

de modo que cuando se proyecten los puntos individuales (X_{d1i} , X_{d2i}) sobre el citado eje W_1 se maximice, tras la proyección, la varianza *entre-grupos* (es decir, entre las medias de los grupos proyectados) en unidades de varianza *dentro de los grupos* (es decir, la media ponderada de las varianzas de los grupos proyectados, con pesos de ponderación iguales a la importancia de cada grupo). Si solo existieran dos grupos, la idea de Fisher es clara: se intenta conseguir que las medias de los grupos (tras la proyección citada) se separen lo más posible (en unidades de las varianzas *dentro de los grupos*). De esta forma, se llega a determinar una primera regla de clasificación: *un individuo se asigna a un grupo determinado si, tras la proyección citada, la distancia entre el individuo y la media del grupo es la mínima entre todas las posibles*. La Figura 2(b) muestra que el citado primer eje óptimo W_1 forma un ángulo de 25° con el eje cartesiano horizontal X_{d1} (a diferencia de la primera componente principal del apartado anterior, para el cual, como se indicó, el ángulo era de 38°). Al igual que en el análisis factorial, pueden buscarse ejes adicionales de discriminación, si bien ahora la condición de perpendicularidad de las distintas transformadas no tiene sentido, puesto que el interés se centra en los grupos. En la Figura 2(b) se ha incluido el segundo eje W_2 de discriminación asociado a los datos considerados en este apartado, donde se observa la citada falta de ortogonalidad con el anterior eje W_1 .

4. CONCLUSIONES

Las técnicas clásicas de *Análisis Multivariantes* engloban toda una serie de técnicas de análisis de datos con objetivos muy dispares. Puesto que la naturaleza del problema que se pretende abordar en cada caso ha motivado el desarrollo de las distintas técnicas citadas, se ha presentado en el apartado 2 una taxonomía donde se pivota a partir de la propia matriz de datos.

Aunque no siempre, una buena parte de dichas técnicas involucran (en parsimonia) transformaciones lineales de los datos originales, admitiendo por ello una sencilla interpretación geométrica. Los ejemplos concretos comentados someramente en el apartado anterior corroboran estos dos últimos puntos, al tiempo que constituyen una muestra no exhaustiva, como es lógico, de los mencionados procedimientos clásicos de análisis de datos.

REFERENCIAS

- Anderson, T.W. *Introduction to Multivariate Statistical Analysis*. Wiley (1958).
- Arnold, S.F. *The Theory of Linear Models and Multivariate Analysis*. Wiley (1981).
- Bock, R.D. *Multivariate Statistical Methods in Behavioral Research*. McGraw-Hill (1975).
- Carroll, J.D. y Green, P.E. *Mathematical Tools for Applied Multivariate Analysis*. Academic Press (1997).
- Eaton, M.L. *Multivariate Statistics*. Wiley (1983).
- Fisher, R.A. The use of multiple measurement in taxonomic problems. *Annals of Eugenics*, **7**, 179-188 (1936).
- Hotelling, H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, **24**, 417-441, 498-520 (1933).
- Hotelling, H. The most predictable criterion. *Journal of Educational Psychology*, **26**, 139-142 (1935).
- Hotelling, H. Relations between two sets of variates. *Biometrika*, **28**, 321-377 (1936).
- Morrison, D.F. *Multivariate Statistical Methods*. McGraw-Hill (1976).
- Spearman, C. "General-intelligence objectively determined and measured". *American Journal of Psychology*, **15**, 201-293 (1904).
- Srivastava, M.S. y Khatri, C.G. *An Introduction to Multivariate Statistics*. North Holland (1979).

