

Ciência no palheiro

Programa em eScience busca extrair novos conhecimentos em meio a volumes gigantescos de dados

A FAPESP lançou uma chamada de propostas que inaugura seu Programa de Pesquisas em eScience, expressão que resume o desafio de pesquisa conjunta em computação e outras áreas do conhecimento, para organizar, classificar e garantir acesso ao gigantesco volume de dados gerados constantemente em todos os campos de pesquisa, visando extrair novos conhecimentos e fazer análises abrangentes e originais. O objetivo principal do programa é integrar grupos envolvidos com pesquisas sobre interfaces, algoritmos, modelagem computacional e infraestrutura de dados com cientistas de áreas em que aplicações de eScience são especialmente necessárias no país, das ciências agrárias às sociais. “A intenção é colocar pesquisadores dessas duas frentes trabalhando juntos para gerar conhecimentos novos tanto em ciência da computação como em aplicações nessas disciplinas”, diz Roberto Marcondes César Júnior, professor do Instituto de Matemática e Estatística da Universidade de São Paulo e coordenador adjunto de Ciências Exatas e Engenharias da FAPESP.

A primeira chamada receberá propostas nas modalidades Projeto Temático ou Auxílio Regular até 28 de abril de 2014. A Fundação espera que as universidades ofereçam contrapartida de recursos humanos, contratando programadores e analistas de banco de dados, entre outros.

Serão disponibilizados R\$ 4 milhões para apoiar projetos que envolvam modelos matemáticos, repositórios digitais e gerenciamento de dados, novos *hardwares*, *softwares*, protocolos, ferramentas e serviços, voltados para atender a demandas de pesquisas nas áreas de ciências agrárias; artes, humanidades e ciências sociais; engenharia e física, clima e ciências da Terra, e à prática e educação em eScience.

As ideias iniciais surgiram de dois *workshops* organizados pela FAPESP que contaram com pesquisadores em computação e em outras áreas (das exatas às humanidades). “Toda a discussão levou à conclusão de que o ideal seria uma iniciativa em eScience”, diz Claudia Bauzer Medeiros, professora do Instituto de Computação da Unicamp e membro da Coordenação da área de Engenharia e Ciência da Computação da FAPESP. “Os detalhes foram sendo amadurecidos e culminaram em uma proposta de

programa redigida pela Coordenação de Computação da FAPESP.”

O edital tem um aspecto pioneiro no país ao exigir dos proponentes a submissão de um Data Management Plan, que descreve como o projeto pretende gerenciar, proteger, preservar e divulgar seus dados. “Muitos países, como Estados Unidos, Alemanha, Reino Unido e Canadá, estão discutindo como garantir tais planos em todo projeto de pesquisa, pois é consenso que uma das partes mais valiosas de qualquer pesquisa são os dados gerados, que precisam ser preservados e divulgados adequadamente”, observa Claudia Bauzer Medeiros.

“Para a FAPESP, avançar em eScience é fundamental, pois ela tem programas que geram uma quantidade inimaginável de dados de interesse científico que necessitam de um tratamento adequado e devem ser compartilhados”, diz Marcondes César, referindo-se a programas como o





Biota, de identificação da biodiversidade paulista, o Bioen, de pesquisa em bioenergia, o Programa FAPESP sobre Pesquisa em Mudanças Climáticas Globais ou o Cinapce, de pesquisas sobre o cérebro. A experiência dos 46 projetos de pesquisa, 32 deles já concluídos, do Instituto Microsoft Research-FAPESP de Pesquisas em Tecnologia de Informação, iniciativa que busca aplicações de alcance social em tecnologia de informação, sinaliza que há uma comunidade de pesquisadores apta a participar do novo programa.

