ligne. Tout l'enjeu est de développer ensemble les mécanismes de chiffrement ou de bruitage de données pour garantir la protection de notre vie privée, et les outils de personnalisation de services opérant sur ces données modifiées.

INVENTER AUJOURD'HUI LE MONDE DE DEMAIN

De la recherche fondamentale à notre vie quotidienne, il n'y a souvent qu'un (petit) pas. Illustration avec quatre questions essentielles pour le monde d'aujourd'hui et de demain, sur lesquelles travaillent les chercheurs du Centre de recherche commun.

Comment **prévenir les « bugs »** quand tout devient numérique ?

Objets connectés, avez-vous une âme ? Les objets qui nous entourent sont de plus en plus autonomes, intelligents et connectés. S'ils se mettent à ressembler à des ordinateurs, alors se pose la question de leur fiabilité ou de leur sécurité : vont-ils se comporter comme prévu ? Un défaut ne peut-il pas être exploité pour tromper leurs utilisateurs ? La recherche du Graal que représente le « zéro bug », c'est le sens des travaux conduits dans le Centre de recherche commun autour des outils de « preuve de programmes » : l'idée n'est plus de repérer le défaut de conception d'un logiciel par un test a posteriori des erreurs qu'il peut produire, mais au contraire de vérifier en amont la logique de sa construction, sa « correction » au sens mathématique du terme. Il s'agit de techniques d'avant-garde déjà utilisées dans des domaines très sensibles tels que l'aéronautique, mais encore très loin d'être parvenues à maturité. Au laboratoire commun, nous travaillons à la fois sur des outils de preuve pour les mathématiciens et sur des applications centrées sur la sécurité du logiciel.

CÉDRIC FOURNET

© MICROSOFT RESEARCH

Comment **donner du sens** au déluge de données ?

© INRIA / J.M. RAMÈS

FRANCIS BACH

Chaque jour, une multitude de données est créée, au point que nous vivons aujourd'hui dans un océan numérique. C'est ce qu'on appelle le Big Data. Ce bruit numérique, que nous produisons spontanément ou inconsciemment, peut nous aider à mieux appréhender le monde. Comment ? En permettant aux ordinateurs de se doter d'une certaine « intelligence » par « l'apprentissage » à partir de ces données, avec des techniques qui sont au croisement des mathématiques appliquées et de l'informatique. On les regroupe sous les termes d'« apprentissage statistique » ou, en anglais, de *machine learning*. Ces techniques permettent de traiter d'immenses quantités de données (des images, des sons, des statistiques…) grâce à des algorithmes capables d'en extraire les informations utiles, de leur donner du sens et donc de les mettre au service de la société dans des projets innovants, aussi bien dans la santé que dans l'éducation, le commerce ou l'environnement.

« Sans cette culture partagée de l'excellence scientifique, le partenariat n'existerait pas »

Laurent Massoulié est le directeur du Centre de recherche commun. Il nous ouvre les portes d'un laboratoire unique en son genre et nous plonge au cœur de ses problématiques. Entretien.

En quoi le Centre de recherche commun Inria-Microsoft Research se distingue-t-il des autres laboratoires en sciences informatiques ? Qu'est-ce qui fait sa spécificité ?

LAURENT MASSOULIÉ : Je dirais qu'il y en a essentiellement deux. Difficile, tout d'abord, de trouver un exemple de partenariat bilatéral public-privé dans le domaine de la recherche en informatique qui ait cette envergure et qui présente une collaboration aussi poussée entre les chercheurs des deux institutions.

L'autre spécificité essentielle du Centre tient à l'importance donnée à la recherche fondamentale et au partage des résultats : ceux-ci sont mis à la disposition de tous, sous la forme de logiciels en open source et de publications scientifiques.

> « Nous nous efforçons de faire de la recherche de très haut niveau pour jouer un rôle moteur au sein de la communauté scientifique. »
>
> Laurent Massoulié

Ce partenariat public-privé a déjà sept ans d'existence. Qu'en avez-vous appris ?

L. M. : La principale leçon est que cette formule fonctionne ! Le pari que nous avions fait lors de la création du Centre était que la collaboration entre les chercheurs d'Inria et ceux de Microsoft serait mutuellement bénéfique, et c'est effectivement ce que nous avons observé. Les chercheurs d'Inria bénéficient d'une très grande souplesse dans le choix de leurs objectifs scientifiques, et cette ouverture est un plus pour leurs collègues de Microsoft Research. Ces derniers, de leur côté, sont au plus proche contact des défis industriels, et peuvent les partager avec leurs confrères d'Inria.

Les chercheurs d'Inria, intégrés dans le tissu de l'enseignement supérieur, ont également un accès privilégié aux étudiants de troisième cycle. Au-delà des résultats scientifiques, cette collaboration permet donc la formation de doctorants bénéficiant d'une double exposition : à la fois au monde de la recherche académique et à celui de la recherche dans l'industrie.

Existe-t-il une culture et une philosophie de recherche propres au laboratoire commun ?

L. M. : C'est en effet l'une des raisons fondamentales du succès de cette collaboration. Les chercheurs des deux instituts partagent une même philosophie, que l'on pourrait résumer ainsi : repousser les limites de leur domaine scientifique pour avoir un impact sur la société. Chacun, à MSR comme à Inria, se fixe des objectifs ambitieux et

© BERNARD LACHAUD

LAURENT MASSOULIÉ, directeur du Centre de recherche commun.

recherche l'excellence, perçue comme la clé pour avoir un véritable impact. D'ailleurs, si cette culture partagée n'existait pas, le partenariat ne fonctionnerait pas.

Le numérique transforme le monde en profondeur. Quels sont, selon vous, les principaux défis à relever dans les années à venir, en particulier dans le domaine de la recherche en sciences informatiques ?

L. M. : Nous ne sommes encore qu'au début du phénomène du « déluge de données ». Avec l'émergence de nouveaux types de capteurs et d'objets connectés, nous allons mesurer et enregistrer des informations de plus en plus variées, allant de nos caractéristiques physiologiques aux données du trafic routier, de la qualité de l'air à nos conversations.
Ce phénomène va rendre possible un grand nombre de nouvelles applications, et il est aujourd'hui difficile de prédire celles qui auront le plus d'impact sur notre manière de vivre et d'interagir. Mais il revient à la science informatique de proposer les architectures de réseaux, de centres de traitement de données ou d'informatique dans les nuages sur lesquelles se déversera ce flux de données. L'informatique va aussi développer les algorithmes d'apprentissage automatique qui transformeront les flux de données brutes en informations utilisables. Mais le bénéfice apporté par la récolte et le traitement de toutes ces données a une contrepartie : le processus pose de nouvelles questions essentielles de fiabilité, de confiance et de protection de la vie privée. La science informatique fournira aussi les solutions pour agir contre ces menaces.
Les projets du Centre de recherche commun cherchent à répondre à ces défis d'une grande complexité. Leurs objectifs vont de l'informatique fondamentale (méthodes formelles pour la vérification, méthodes et algorithmes pour l'apprentissage statistique) à ses applications-clés (compréhension du fonctionnement du cerveau humain, santé, recherche de vidéos, réseaux sociaux, cloud, protection de la vie privée). Nous ne couvrons donc pas tous les principaux défis des sciences informatiques (et notamment pas ceux de la robotique), mais nous avons déjà un beau programme !

> « Nous ne sommes encore qu'au début du phénomène du déluge de données. »
> Laurent Massoulié

Justement, comment ont été choisis les nouveaux axes de recherche pour les prochaines années ?

L. M. : Nous avons commencé par définir les grands thèmes auxquels seraient attachés les projets, en mettant d'emblée l'accent sur les méthodes formelles – un domaine dans lequel Inria et Microsoft avaient déjà collaboré et où ils avaient tous deux déjà bien avancé –, mais également sur le *machine learning*, la sécurité et le respect de la vie privée, afin de couvrir largement le spectre des questions soulevées par ce déluge de données que j'évoquais à l'instant.
Cette démarche a été complétée par une approche de terrain, où nous avons essayé d'identifier, à l'intérieur de chacun de ces grands thèmes, les tandems de chercheurs des deux institutions qui souhaitaient travailler ensemble sur des objectifs précis.

Qu'attendez-vous de ces projets ? Et au-delà, comment est-il possible d'étendre l'impact du Centre de recherche commun sur son écosystème et sur la société ?