Teófilo Valdés

Dpto. de Estadística

Facultad de Matemáticas

Universidad Complutense de Madrid

Beowulf Superconductores caseros

INTRODUCCIÓN

*To Beowulf now the glory was given,
and Grendel thence death-sick his
den in the dark moor sought, noisome
abode: he knew too well that here
was the last of life, and end of his
days on earth.*

Beowulf, Anónimo

Beowulf, guerrero escandinavo del siglo VI, es famoso por ser el único héroe capaz de derrotar al aterrador monstruo Grendel. Su historia se relata en un manuscrito, del que sólo existe un ejemplar, que data de algo antes del siglo X. Este manuscrito constituye la pieza épica en Inglés Antiguo más antigua que se conserva [1, 2].

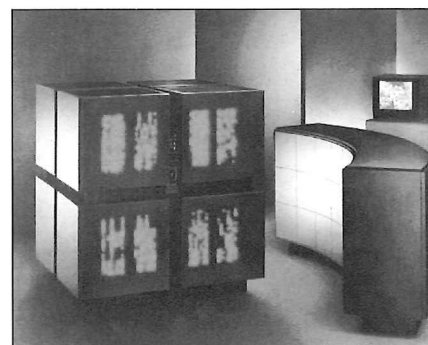


Figura 1. *Connection Machine*.

En el siglo XXI, los modernos guerreros de la computación desafían problemas que manejan ingentes cantidades de datos y que realizan cantidades astronómicas de cálculos. Para ello utilizan un computador de alto rendimiento construido con medios caseros y que recibe su nombre de este héroe épico: Beowulf.

La clave de Beowulf es el sistema operativo Linux [3], una versión libre de Unix que funciona en sistemas basados en procesadores Intel-compatibles, DEC Alpha o Power PC, entre otros. El código fuente de Linux se distribuye libremente, por lo que cualquiera puede mejorarlo, extenderlo o incluso contribuir a su desarrollo.

¿QUÉ ES UN SUPERCOMPUTADOR?

Para empezar, definamos lo que es un computador. Sin entrar en detalles, es una máquina destinada a procesar datos. El conjunto de instrucciones que definen las operaciones a realizar sobre los datos constituyen un *programa*. El encargado de ejecutar esas instrucciones, el “cerebro” del computador, recibe el nombre de *procesador*. Los programas y los datos sobre los que éstos trabajan se almacenan en la *memoria* del computador.

Las aplicaciones científicas a las que se pueden destinar los computadores son muy variadas: simulación de los brazos espirales de una galaxia, física de partículas, modelos climáticos, análisis de datos económicos y bursátiles, “*data mining*”, búsqueda y almacenamiento de información en Internet, etc. A medida que estos problemas se hacen más complejos (tanto por lo que se refiere al número de operaciones a realizar como a la cantidad de datos a manejar) se hacen necesarios computadores más y más potentes.

La potencia de un computador queda determinada, en gran medida, por la de su procesador, aunque no es el único factor que influye. Si bien la potencia de los procesadores mejora de día en día, en un momento dado obtener un procesador diez veces más potente supone un desembolso económico mucho más de diez veces superior.

Una manera de hacer frente a este problema consiste en utilizar muchos procesadores menos potentes en lugar de un único procesador muy potente y caro. El problema a resolver se reparte entonces entre todos los procesadores disponibles, de manera que la solución se obtiene de manera cooperativa. Normalmente nos referimos a estos sistemas como *computadores paralelos*, ya que el trabajo de cada procesador se realiza de forma paralela al de los restantes. Y éste es, de manera intuitiva, el concepto de *supercomputador*.

Un ejemplo muy significativo es *Connection Machine* (Figura 1) [4], con sus 65.536 procesadores. *Deep*

Blue, de IBM, con 256 procesadores especializados en ajedrez, se enfrentó a Kasparov a seis partidas en 1997. El resultado: una victoria de Kasparov, dos victorias de *Deep Blue* y tres tablas [5].

