



Una avalancha de datos

El avance de la *eScience* altera el modo tradicional de hacer ciencia

Yuri Vasconcelos

PUBLICADO EN NOVIEMBRE DE 2014

Ya se fue el tiempo en que el problema de los científicos era la dificultad en obtener datos para dar prosequimiento a sus investigaciones. En muchas áreas del conocimiento, los recientes avances en el área de tecnología de la información, con la democratización de la informática, la ampliación de las redes computacionales y la multiplicación de las fuentes de información, tuvieron como efecto directo una producción intensiva de datos. Esto sucede en campos tan disímiles como la astronomía, inundada diariamente por miles de imágenes e información de cuerpos celestes captados por potentes telescopios, la biología molecular, beneficiada por el surgimiento de máquinas de secuenciación genética de alto desempeño, y la ecología, favorecida por una serie de tecnologías y sensores capaces de transformar con precisión las transformaciones por las que pasan distintos biomas. Todo esto lleva a los científicos a depararse con un nuevo problema: cómo hacer para procesar, elaborar y visuali-

zar la avalancha de datos adquiridos a través de los más diversos medios. Para dar una respuesta a este dilema, una nueva rama de la ciencia ha cobrado cada vez más importancia: la *eScience*, que recurre a modelos matemáticos y herramientas informáticas para analizar información y acelerar la investigación en otros dominios del conocimiento.

“La idea de conectar la práctica científica tradicional con el acceso, el uso y el procesamiento de grandes cantidades de datos alterará el modo de hacer ciencia y aumentará su potencialidad. La FAPESP se ubica a la vanguardia de este proceso y al final del año pasado lanzó el Programa *eScience*”, dijo Carlos Henrique de Brito Cruz, director científico de la Fundación, durante el Microsoft *eScience* Workshop 2014, realizado entre los días 20 y 22 de octubre en la localidad de Guarujá, en el litoral paulista. El objetivo del programa consiste en organizar el integrar grupos que toman parte en la investigación sobre algoritmos, modelado computacional e infraestructura de datos con equipos de

científicos que actúan en otras áreas del conocimiento, tales como biología, ciencias sociales, medicina y humanidades.

UN DESAFÍO MUNDIAL

“Una de las principales barreras que enfrentaremos es quizá la dificultad de comunicación entre los científicos de los equipos necesarios para hacer ciencia de esta manera, fuertemente basada en datos o en grandes cantidades de datos. Esto requiere una comunicación sumamente efectiva entre investigadores del área de ciencias de la computación y científicos de otros dominios. Es un desafío en Brasil y en cualquier lugar”, dijo Brito Cruz. El director científico de la FAPESP fue uno de los disertantes de la mesa redonda intitulada “La importancia estratégica de la *eScience*”, que también contó con la presencia de los científicos Jason Rhody, director sénior del Office of Digital Humanities da National Endowment for the Humanities, y Chris Mentzel, director de la Gordon and Betty Moore Foundation, dos organizaciones estadounidenses que mantienen programas de apoyo a la ciencia.

“En la actualidad, todos los campos de la investigación científica se ven afectados por la escala moderna de pro-

ducción de datos”, dijo Mentzel, quien hizo hincapié en la importancia de los denominados *data scientists* –o científicos de datos–, denominación dada a los profesionales que trabajan con el gran volumen de datos que se generan en el marco de las investigaciones, para producir con éstos nuevos conocimientos. “Son investigadores que trabajan entre las disciplinas. Son establecedores de puentes”, dijo. En la Gordon and Betty Moore Foundation, Mentzel se encuentra al frente de un programa que cuenta con un presupuesto de 60 millones de dólares destinado a incentivar iniciativas en el área de *eScience*. Para Rhody, los científicos afrontan un cambio de paradigma. “Estamos pasando de una cultura de escasez a una cultura de abundancia de datos.”