L. M. : Avant tout, nous nous efforçons de faire de la recherche de très haut niveau pour jouer un rôle moteur au sein de la communauté scientifique. Notre conviction est que les fruits de cet effort, qu'il s'agisse de publications scientifiques ou de prototypes de logiciels en open source, induiront de l'innovation qui bénéficiera largement à la société.
Cela peut prendre une forme très évidente, lorsqu'un entrepreneur exploite les résultats d'un projet pour développer une start-up, ou quand des médecins interagissent avec nos chercheurs pour améliorer leurs outils de diagnostic des maladies cardiaques.
Cela peut aussi se faire de manière moins directe : quand nos chercheurs inventent de nouvelles manières d'indexer et de chercher dans des vidéos, ils réduisent le temps d'émergence de nouvelles fonctionnalités dans les moteurs de recherche, ce qui augmente en conséquence la productivité de tout un chacun.
Enfin, l'innovation naît souvent d'opportunités imprévues : notre travail fondamental sur les méthodes formelles a ainsi ouvert la voie à nos chercheurs pour mieux sécuriser les protocoles Internet comme HTTPS. Un développement que nous n'avions pas anticipé au démarrage du projet !

> « L'innovation naît souvent d'opportunités imprévues. »
> Laurent Massoulié

INRIA

À l'interface des sciences informatiques et des mathématiques, les chercheurs d'Inria établissent depuis quarante ans les bases scientifiques d'un nouveau champ de connaissances : les sciences du numérique. En interaction avec les autres disciplines scientifiques, les sciences du numérique proposent de nouveaux concepts, langages, méthodes et objets d'enseignement, qui ouvrent des perspectives inédites dans l'appréhension des phénomènes complexes. Réunis en équipes-projet, les chercheurs d'Inria croisent avec créativité recherche fondamentale et recherche appliquée. Les 174 équipes, pour la plupart communes avec de grandes institutions de recherche françaises ou internationales, rassemblent autour d'un projet une vingtaine de chercheurs sur une durée de quatre à huit ans. Inria est, en France, l'institut public de recherche entièrement dédié aux sciences du numérique. Il accueille chaque année plus de 1000 jeunes chercheurs.

www.inria.fr

MICROSOFT RESEARCH

Depuis sa création, en 1991, Microsoft Research est devenu l'une des premières organisations de recherche en logiciels au monde. Aujourd'hui, ce sont plus de 1 100 chercheurs et ingénieurs dans 13 laboratoires répartis sur quatre continents, et plus de 50 domaines de recherche. Leur mission : repousser les frontières de la recherche en informatique et inventer le futur du numérique.
Microsoft Research Cambridge, au Royaume-Uni, est actuellement l'un des plus importants centres de recherche en sciences informatiques d'Europe. En France, Microsoft Research déploie également l'initiative « Windows Azure for Research », dont la finalité est d'aider la communauté de recherche à utiliser le cloud computing comme instrument au service du progrès scientifique. Outre Inria, cette initiative bénéficie à des institutions comme l'École centrale de Paris, l'université de Grenoble, l'université de Nice ou l'université Paris-Descartes.

www.research.microsoft.com

© B. LACHAUD

UN CENTRE DE RECHERCHE **DE RANG MONDIAL**

C'est à une vingtaine de kilomètres au sud de Paris, au cœur de la future université de Paris-Saclay, que se trouve **le Centre de recherche commun**. Inauguré en janvier 2007 par Inria et Microsoft, il **accueille des chercheurs du monde entier** qui y mènent conjointement des recherches à long terme dans quatre grands domaines :