PLANILHAS

O advento do eScience se relaciona com mudanças na forma de fazer ciência. “Até poucos anos atrás, um doutorado em biologia, por exemplo, dependia de um conjunto de experimentos nos dois ou três primeiros anos, compilados depois numa planilha cujos dados eram analisados no último ano. Já hoje é comum que

Pesquisas sobre a internet

A FAPESP e os ministérios da Ciência, Tecnologia e Inovação (MCTI) e das Comunicações firmaram um convênio de cooperação no valor de R\$ 98 milhões, no dia 18 de dezembro passado, para apoiar pesquisas científicas e tecnológicas que contribuam para o desenvolvimento da internet no Brasil. O valor corresponde aos recursos remanescentes do período em que a FAPESP, por delegação do Comitê Gestor da Internet no Brasil (CGI.br), geriu as atividades de registro de domínio e alocação de endereços IP no país, entre 1998 e dezembro de 2005, quando essa tarefa foi assumida pelo Núcleo de Informação e Coordenação do

Ponto BR (NIC.br). “Em 1998, a internet no Brasil contava com 27 mil domínios. Hoje são mais de 3 milhões”, pontuou o presidente da FAPESP, Celso Lafer. “A FAPESP abriu-se para ajudar o CGI num momento em que ainda não existia o NIC.br, e esse apoio foi fundamental”, ressaltou o ministro Marco Antonio Raupp, do MCTI. Os recursos serão distribuídos entre projetos apresentados por pesquisadores de todo o país, proporcionalmente ao número de registros de domínios solicitados por cada estado naquele período. São Paulo contará com 47% dos R\$ 98 milhões para apoiar projetos no âmbito do convênio que envolve os dois ministérios e a FAPESP.

no primeiro ano o aluno tenha acesso a centenas de planilhas sobre um determinado experimento disponíveis na internet e seu desafio seja descobrir conhecimento novo em meio àquele volume de informações”, explica o professor. “O problema de pesquisa se transformou. Trata-se de extrair conhecimento de um volume grande de informações, na maioria das vezes com dados heterogêneos.”

Diversas iniciativas internacionais enfrentam os desafios de eScience, caso do Institute for Data Sciences and Engineering, da Universidade de Colúmbia. Organizado em temas interdisciplinares, pesquisa formas de extrair informações de mídias *on-line*, de monitorar e melhorar o uso da infraestrutura urbana ou de aprimorar o sistema de atendimento de saúde utilizando dados de pacientes e registros de saúde pública. Outro exemplo é o eScience Institute, da Universidade de Washington, EUA, que abriga pesquisas que vão desde astronomia até biologia marinha. Há também telescópios virtuais, como o SkyView, da Nasa, que permitem que milhões de pessoas acessem dados astronômicos do universo. Existe ainda a iniciativa mundial de observatórios astronômicos virtuais, para processar dados astronômicos mundialmente, permitindo que milhões de pessoas acessem o universo. Outro exemplo é o LHC, o maior acelerador de partículas do mundo, cujos resultados são processados mundialmente.

A *expertise* que os cientistas buscam para extrair informações em meio a gran-

des volumes de dados já foi alcançada pelo setor privado, observa Gilberto Câmara, pesquisador de geoinformática e modelagem ambiental do Instituto Nacional de Pesquisas Espaciais (Inpe) e diretor da instituição entre 2005 e 2012. “Bancos e *sites* de comércio eletrônico criaram uma estrutura de gerenciamento de dados gigantesca para obter informações sobre os clientes e seus hábitos e prestar serviços, mas a forma como estão organizados arquivos científicos mudou muito pouco”, afirma, citando como exemplo as 500 mil imagens de satélite oferecidas gratuitamente pelo Inpe. “Eu posso baixar uma por uma, mas quem precisa gerenciar mais de 200 fotos pessoais sabe a dificuldade que isso gera. Um desafio do eScience é fazer com que o conjunto delas possa ser analisado sem que eu precise baixá-las”, afirma.

Câmara cita um artigo da revista *Science*, publicado em novembro passado, com os resultados da pesquisa de um grupo que mapeou as mudanças na cobertura de florestas do mundo entre 2000 e 2010, a partir de 650 mil imagens do satélite Landsat. “O Google, que tem uma infraestrutura poderosa em sistemas de armazenamento e está na liderança nesse campo, franqueou ao grupo de pesquisa o acesso às imagens. Há uma revolução em andamento no tratamento de dados científicos. Ela vai permitir que se faça ciência de um jeito diferente, em que será possível levar o pesquisador até os dados e não mover os dados para o pesquisador.” ■

Fabrcio Marques