El término *eScience*, acuñado en 1999 por el director de la Oficina de Ciencia y Tecnología del Reino Unido, John Taylor, tiene también otras denominaciones, tales como ciencia impulsada por datos (*data-driven science*) y computación fuertemente impulsada por datos (*data-intensive computing*). Algunos países, tales como Estados Unidos e Inglaterra, cuentan con programas con apoyo del gobierno que se abocan al desarrollo



de esta nueva área científica. En Brasil, cabe destacar al Núcleo de Investigación en *eScience* de la Universidad de São Paulo (USP), institucionalizado en 2012. Coordinado por el profesor Roberto Marcondes Cesar Junior, del Instituto de Matemática y Estadística (IME) y uno de los coordinadores adjuntos del área de ciencias exactas e ingenierías de la Dirección Científica de la FAPESP, está integrado por 20 investigadores.

El Microsoft *eScience* Workshop 2014 se realizó en simultáneo con la 10ª Conferencia Internacional IEEE en *eScience*, organizada por la Computer Society del Institute of Electrical and Electronics Engineers (IEEE), una entidad creada en Estados Unidos por ingenieros electricistas y electrónicos. Durante dicho evento tuvo lugar un panel con científicos contemplados con becas del Instituto Virtual de Investigaciones FAPESP-Microsoft Research, que relacionan las aplicaciones de las ciencias de la computación con los desafíos de la ciencia básica en áreas vinculadas con los cambios climáticos y otras disciplinas asociadas al medio ambiente. Uno de los trabajos presentados explora soluciones innovadoras para el monitoreo de plantas en los trópicos, combinando investigación en ciencias de la computación y fenología. Esta última, una de las ramas más antiguas de la ciencia, es un área de la ecología en la cual se estudian los fenómenos cíclicos de las plantas, tales como la aparición de hojas, pimpollos, flores y frutos, y su relación con las condiciones ambientales.

El proyecto, coordinado por la profesora Leonor Patricia Morellato, del Laboratorio de Fenología del Instituto de Biociencias de la Universidade Estadual Paulista (Unesp) de la ciudad de

Hay que conectar las aplicaciones de las ciencias de la computación con la fenología, que estudia los fenómenos cíclicos de las plantas



Rio Claro (São Paulo), prevé la combinación de tecnologías para el monitoreo de las alteraciones que sufren en el transcurso del tiempo las vegetaciones nativas de sabana o cerrado, de bosque atlántico, de matorrales de altura e incluso de *caatinga*. El área central de la investigación está en la zona de Itirapina, en el interior de São Paulo. “Aparte de la observación directa de las plantas en el suelo, instalamos una cámara en la cima de una torre, a 18 metros del suelo, para fotografiar diariamente la vegetación, y una estación meteorológica. También contaremos con un vehículo aéreo no tripulado (vant) equipado con un sensor hiperespectral y una cámara para agregar una escala espacial a la recolección de datos”, dice la investigadora. Los sensores hiperespectrales, provistos de una alta resolución espacial, pueden suministrar detalles sobre las características fisicoquímicas y respuestas fisiológicas de las imágenes de las plantas. Para Morellato, la fenología es una de las mejores herramientas para entender los efectos de los cambios climáticos en las plantas. “Esto ya está establecido en zonas templadas, donde el disparador de la fenología es la temperatura ambiental y la duración del día. Pero poco sabemos acerca de qué sucede en las vegetaciones tropicales. Con los datos de la cámaras y del sensor hiperespectral, pretendemos definir cuáles son los disparadores de la fenología en el trópico, es decir, qué es lo que lleva al surgimiento de flores, frutos y hojas en las plantas en determinados momentos”, dice Morellato.

PARA ANALIZAR IMÁGENES

De acuerdo con Morellato, sin la ayuda de investigadores y recursos de las ciencias de la computación sería imposible realizar la investigación científica. “El volumen de datos que recolectaremos es gigantesco. Una cámara digital registra 60 fotos por día. Tenemos 11 cámaras que monitorean seis tipos de vegetación y debemos observar la evolución durante al menos una estación de crecimiento, para posteriormente relacionarla con el clima. Luego hay que procesar y analizar todas las imágenes, lo que sería imposible con una planilla electrónica simple. Es necesario contar con apoyo para trabajar con ese *big data*. A tal fin, una maestranda creó un banco de datos especial para el proyecto, y una posdoctoranda trabajó

Los sensores inalámbricos instalados en bosques generan cuantiosos datos sobre procesos naturales

en un *software* para la visualización y la organización de las imágenes.”