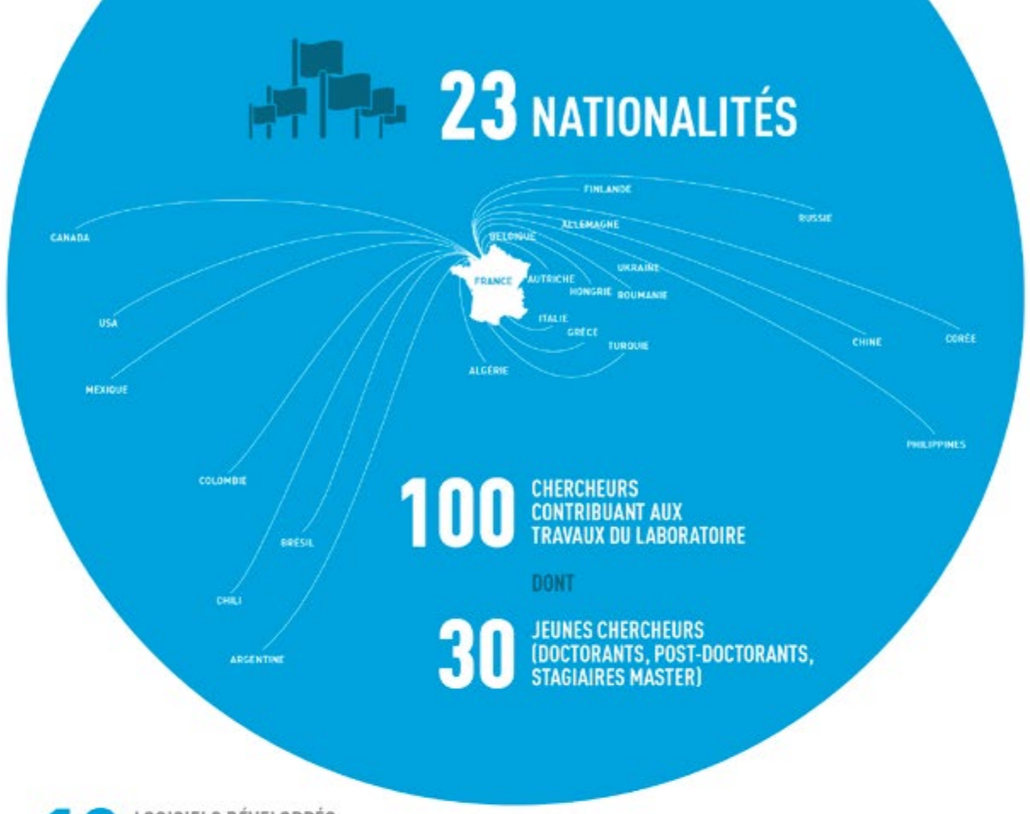
VISION NUMÉRIQUE ET IMAGERIE MÉDICALE

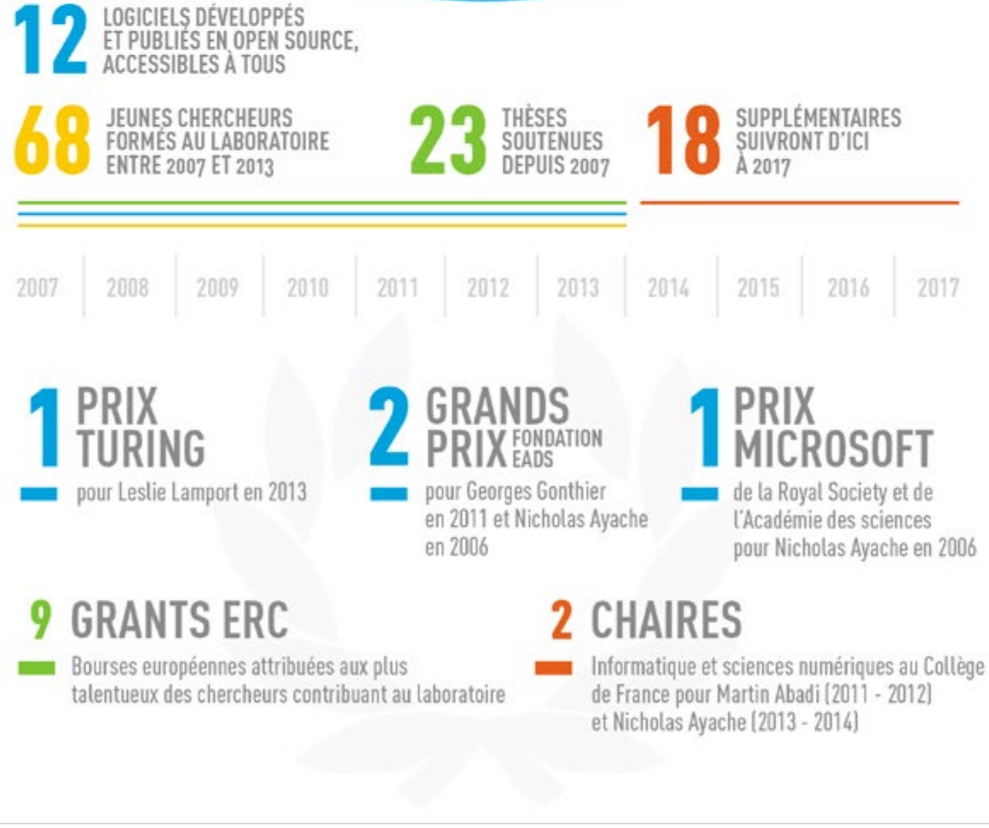
MÉTHODES FORMELLES POUR LES MATHÉMATIQUES, SYSTÈMES DISTRIBUÉS ET SÉCURITÉ

MACHINE LEARNING ET BIG DATA

RÉSEAUX SOCIAUX D'INFORMATION ET PROTECTION DE LA VIE PRIVÉE

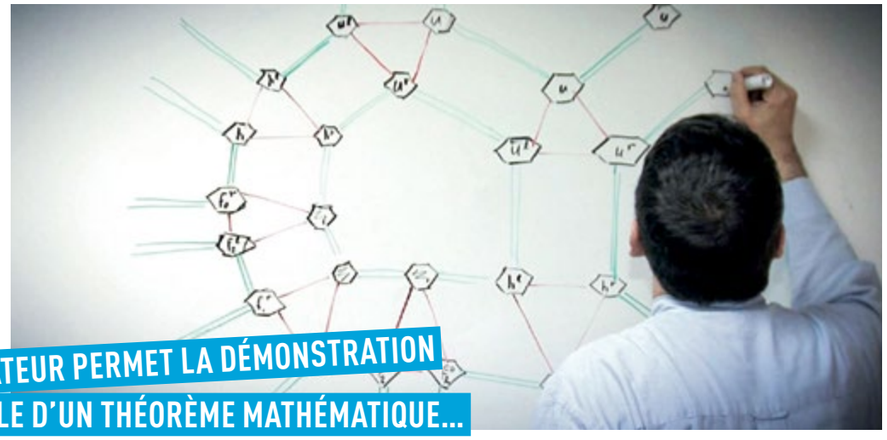
Les résultats des travaux du Centre de recherche commun Inria-Microsoft Research sont publics et mis à la disposition de la communauté scientifique nationale et internationale.





UN IMPACT SCIENTIFIQUE MAJEUR
AU NIVEAU MONDIAL

« This is really the End »[1]. Le 20 septembre 2012, à 17 h 46, Georges Gonthier signalait d'un mail laconique l'aboutissement de six ans de travail acharné : la preuve informatique de l'emblématique théorème de Feit-Thompson. Une prouesse saluée au niveau mondial, et qui est loin d'être unique : partez à la découverte des projets les plus marquants du laboratoire.



LORSQU'UN ORDINATEUR PERMET LA DÉMONSTRATION IRRÉFUTABLE D'UN THÉORÈME MATHÉMATIQUE…

170 000 lignes de démonstration, 1 h 40 de compilation, et une certitude, formelle : la démonstration faite en 1963 est correcte et complète.

Georges Gonthier, de Microsoft Research, a un grand projet : apprendre aux ordinateurs à « faire des maths ». Plus précisément, construire des briques logicielles capables de démontrer de manière irréfutable des théorèmes mathématiques complexes. L'objectif du projet est simple : fournir aux mathématiciens des outils informatiques pour les assister efficacement, aussi bien dans la vérification de leurs preuves que dans la structuration de leur savoir, et leur apprendre à collaborer autrement pour mieux travailler ensemble.

Quoi de mieux, pour prouver l'efficacité du procédé, que de s'attaquer aux 250 pages (!) de l'un des théorèmes phares de l'algèbre du xx[e] siècle, celui de Feit-Thompson ?

Fin septembre 2012, mission accomplie pour Georges Gonthier et ses collègues d'Inria, qui ont réussi à vérifier par ordinateur la preuve du théorème. Six années auront été nécessaires pour faire comprendre la « démonstration » à l'ordinateur, et donc pour lui apprendre, en amont, une partie substantielle des mathématiques modernes. Les objets mathématiques, les théorèmes et les démonstrations bien sûr, mais aussi et surtout une compétence essentielle et implicite : les manières de s'en servir. Un succès majeur qui, au-delà de la prouesse technique, présage une façon radicalement nouvelle de faire des mathématiques.

1. « C'est vraiment la fin »

FEIT-THOMSON

SÉCURITÉ

… ET VÉRIFIE INDISCUTABLEMENT LA FIABILITÉ D'UN LOGICIEL

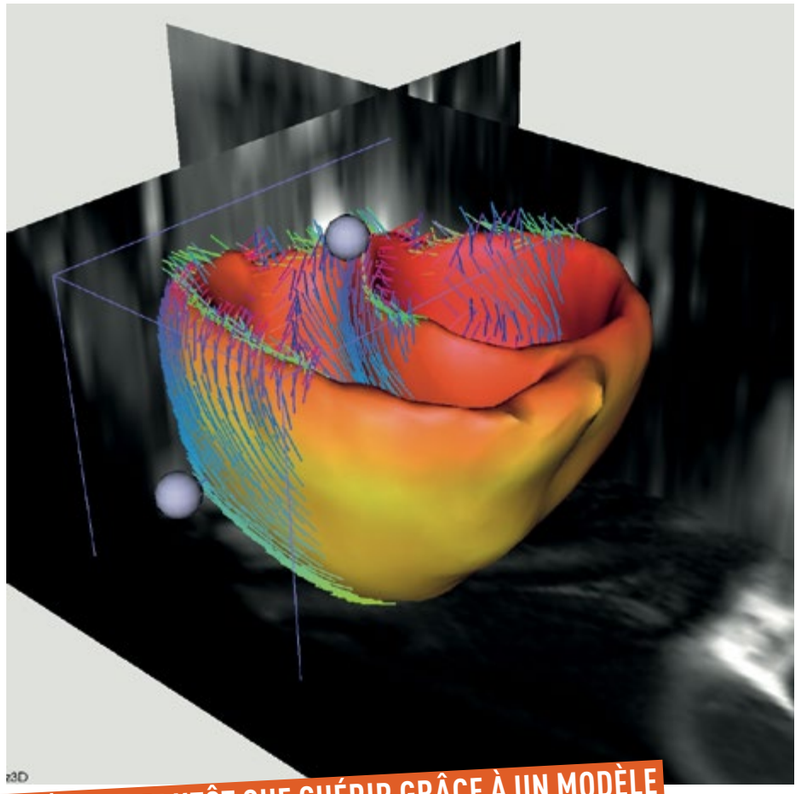
Les méthodes formelles permettent d'améliorer la qualité et la fiabilité des applications tout en gagnant un temps précieux dans leur conception et leur développement.

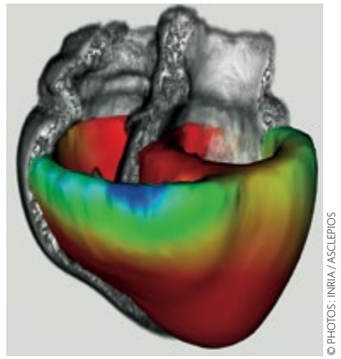
Cédric Fournet, de Microsoft Research, et son équipe travaillent sur un projet fondé lui aussi sur les méthodes formelles, mais avec des applications dans le domaine de la sécurité informatique. L'objectif concret ? **Limiter l'apparition de failles de sécurité et construire une informatique de confiance capable de garantir la sécurité des transactions d'ordinateur à ordinateur à travers Internet.**

L'équipe s'est focalisée sur le code mettant en œuvre TLS, le protocole de sécurité présent derrière l'universellement connu « https:// », qu'un internaute utilise pour se connecter à un site sécurisé, comme celui de sa banque.

Grâce à leurs outils de vérification, les chercheurs ont pu certifier la conformité d'un code appliquant le protocole TLS. Au passage, ils ont découvert, avec ces mêmes outils, quelques failles dans des implémentations de TLS couramment déployées, et des attaques exploitant judicieusement ces failles pour compromettre un site soi-disant sécurisé !

©INRIA / KAKSONEN





© PHOTOS : INRIA / ASCLEPIOS

CI-DESSUS
Visualisation des deux ventricules cardiaques et de l'architecture des fibres cardiaques à partir d'images médicales.

À DROITE
Modélisation 3D à partir d'une image médicale d'un cœur.

PRÉVENIR PLUTÔT QUE GUÉRIR GRÂCE À UN MODÈLE NUMÉRIQUE ET PERSONNALISÉ DU CŒUR HUMAIN

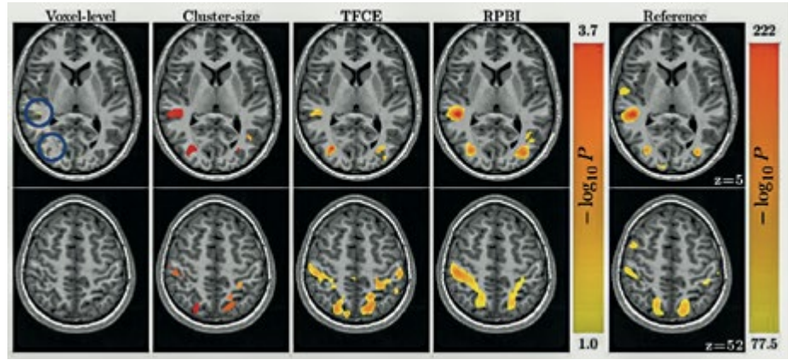
Dans leur projet commun, les chercheurs Nicholas Ayache, d'Inria, et Antonio Criminisi, de Microsoft Research, ainsi que leurs doctorants, construisent un modèle numérique du cœur humain capable d'intégrer toutes les caractéristiques de celui d'un véritable patient. Pour cela, ils ont travaillé avec une large base de données d'images médicales cardiaques, commentées et classées en fonction du diagnostic réalisé par un expert.