Ya que en un supercomputador varios procesadores van a trabajar en paralelo en la resolución de un problema se hace necesaria la comunicación entre ellos. Hay problemas en los que los procesadores pueden tra-



Figura 2. *Cplant*.

bajar gran parte del tiempo “en solitario”, sin comunicarse con los demás, mientras que en otros problemas la comunicación permanente entre procesadores es fundamental. Por este motivo el diseño de una red de comunicaciones entre los procesadores es fundamental: ha de ser capaz de soportar el tráfico de información sin convertirse en el cuello de botella del sistema. Existen numerosas topologías de red, cada una con unas características propias que las hacen adecuadas para ciertos tipos de problemas. La programación eficaz de un supercomputador es complicada y en ningún momento se puede perder de vista la topología específica del sistema que se está utilizando: “Se escriben programas determinados para máquinas determinadas”.

El principal problema de un supercomputador es su alto coste: su adquisición es cara, los programas son caros, resultan caros de mantener y caros y difíciles de ampliar. Además, ya que son máquinas “con

nombre y apellidos”, se depende en exclusiva de un único fabricante, dependencia que en muchas ocasiones puede no resultar conveniente.

Dados los inconvenientes, la conclusión es obvia: el científico pobre no puede disponer de un supercomputador; sólo las grandes instituciones pueden permitírselo. Entonces, ¿por qué no construirse uno mismo un supercomputador? ¿Es esto posible? ¡Sí!

EL PROYECTO BEOWULF

Un supercomputador de clase Beowulf consta de una serie de computadores tipo PC interconectados entre sí por medio de una red de comunicaciones de alta velocidad. Cada uno de estos PC's se comunica con los demás mediante el paso de mensajes a través de la red.

El Proyecto Beowulf fue puesto en marcha por el Dr. Thomas Sterling y Donald Becker, del Center of Excellence and Information Systems Sciences (CESDIS) del Centro Espacial Goddard de la NASA, en 1994. Su objetivo era crear los programas necesarios para permitir el funcionamiento del grupo de PC's como un único *computador paralelo virtual* [6].

El punto más difícil de resolver fue la existencia de diferentes adaptadores de red para PC, cada uno de los cuales necesitaba programas controladores específicos [7]. Hoy en día esto ya no es un problema, pues un Beowulf puede trabajar con cualquiera de las redes que se pueden encontrar en la red de área local de una oficina: Ethernet (estándar, Fast y Gigabit), ATM, Myrinet, etc. Esta capacidad está perfectamente inte-



Figura 3. *El proyecto Active Mural*.



Figura 4. Sala de computadores de Digital Domain.

grada en Linux, instalado actualmente en varios millones de computadores de todo el mundo.

Los escépticos no creyeron que un supercomputador se pudiera construir con estos componentes, que se pueden encontrar en cualquier tienda de informática, y con Linux como sistema operativo, que está programado por voluntarios de todo el mundo que no cobran por hacerlo.

Bajo los auspicios del proyecto *ESS* (Earth and Space Sciences) se construyeron varios Beowulfs de 16 procesadores (16 PC's) cada uno. Fueron capaces de funcionar 290 días seguidos sin fallos y en noviembre de 1996 dos de ellos, con procesadores Pentium Pro a 200 MHz, excedieron la velocidad de 2 gigaflops (dos mil millones de operaciones en coma flotante por segundo). Esta hazaña mereció el Gordon Bell Prize a la mejor relación precio/rendimiento en 1997.

Esto inspiró a otros laboratorios e instituciones académicas la construcción de sus propios computadores Beowulf. Desde sistemas de cuatro procesadores con fines educativos a verdaderos supercomputadores con varios miles de procesadores en grandes laboratorios.

Hoy, los científicos que se embarcan en la construcción de su propio supercomputador pueden obtener todos los programas y la documentación necesarios, así como Linux, de manera totalmente libre y gratuita en Internet. Y, "a pesar de su precio", estos programas son altamente fiables y están muy optimizados,

pues al estar su código fuente disponible a todos son muchos los que los ponen a prueba y los mejoran continuamente.

LA RECETA PARA EL ÉXITO

Las claves de su éxito son numerosas. En primer lugar, a diferencia de los supercomputadores comerciales, se trata de una tecnología familiar (simples PC's), de manera que es fácil emprender la construcción del sistema, así como fácil resulta su uso y mantenimiento.