La investigación de la profesora de la Unesp cuenta con la colaboración del científico Ricardo Silva Torres, director del Instituto de Computación de la Universidad de Campinas (Unicamp), también contemplado con un proyecto en el marco del acuerdo FAPESP-Microsoft Research. Silva Torres está al frente de un estudio cuyo objetivo consiste en desarrollar nuevas técnicas analíticas y herramientas informáticas para el procesamiento de imágenes de monitoreo remoto, con el fin de analizar la dinámica de algunos biomas a escalas regionales y continentales. El enfoque del trabajo, realizado en colaboración con la profesora Marina Hirota, del Departamento de Física de la Universidad de Santa Catarina (UFSC), recae sobre los biomas tropicales sudamericanos.

Otro trabajo presentado en el evento de Guarujá está encabezado por el ecólogo de la Unicamp Rafael Silva Oliveira, quien trabaja en colaboración con los investigadores Antonio Alfredo Ferreira Loureiro, del Departamento de Ciencias de la Computación de la Universidad Federal de Minas Gerais (UFMG), y Stephen Burgess, de la University of Western Australia. “Nuestro estudio tiene por meta investigar la dinámica de agua y carbono en bosques nubosos, pasturas y áreas de transición entre ambos”, dice Silva Oliveira. Los bosques nubosos se encuentran en lo alto de las montañas tropicales. “Pretendemos entender de



qué modo procesos claves, tales como la absorción y el almacenamiento de carbono, la traspiración de los árboles y la captación de agua de la neblina en la vegetación, se ven afectados por los cambios en el uso de la tierra y por las variaciones del clima.”

Los estudios de campo se realizan en una región de bosque ubicada en Serra da Mantiqueira, en la zona serrana de Campos do Jordão, interior paulista. Según Silva Oliveira, se está instalando en ese lugar una red de sensores inalámbricos para monitorear en tres capas del ecosistema (la atmósfera, la vegetación y el suelo), parámetros microclimáticos del metabolismo de la vegetación y de la dinámica del agua en el suelo. “Estos datos podrán mejorar el pronóstico de los impactos ambientales generados por los cambios en la utilización de la tierra y, al mismo tiempo, harán posible el desarrollo de modelos hidrológicos y de circulación biosfera-atmósfera con mejor capacidad predictiva”, explica Silva Oliveira. ■

Proyectos

1. Towards an understanding of tipping points within tropical South American biomes (nº 2013/ 50169-1); **Modalidad** Programa de Investigación en Asociación para la Innovación Tecnológica (Pite) y Acuerdo FAPESP-Microsoft; **Investigador responsable** Ricardo da Silva Torres (Unicamp); **Inversión** R\$ 384.838,38 (FAPESP).
2. Combining new technologies to monitor phenology from leaves to ecosystems (nº 2013/ 50155-0); **Modalidad** Programa FAPESP de Investigaciones sobre Cambios Climáticos Globales – Programa de Investigación en Asociación para la Innovación Tecnológica (Pite) y Acuerdo FAPESP-Microsoft; **Investigadora responsable** Leonor Patrícia Cerdeira Morellato (Unesp); **Inversión** R\$ 1.115.752,48 y US\$ 535.902,72 (FAPESP).
3. Interacciones entre el suelo, la vegetación y la atmósfera en un paisaje tropical en transformación (nº 2011/ 52072-0); **Modalidad** Programa de Investigación en Asociación para la Innovación Tecnológica (Pite) y Acuerdo FAPESP-Microsoft; **Investigador responsable** Rafael Silva Oliveira (Unicamp); **Inversión** R\$ 644.800,74 y US\$ 663.429,82 (FAPESP).