Le but ? Réussir à sélectionner automatiquement des cas similaires déjà traités pour anticiper une éventuelle évolution pathologique, tenter de la prévenir, ou, le cas échéant, choisir le bon traitement et identifier le moment le plus propice à une opération.

En parallèle, l'équipe Inria de Nicholas Ayache, Asclepios, est déjà introduite dans certains hôpitaux, travaillant en partenariat avec plusieurs praticiens qui comparent au jour le jour les mesures faites sur les patients pendant la consultation avec les diagnostics prédictifs des outils mis au point dans le laboratoire.

EXPLORER LE FONCTIONNEMENT DU CERVEAU GRÂCE AU CLOUD

À l'origine du projet « A-Brain », un constat : nos cerveaux sont tous un peu différents les uns des autres, et cette variabilité complique le diagnostic de certaines maladies du système nerveux.
Les chercheurs ont eu l'idée de corréler les vues du cerveau et le patrimoine génétique de chaque patient, pour voir si on ne pouvait pas relier certains gènes à des fonctionnements différents du cerveau.
Pour cela, ils ont utilisé la plateforme d'imagerie de Neurospin, un outil d'IRM surpuissant, afin d'acquérir les images du cerveau, puis ils ont développé des algorithmes spécifiques pour les analyser. Ils ont ensuite croisé les informations recueillies avec une base de données génétiques, ce qui leur a permis d'obtenir des analyses fines des mécanismes et des comportements du cerveau, ainsi que de ses pathologies. Il faut savoir que le croisement statistique des informations visuelles et génétiques nécessite un volume de données de l'ordre du péta-octet pour chaque patient, et un temps de calcul se chiffrant en années pour un calculateur conventionnel. Bertrand Thirion et son équipe ont donc utilisé la puissance de calcul et de stockage permise par la plateforme de Cloud Computing Windows Azure pour conduire ces expériences à grande échelle, en collaboration étroite avec l'équipe rennaise KerData du centre de recherche Inria Rennes–Bretagne Atlantique et avec l'Advanced Technology Lab Europe de Microsoft Research, situé à Munich. Un défi remarquable par l'échelle du problème.



© ICONEM

FAIRE RENAÎTRE LES DÉCORS ÉTINCELANTS D'UNE VILLA DE POMPÉI

Produire un outil logiciel de « restauration numérique » du bâti archéologique à partir de simples photos numériques : c'est le projet de Jean Ponce, professeur à l'École normale supérieure de Paris (ENS) et directeur du projet Inria « Willow », qui a rassemblé autour de lui une équipe pluridisciplinaire issue de l'ENS, du CNRS et du laboratoire commun Microsoft-Inria.
C'est en collaboration étroite avec Hélène Dessales, maître de conférences à l'ENS et spécialiste de l'architecture romaine, que Jean Ponce et son équipe ont reconstitué **deux vastes villas de Pompéi – plusieurs milliers de mètres carrés pour la villa de Diomède – en extrapolant le décor d'origine de ces architectures anciennes de plus de deux mille ans** à partir de données très hétéroclites.

Des photos bien sûr, mais aussi une très riche iconographie comprenant des aquarelles et des croquis des XVIIIe et XIXe siècles, des peintures, des relevés d'architectes et même des objets empruntés aux musées archéologiques.
Cette véritable résurrection virtuelle ouvre déjà de nouvelles perspectives en matière d'archéologie des sites en danger et, plus généralement, d'urbanisme. L'outil développé permet, par exemple, de générer automatiquement un modèle 3D d'un bâtiment ou d'un quartier à partir de photos.
Un service très utile pour les géomètres et les architectes, qui pourrait à terme remplacer définitivement les laborieux relevés de métré et de report dans les logiciels de 3D.

Un outil logiciel de restauration numérique qui ouvre de nouvelles perspectives en matière d'urbanisme et d'archéologie.

© PHOTOS : INRIA/KAKSONEN - NETWORK PAR SIMON COCKELL (FLICKR) LICENCE CC

ARCHÉOLOGIE

QUATRE GRANDS AXES DE RECHERCHE

Dans ses premières années, le Centre de recherche commun s'est construit naturellement autour d'une dynamique de collaboration préexistante et d'un intérêt partagé par les équipes des deux institutions pour le thème central de la logique mathématique, les méthodes formelles et la conception des langages de programmation.

En 2013, une dynamique similaire se développe autour de l'apprentissage statistique et du Big Data, des défis théoriques et techniques qu'ils recèlent, et de certaines de leurs applications très concrètes au monde numérique qui nous entoure.

APPRENTISSAGE STATISTIQUE ET BIG DATA

Comment les machines doivent-elles apprendre ? Jusque dans les années 80, les chercheurs en intelligence artificielle ont tenté de rendre les machines « intelligentes » en modélisant l'intelligence humaine sous la forme de règles dites « de décision », à partir desquelles les machines étaient censées construire leur propre raisonnement.

RENDRE LES MACHINES INTELLIGENTES

Avec le *machine learning*, « apprentissage statistique » en français, l'idée est que les logiciels apprennent par l'expérience. En ingérant de très grands volumes de données, la machine va les analyser pour en extraire l'information structurelle qui lui permettra de reconnaître des situations ultérieures et de réagir en conséquence. Cette forme d'apprentissage repose sur l'utilisation de puissants algorithmes d'analyse de données combinant des outils de statistique et d'optimisation qui nécessitent eux-mêmes des ressources de calcul importantes.

Un exemple ? En faisant analyser à un ordinateur les symptômes d'un grand nombre de patients atteints de la même maladie, on peut le rendre capable de déceler cette même maladie avec une marge d'erreur connue. Plus près de nous, nous utilisons déjà l'apprentissage statistique au quotidien : quand nous cherchons sur Internet, quand nous utilisons un service de traduction automatique, quand notre messagerie filtre un spam… l'ordinateur apprend de nos actions et de nos données pour nous aider à mieux appréhender le monde.

Qu'il s'agisse de santé, d'éducation, de commerce ou encore d'environnement, l'apprentissage statistique fondé sur le Big Data a déjà rendu possible une multitude de nouveaux services. Mais ce n'est qu'une infime partie de ce qui nous attend !

> « Nous partons toujours des données de problèmes concrets pour ensuite naviguer des applications à la théorie et à l'algorithmie. Avec en tête un objectif essentiel : trouver des solutions ayant un temps de calcul raisonnable. »
>
> Francis Bach

© INRIA / J.M. RAMÈS

Polytechnicien, ingénieur du corps des Mines et titulaire d'un doctorat en informatique de l'université de Berkeley sous la direction du professeur Michael Jordan, **Francis Bach** est un spécialiste mondialement connu de l'apprentissage statistique. Ses recherches, à l'interface des mathématiques appliquées, des statistiques et de l'informatique, trouvent des applications dans de nombreux domaines, de la vision artificielle à l'imagerie cérébrale, en passant par le traitement du signal audio et la bioinformatique.

AUTOMATISER L'INGÉNIERIE LOGICIELLE

Les méthodes d'optimisation mathématique sont devenues essentielles dans de nombreux domaines des sciences et techniques et même de la vie quotidienne (la circulation des trains, la gestion des centrales électriques, par exemple). Simultanément, si la révolution industrielle au XIXe siècle a permis l'éclosion du métier d'ingénieur, la révolution numérique impose l'idée de « génie logiciel » ou d'« ingénierie logicielle ». Ce projet vise à introduire les techniques d'optimisation dans l'ingénierie logicielle. L'ambition ? Assurer à moindre coût la production de logiciels efficaces de la meilleure qualité possible, avec le minimum d'erreurs et de bugs, notamment en automatisant la création de jeux de test de bon comportement du logiciel. Cet objectif s'inscrit dans la continuité d'un projet précédent, qui s'attachait à augmenter automatiquement la performance des algorithmes d'optimisation grâce à des méthodes de paramétrage automatique.

APPRENTISSAGE STATISTIQUE ET DONNÉES STRUCTURÉES À GRANDE ÉCHELLE

Le Big Data et le niveau d'échelle qu'il induit mettent au défi les techniques actuelles d'apprentissage statistique. Pour obtenir un « modèle » numérique d'un problème, il faut en effet traiter des millions de données, depuis des vidéos jusqu'à des séquences d'ADN, sachant que, pour chaque donnée examinée, on cherche à déterminer des milliers d'attributs, comme la liste des objets visibles dans une vidéo, ou des prédispositions pathologiques apparaissant dans le génome humain. En conséquence, si l'on veut atteindre un modèle statistique de qualité suffisante, les temps de calcul peuvent devenir rédhibitoires.

L'ambition de ce projet est de trouver de nouveaux algorithmes, ou de nouvelles familles d'algorithmes, permettant de rester efficace dans un contexte de Big Data. En d'autres termes, il s'agit d'obtenir un compromis optimal entre la qualité de l'apprentissage et le temps de calcul. Il s'agit aussi de développer des algorithmes visant à un compromis optimal entre les objectifs contradictoires d'« exploration » (par exemple, proposer à un patient des traitements différents afin d'explorer leur efficacité) et d'« exploitation » (soigner au mieux ce même patient).

MACHINE LEARNING

TECHNIQUES DE STOCKAGE ET DE TRAITEMENT AVANCÉS DES DONNÉES DANS LES NUAGES

Plus de science aujourd'hui qui ne soit à un degré ou un autre numérique – modélisation, simulation, etc. Se pose donc la question de fournir aux scientifiques des outils modulaires de gestion de leurs « instruments de laboratoire » numériques. Les scientifiques ont besoin d'exprimer aisément les enchaînements de leurs tâches de calcul et d'en compiler les résultats. **Ces processus sont complexes, génèrent souvent des masses de données, les calculs peuvent durer des jours et sur des ordinateurs géographiquement distants.**
L'orchestration de ces « appareils numériques de laboratoires » réside dans ce que l'on appelle les « workflows scientifiques ».
Dans un contexte d'expérimentation à grande échelle et de données massives, il est critique de traiter les étapes de calcul en parallèle et d'inventer de nouvelles approches de gestion des données « dans les nuages » qui soient d'un même niveau d'efficacité et d'une manipulation aisée pour les scientifiques. Ce projet se propose donc de se focaliser sur la gestion des masses de données dans un tel cadre : Comment adapter les modèles de workflows actuels à un environnement cloud ? Comment optimiser le transfert de données dans un délai raisonnable ? Comment gérer les données et les tâches de façon à ce qu'elles soient toujours accessibles ou exécutables au moment où on en a besoin ?

VOUS AVEZ DIT BIG DATA ?

Un milliard de milliards d'octets (1 exaoctet), soit un milliard de gigaoctets : c'est l'unité qui permet aujourd'hui de mesurer la quantité de données produite du fait de la généralisation des outils du numérique. Les sciences génèrent de grands volumes de données, que ce soit dans le domaine du décryptage de l'ADN, de la climatologie ou de la mécanique des fluides… Pour les gérer, il faut non plus quelques dizaines de serveurs, mais des milliers.
La capacité à traiter cette explosion des données, c'est ce qu'on appelle le Big Data.
Une notion qui a des implications dans des domaines très variés et qui présente un enjeu de taille : passer de la profusion de données à la connaissance. Outre l'impératif de puissance des calculateurs, le traitement des données en version Big Data implique la création de nouveaux outils d'analyse. Aujourd'hui, le défi est de créer des modèles et des algorithmes capables de traiter les Big Data pour les transformer en information et en connaissance, et ce à un coût de calcul acceptable.

© INRIA / KAKSONEN

VISION NUMÉRIQUE ET IMAGERIE MÉDICALE

Mieux naviguer dans une vidéo en la découpant en de multiples séquences automatiquement, savoir « lire » une image scientifique ou un cliché médical pour repérer à temps une anomalie susceptible de se transformer en maladie grave, modéliser un objet, un lieu ou un organe, reconstruire virtuellement en 3D, en partie ou en totalité, un lieu historique disparu…
Tout cela est aujourd'hui possible, et bien plus encore, grâce à l'analyse et à la synthèse numériques des images fixes ou animées.
Mais il s'agit d'un domaine dans lequel les défis sont immenses et où la recherche est extrêmement active. C'est un champ d'application particulièrement fécond en techniques avant-gardistes d'apprentissage statistique. Le Centre de recherche commun se concentre ici sur deux aspects principaux : d'un côté, l'analyse et la compréhension de contenus vidéo, de l'autre, l'analyse et la synthèse d'images médicales.

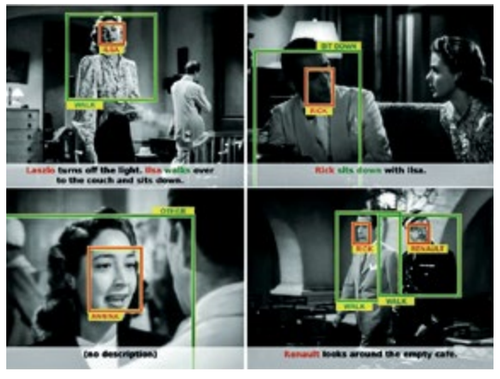
© COLLÈGE DE FRANCE / PATRICK IMBERT

Ingénieur civil des Mines, titulaire d'un Master de l'université de Californie, Los Angeles, ainsi que d'un doctorat et d'une thèse d'état de l'université Paris-Sud, **Nicholas Ayache** rejoint Inria en 1981. Il y crée l'équipe Épidaure au centre de Paris-Rocquencourt, puis Asclepios, au centre de Sophia Antipolis-Méditerranée. Il est directeur de recherche de classe exceptionnelle depuis 2006 et directeur scientifique de l'institut hospitalo-universitaire de Strasbourg depuis 2012. Élu professeur au Collège de France, titulaire de la chaire Informatique et sciences numériques pour l'année académique 2013-2014, il a été récompensé par le prix de la fondation EADS en 2006 et il a également cofondé la conférence internationale MICCAI.

> « Les images médicales d'aujourd'hui recèlent une telle quantité d'informations à la fois anatomiques et fonctionnelles qu'elles ne peuvent plus être exploitées sans l'aide d'algorithmes spécifiques. La production des images n'est donc plus une fin en soi mais le début d'un processus d'analyse qui repose sur des logiciels complexes. »
>
> Nicholas Ayache

COMPRENDRE, EXPLORER ET ORGANISER LES CONTENUS VIDÉO DYNAMIQUES

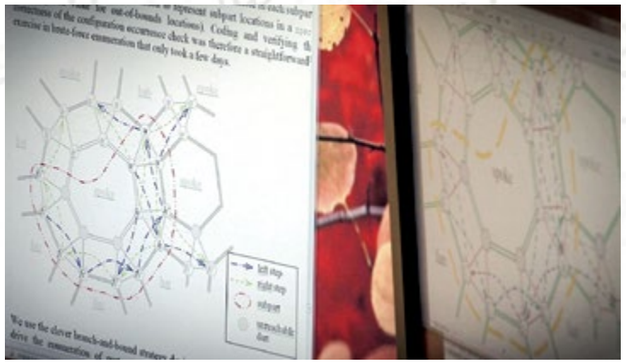
Comprendre et interpréter automatiquement un contenu vidéo sont devenus aujourd'hui un enjeu essentiel. Une fois ce verrou levé, la machine nous permettra de trouver un passage précis dans un film, de sélectionner des contenus plus pertinents, ou encore de détecter et de diagnostiquer une situation anormale. **L'ambition de ce projet est de rendre la machine capable d'une analyse fine et détaillée de chaque vidéo,** déterminant les objets, les personnes, les actions qui constituent une scène, ainsi que le lieu où elle se déroule.
Une fois les films enrichis de toutes ces « métadonnées structurées », il devient possible de faire des recherches précises dans un grand corpus de vidéos, ce qui ouvre la voie à toute une variété d'applications innovantes.
De nouvelles méthodes d'apprentissage statistique seront développées pour atteindre l'objectif du projet.



© INRIA / KAKSONEN

DIAGNOSTIQUER ET CONTRÔLER LES MALADIES GRÂCE À DES IMAGES DE SYNTHÈSE

En croisant des milliers d'images, un ordinateur peut « apprendre » à détecter automatiquement une anomalie médicale chez un nouveau patient. Mais, pour qu'il y arrive, il faut que les images de référence ayant servi à « l'entraîner » soient correctement « annotées », c'est-à-dire qu'un expert – un médecin – ait explicitement désigné sur l'image IRM l'anomalie (une tumeur, par exemple) et les régions associées. Il s'agit alors d'un problème de *machine learning*, et plus précisément de « classification automatique ». Il n'est, hélas, pas envisageable de solliciter les experts pour annoter l'énorme corpus d'images nécessaire au traitement de toute la diversité des cas possibles. Pour s'affranchir de cet obstacle, il a été récemment démontré que, sous certaines conditions, on peut automatiser la génération d'images synthétiques annotées, c'est-à-dire des images médicalement réalistes mais n'appartenant à aucun patient réel. C'est l'approche employée, par exemple, par Microsoft et Xbox avec le « moteur » du capteur Kinect : l'algorithme permettant de séparer les parties d'un corps humain a été « entraîné » avec des millions d'images de corps humains de toute corpulence et de toute taille, générées automatiquement et correctement annotées. L'ambition de ce projet est donc de relever ce défi dans le contexte d'images médicales comportant des anomalies tumorales (IRM et tomographie), en particulier du cœur et du cerveau. Les applications potentielles sont considérables : du diagnostic au traitement spécifique, en passant par la prévention et la quantification des effets d'un médicament, pour faire un pas de plus vers une médecine personnalisée et préventive.



MÉTHODES FORMELLES

Les méthodes formelles sont des techniques fondées sur la logique mathématique qui permettent de raisonner rigoureusement sur des programmes informatiques ou des matériels électroniques. Il s'agit d'un champ de recherche très actif depuis de nombreuses années, tout particulièrement au sein d'Inria, qui a développé le « langage de preuves » Coq. Elles visent à garantir irréfutablement l'absence de bugs ou de défauts, et donc à rendre l'informatique bien plus fiable : il ne s'agit plus de repérer le défaut de conception d'un logiciel par les erreurs qu'il peut produire, mais de vérifier en amont la logique de sa construction.
De nombreux domaines bénéficient de ces techniques : de l'automobile à l'aviation, en passant par les équipements médicaux, elles permettent d'assurer une plus grande fiabilité et une plus grande sécurité tout en réduisant grandement les coûts de développement et de test.
Le Centre de recherche commun travaille depuis plusieurs années déjà à une génération révolutionnaire d'outils de vérification basés sur les preuves formelles : il élabore de nouveaux langages et environnements de vérification capables de traiter des énoncés de mathématiques extrêmement complexes. Il exploite ces développements pour la vérification de mécanismes de sécurité, et en particulier de cryptographie, avec comme enjeu la sécurisation des transactions sur Internet. Enfin, il développe des méthodes pour certifier les propriétés des protocoles dans les systèmes distribués, une tâche particulièrement complexe.

© INRIA / KAKSONEN

Récompensé par le prix de la fondation EADS décerné par l'Académie des sciences en 2011, **Georges Gonthier** est l'un des informaticiens les plus brillants de sa génération : après un début de carrière aux Bell Laboratories et à Inria, il devient chercheur à Microsoft Research Cambridge et au Centre de recherche commun. Ses recherches portent notamment sur l'utilisation des assistants de preuve dans des domaines étendus des mathématiques, avec, à son actif, deux succès majeurs : la formalisation de la preuve du théorème des quatre couleurs en 2005, et celle du théorème de Feit-Thompson en 2012.

> « Les méthodes formelles visent à garantir **irréfutablement l'absence de bugs ou de défauts,** et donc à rendre l'informatique bien plus fiable. »

VÉRIFIER LES MATHÉMATIQUES PAR ORDINATEUR

L'équipe de Georges Gonthier a développé un langage informatique ainsi que toute une bibliothèque de briques logicielles qui permettent de représenter un riche ensemble de propositions mathématiques, et, mieux encore, d'en faire vérifier automatiquement la véracité par ordinateur. Le point d'orgue de cet effort de six ans a été la vérification par ordinateur du théorème de Feit-Thompson, résultat phare des mathématiques modernes, en septembre 2012 (voir page 10). En principe, l'effort accompli permet aux mathématiciens de faire vérifier par ordinateur les preuves de leurs théorèmes, au moins dans le sous-domaine des mathématiques que constitue l'algèbre. Il présage ainsi une nouvelle façon de faire des maths à l'aide de l'ordinateur. L'équipe a aujourd'hui pour ambition de développer et d'optimiser les outils existants pour les rendre plus faciles d'accès aux mathématiciens et, ainsi, en favoriser l'adoption. Elle va aussi s'attacher à développer un nouvel ensemble de briques logicielles pour étendre l'applicabilité des méthodes formelles à d'autres branches des mathématiques : après l'algèbre, il s'agit de couvrir l'analyse, sur laquelle repose notamment tout une partie de la physique théorique.

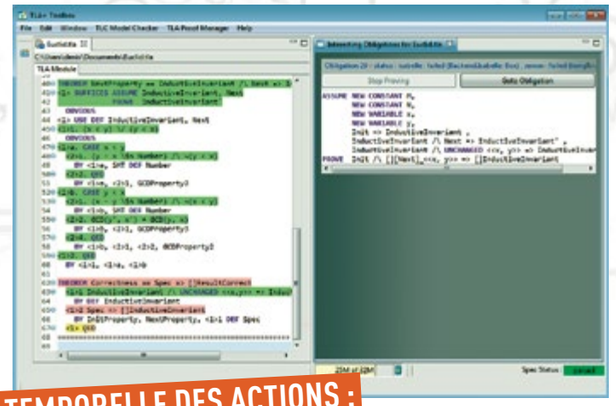
> « Cette extension des composants mathématiques à l'analyse est **absolument cruciale.** Elle pourrait nous permettre de faire la preuve de la conjecture de Poincaré, de vérifier la théorie de la renormalisation et la théorie quantique des champs, ou encore de valider les calculs du LHC, le plus grand accélérateur de particules au monde, qui a permis la découverte du boson de Higgs. »
>
> Georges Gonthier

© PHOTOTHÈQUE CNRS – SÉBASTIEN GODEFROY

LES MÉTHODES FORMELLES, BIENTÔT INDISPENSABLES À NOTRE QUOTIDIEN

CÉDRIC VILLANI, MÉDAILLE FIELDS

« La vérification informatique de théorèmes mathématiques est une manière astucieuse de tester ces méthodes et de démontrer, par la preuve, qu'elles s'appliquent à des cas compliqués, pour des démonstrations longues et complexes. C'est le prélude d'un phénomène qui pourrait à court terme devenir industriel et s'appliquer à tous les systèmes complexes de notre quotidien (transports, médecine, électronique, informatique, téléphonie, consoles de jeu, alarmes, système bancaire, etc.). »



LOGIQUE TEMPORELLE DES ACTIONS : VÉRIFIER LES SYSTÈMES DISTRIBUÉS

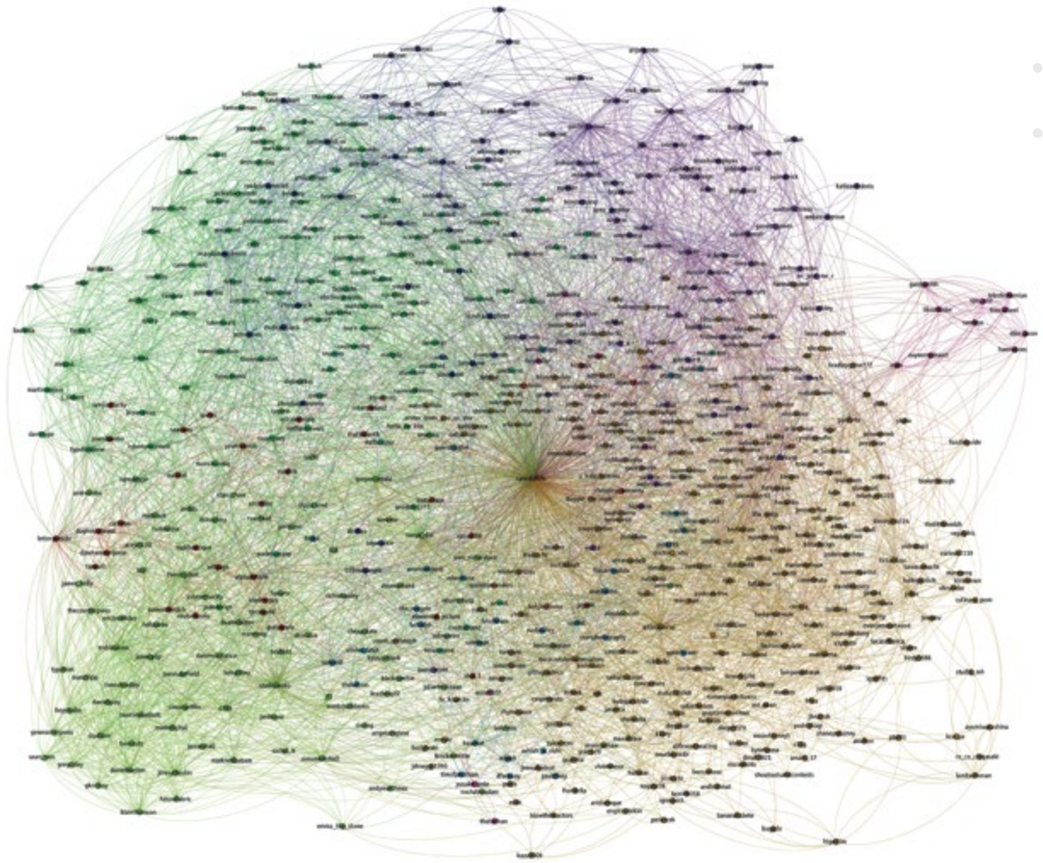
Si l'écriture d'un programme est une activité potentiellement source de bugs, la définition du programme à écrire à partir d'un cahier des charges – la spécification – est également une activité critique. Si un besoin y est décrit de façon partielle ou ambiguë, il y a fort à parier que la solution ne sera pas satisfaisante. Et le problème est encore plus complexe s'il s'agit de logiciels temps réel ou parallélisés sur de nombreuses machines (ce que l'on appelle des « systèmes distribués »). Comme pour la programmation, les méthodes formelles peuvent constituer une approche valide pour garantir la meilleure expression d'une spécification, c'est-à-dire pour faire des « spécifications formelles ».
TLA+ est un langage proposé par Leslie Lamport, de Microsoft Research, et qui est particulièrement bien adapté aux spécifications formelles des systèmes parallèles et distribués. Ces dernières années, les chercheurs du Centre commun ont construit des méthodes et des outils pour faciliter l'écriture et vérifier formellement les spécifications complexes écrites en TLA+. Ils ont illustré le potentiel de leur approche en certifiant les propriétés d'un protocole qui est au cœur de tout système distribué, celui qui est censé garantir que plusieurs ordinateurs distants se « mettent d'accord » sur la valeur de quantités d'intérêt commun (comme lorsque plusieurs agences de voyage gérant des réservations sur les mêmes places d'avion doivent se mettre d'accord sur qui a vendu quelle place). **Aujourd'hui, l'équipe souhaite aller plus loin en développant et en mettant en œuvre des outils capables de prouver des propriétés « temporelles »**, puis en les testant sur des algorithmes complexes. Une première collaboration a d'ores et déjà été initiée avec le CEA pour prouver des propriétés temporelles de son nouveau système d'exploitation temps réel, PharOS.

MATHÉMATIQUES

VÉRIFIER LA SÉCURITÉ DES TRANSACTIONS EN LIGNE

La sécurité de notre utilisation d'Internet présente deux maillons faibles. D'une part, les communications peuvent être compromises si les protocoles de chiffrement censés les protéger présentent des failles. D'autre part, le code téléchargé sur des pages Web pour être exécuté localement sur nos machines peut compromettre notre machine, soit délibérément, soit à cause d'un bug qu'il contient. C'est particulièrement fréquent avec du code écrit en JavaScript, un langage intrinsèquement difficile à débugger. Ce projet s'attache à renforcer ces deux maillons faibles en utilisant les méthodes formelles.

Concernant le premier maillon, Cédric Fournet et son équipe ont développé un logiciel qui génère automatiquement le code d'un protocole cryptographique à partir de ses spécifications, et qui, au passage, certifie leur bon respect. L'équipe va développer cette approche pour rendre possible l'utilisation de fonctions de cryptographie, afin de pouvoir stocker ses données personnelles dans le cloud sous une forme cryptée et de les consulter sans que l'opérateur du cloud puisse y accéder. Concernant le deuxième maillon, l'équipe a mis au point le langage F*, pour lequel il est facile de vérifier les propriétés du code qu'on y écrit, et a développé un compilateur synthétisant automatiquement à partir d'un code écrit en F* un second code écrit en JavaScript. L'équipe va raffiner les outils correspondants et s'employer à populariser cette approche auprès des développeurs Web.



© NETWORK PAR SIMON COCKELL (FLICKR) LICENCE CC

RÉSEAUX SOCIAUX ET CONFIDENTIALITÉ

Les réseaux sociaux proposent une nouvelle manière d'accéder et de traiter collectivement l'information, les utilisateurs produisant, recommandant ou partageant des contenus avec leurs connaissances. Comment améliorer alors la pertinence des contenus proposés en fonction des goûts de chacun ? Et comment favoriser l'émergence de communautés d'intérêt tout en préservant la vie privée et en garantissant la confidentialité des données personnelles des utilisateurs ?
Ce sont là deux questions-clés pour les chercheurs, qui s'emploient à développer des réseaux sociaux et des systèmes proposant une information mieux ciblée à leurs utilisateurs tout en étant plus respectueux de leur vie privée.

FLUX D'INFORMATIONS DANS LES RÉSEAUX SOCIAUX EN LIGNE

Ce nouveau projet, initié par Laurent Massoulié, directeur du Centre commun, s'intéresse plus particulièrement à l'accès à l'information via les réseaux sociaux en ligne, comme Twitter et Facebook. Pour améliorer la précision, la pertinence et la finesse de la recherche d'information sur les réseaux sociaux, **Laurent et son équipe vont concevoir et développer des algorithmes qui croiseront automatiquement plusieurs dimensions, notamment la typologie des contenus, les goûts, les centres d'intérêt et les domaines d'expertise des utilisateurs.** Ces algorithmes permettront la recommandation automatisée aux usagers de contacts potentiels ayant des goûts similaires aux leurs, et ainsi l'émergence d'interconnexions entre utilisateurs qui optimiseront la diffusion d'informations. Ils permettront aussi la recommandation proactive de contenus récents en recherchant un compromis optimal entre catégorisation rapide du type de contenu et limitation du spam. Enfin, l'équipe cherchera à élaborer des mécanismes de réputation des usagers, facilitant l'évaluation de leurs domaines d'expertise et visant à les encourager à faire le meilleur filtrage possible des informations auxquelles ils accèdent.

© B. LACHAUD

Laurent Massoulié est diplômé de l'École polytechnique. Il a débuté comme chercheur à France Télécom R&D avant de rejoindre, en 1999, Microsoft Research Cambridge, où il travaillait sur les algorithmes distribués pour des systèmes pairs-à-pairs. Après un passage par Thomson, il a pris la tête du Centre de recherche commun en 2012.

> « Comment réussir à garantir l'accès à une information à la fois plus pertinente et plus personnalisée alors même que le travail éditorial d'agrégation, de tri et de filtrage de l'information est fait par le réseau ? C'est tout l'enjeu scientifique de notre recherche. »
>
> Laurent Massoulié

© SOOGLS

SERVICES EN LIGNE RESPECTANT LA VIE PRIVÉE

Nos appareils génèrent un flux constant de données potentiellement compromettantes pour notre vie privée. C'est le cas de nos signaux biométriques (ECG), de nos achats sur Internet, ou encore de notre localisation (GPS). Aujourd'hui, nous n'exerçons aucun contrôle sur ces données : notre localisation GPS est, par exemple, connue des opérateurs d'applications de smartphones. Cependant, il peut être souhaitable de divulguer une partie de ces données privées quand elles sont nécessaires à certaines fonctionnalités, comme la mesure fine du trafic dans une zone urbaine. Catuscia Palamidessi et son équipe vont concevoir des modèles de divulgation partielle, dans lesquels les données privées sont transmises soit sous forme cryptée, soit après distorsion délibérée par rajout de « bruit ».
Ces modèles **permettront de réaliser un compromis entre préservation de la vie privée pour les utilisateurs et capacité de fournir un service utile pour les opérateurs exploitant ces données.**
Les chercheurs vont développer les solutions logicielles correspondantes pour le partage de données de géolocalisation depuis les smartphones en intégrant la contrainte du respect de la vie privée.

VISUALISATION INTERACTIVE DES RÉSEAUX SOCIAUX

La représentation des données sous forme de réseaux et d'interactions visuelles est devenue aujourd'hui essentielle dans de nombreux domaines : de la visualisation de l'activité du cerveau aux similitudes entre des gènes, en passant par les interactions sociales entre individus ou organisations.
Elle permet en effet de voir, d'explorer et de comprendre des structures complexes que les statistiques et la modélisation ne peuvent pas révéler.
Le groupe de Microsoft Research VIBE à Redmond, aux Etats-Unis, et l'équipe projet AVIZ d'Inria sont parmi les leaders mondiaux dans le domaine de la visualisation interactive de réseaux. Ils ont publié ensemble près de trente articles dans de grandes revues internationales au cours des six dernières années. Dans ce nouveau projet, ils vont explorer plus particulièrement la visualisation des réseaux temporels dynamiques et multidimensionnels. Pour cela, ils vont étudier à la fois les réseaux sociaux en ligne et la connectivité cérébrale.

LES TROIS CLÉS D'**UN LABORATOIRE OUVERT SUR LE MONDE**

Dès sa création, le laboratoire commun s'est inscrit dans une logique d'ouverture et de partage avec la communauté scientifique mais également académique et entrepreneuriale. Avec un objectif clairement affiché : faire le lien entre recherche fondamentale et applications concrètes pour construire ensemble le monde de demain.

Des collaborations académiques fertiles avec la recherche française et internationale

Si le Centre de recherche commun Inria-Microsoft est avant tout constitué d'équipes issues des deux organisations, il est également profondément immergé dans l'ensemble de l'écosystème de la recherche publique française. Certains de ses projets peuvent ainsi compter sur la contribution de chercheurs et enseignants-chercheurs d'institutions comme l'École normale supérieure, l'École polytechnique, les universités Paris-Sud, Rennes et Montpellier, ou encore le CNRS. **Cette organisation est le reflet de l'ancrage des équipes de recherche dans le tissu universitaire qui les entoure.**
Elle est également l'illustration de la vocation d'ouverture scientifique qui a présidé à la création de ce laboratoire commun : tous les travaux de recherche menés sont publics, et les résultats sont mis à la disposition de la communauté scientifique, tout particulièrement les logiciels, qui sont publiés en licence open source.
C'est une approche fondamentale qui contribue au dynamisme de la recherche menée dans le laboratoire commun, mais aussi à sa notoriété et à son attractivité : en témoignent les 68 doctorants de 23 nationalités différentes ayant rejoint le Centre depuis 2006 pour y effectuer leur thèse.

© COLLÈGE DE FRANCE / PATRICK IMBERT

MARTIN ABADI, titulaire de la chaire Informatique et sciences numériques (2010–2011) au Collège de France, lors de sa leçon inaugurale.

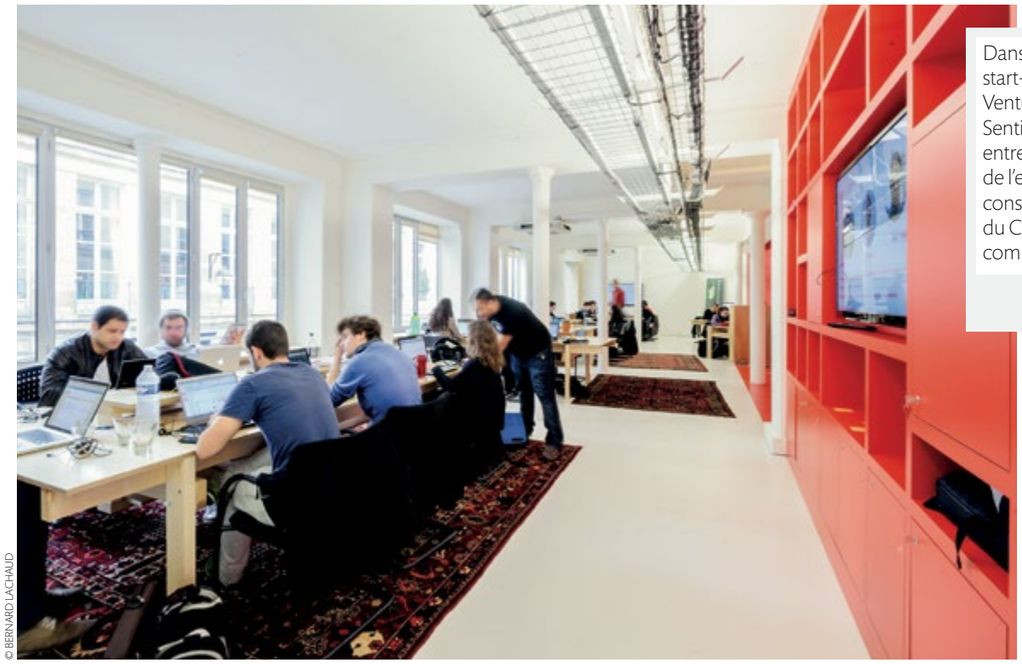
RECONSTRUCTION DU SITE DE BAT
Classé au patrimoine mondial de l'Unesco en 1988, le site de Bat, dans le sultanat d'Oman, présente de nombreux tumuli datant du troisième millénaire avant notre ère. Ces amoncellements de terre et de pierres sont très difficilement identifiables au milieu d'un paysage fortement accidenté. Les archéologues ont fait appel aux services d'Iconem pour réaliser une cartographie précise du site. Un drone de nouvelle génération a assuré le survol d'une zone de près de 400 000 mètres carrés, ce qui a ensuite permis une recomposition pierre à pierre.

© PHOTOS : ICONEM - BENJAMIN SYLVANDER/STRATOSPHERE

De la recherche à la start-up : accélérer le transfert

Il n'y a parfois qu'un pas entre un projet de recherche et la naissance d'une ambitieuse start-up. Car si le Centre de recherche commun s'est construit autour de travaux de recherche fondamentale, les applications pratiques sont immédiatement envisagées, que ce soit dans le domaine des sciences, dans le monde du Web ou la vie numérique.
Dans ce domaine, les évolutions du projet Digital Image & Video Mining, commencé en 2008 sous la direction de Jean Ponce, professeur à l'École normale supérieure, en sont une très belle illustration. C'est à l'origine un projet de « photogrammétrie » à grande échelle : reconstruire automatiquement, en 3D et à partir de simples photos, les décors disparus d'une des plus grandes villas patriciennes de Pompéi, la villa de Diomède (voir page 14). Le projet a ensuite donné lieu à la naissance d'une start-up, Iconem, hébergée au sein de l'incubateur public parisien Agoranov avant de prendre son envol.
C'est sous l'impulsion d'Yves Ubelmann, architecte et collaborateur du projet de recherche initial, qu'Iconem a vu le jour. Avec une idée un peu folle en arrière-fond : mettre les technologies de photogrammétrie au service de la sauvegarde des grands sites archéologiques et historiques mondiaux menacés de disparition par l'érosion ou l'activité humaine. De l'Afghanistan à l'Albanie, en passant par la France ou le Pakistan, **la start-up accompagne des organismes de recherche, des collectivités locales et des entreprises privées dans leurs missions d'archéologie et de conservation du patrimoine.**
La start-up a pu ainsi pleinement bénéficier du projet autour de la villa de Diomède : il a constitué à la fois un banc d'expérimentation sur le terrain, un apport technique et logiciel novateur, ainsi qu'une collaboration fructueuse entre informaticiens, architectes et archéologues pour contribuer à transformer en profondeur la protection de notre patrimoine.

Faire germer l'étincelle créative



© BERNARD LACHAUD

Dans l'accélérateur de start-up de Microsoft Ventures, situé dans le Sentier à Paris, de jeunes entrepreneurs bénéficient de l'expertise et des conseils des chercheurs du Centre de recherche commun.

Les chemins de l'innovation sont multiples et tout sauf linéaires. Ils naissent souvent de la rencontre entre des chercheurs et des créateurs qui ne partagent pas toujours la même culture : entre le temps long de la recherche et les cycles de recherche et développement très courts d'une partie de la génération actuelle des start-up, le fossé peut parfois sembler insurmontable. Et pourtant…

De plus en plus en souvent, des start-up à la pointe de l'innovation se trouvent confrontées à une barrière technologique relevant de la recherche pure. Elles se retrouvent ainsi bloquées, soit parce qu'elles ne peuvent plus avancer seules, soit parce que leur méconnaissance des dernières avancées les handicape par rapport à leurs concurrents. C'est pour répondre à ce problème et à ce manque de partage que le Centre de recherche commun est partenaire du Microsoft Ventures Accelerator Paris, l'accélérateur de start-up de Microsoft France. Laurent Massoulié, directeur du Centre de recherche commun, est ainsi membre du comité de sélection des promotions de start-up, et il apporte son regard de chercheur. Il peut être critique (la start-up ne risque-t-elle pas de se retrouver dans une impasse faute de technologie adaptée ?), mais sa mission première est de transmettre et de créer des ponts entre les chercheurs et les porteurs de projets innovants.
Des sessions de coaching, mais également des présentations et des entretiens, sont ainsi régulièrement organisés entre chercheurs et créateurs.
Au programme : Big Data, apprentissage statistique et… applications concrètes pour inventer le monde de demain !

CENTRE DE RECHERCHE COMMUN
Inria - Microsoft Research

Centre de recherche commun Inria-Microsoft Research
Campus de l'École polytechnique, bâtiment Alan Turing
1, rue Honoré-d'Estienne-d'Orves - 91120 Palaiseau, France
Tél. : 01 65 35 69 69 – Fax : 01 69 35 69 69
www.msr-inria.fr