El coste inicial es bajo: se puede empezar con PC's y redes baratos. ¡Basta con dos PC's, y los programas son gratis! Se pueden utilizar computadores de "penúltima generación", que se tienen infrutilizados o abandonados, o incluso computadores que se están utilizando simultáneamente para otras labores. De hecho, atendiendo al uso que se le da a los computadores individuales, se clasifican estos supercomputadores caseros en dos grandes grupos:

- **Red de estaciones de trabajo (o de PC's).** Los PC's que la integran se dedican simultáneamente a otros fines. Parte de la potencia computacional de cada uno de ellos se dedica al computador paralelo virtual. Esto es transparente al usuario habitual del computador, que lo sigue utilizando de la manera habitual.
- **Beowulf:** Este término se reserva habitualmente para los sistemas en que los PC's se dedican en exclusiva al computador paralelo virtual. Normalmente el usuario sólo puede acceder al sistema desde uno de los PC's, que es el único que tiene monitor y teclado, mientras que el resto son computadores "minimalistas", en el sentido que carecen de todo lo que resulte superfluo para el cálculo puro y duro (monitores, teclados, carcasas, incluso discos, etc.). Son, por tanto, sistemas más baratos y de mayor rendimiento que las redes de estaciones de trabajo.

La distinción entre estas dos clases de sistemas no siempre es clara y

resulta irrelevante para los objetivos de este artículo, por lo que seguiremos empleando, sin hacer distinciones, el término Beowulf para referirnos a ambas.

La ampliación del número de procesadores de un supercomputador comercial es muy costosa. Además, depende de los caprichos de un único fabricante, al que se está supe-ditado. Los Beowulf's, sin embargo, se amplían, o incluso se reducen, fácilmente, de acuerdo con las necesidades de cada momento. Basta con conectar más o menos PC's a la red.

A diferencia de los programas escritos para los supercomputadores comerciales, que son difíciles de portar de un sistema o otro, los programas escritos para un Beowulf se pueden llevar a otro con pocos o ningún cambio. Esto es cierto independientemente de los avances que se produzcan en los procesadores de los PC's de dentro de unos años. Mientras que reescribir un programa para un supercomputador comercial es difícil y caro, escribir un programa para un Beowulf es una inversión: no sabemos cómo serán los PC's de dentro de 10 años, pero los programas de los Beowulf's de hoy seguirán funcionando.

Una consideración final: dado un Beowulf, el precio de un computador comercial con un rendimiento similar es al menos un orden de magnitud superior. No hay que pensárselo mucho...

ALGUNAS MÁQUINAS

En la actualidad el número de Beowulf's en funcionamiento se puede medir en, al menos, centenas. Cada día aparecen sistemas más potentes y con mayor número de procesadores. En esta sección se van a mencionar algunos de los Beowulf's más representativos, abarcando tanto los grandes sistemas, que poseen los grandes laboratorios, como los más modestos, al alcance de un departamento universitario. Algunos de estos nombres ya son leyenda.

En primer lugar, y por derecho propio, se encuentra *Avalon* [8].

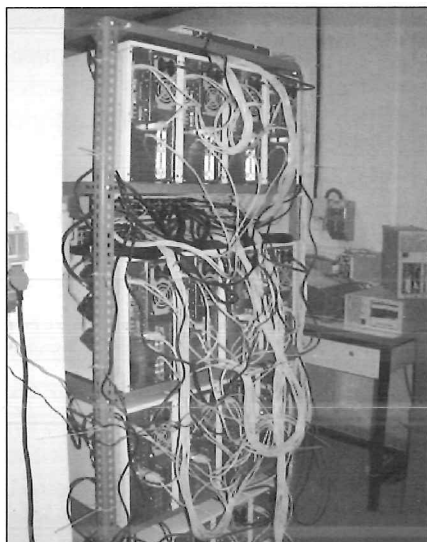


Figura 5. Vista de *Smaug*, por el lado de los cables.

Hasta hace poco, el Beowulf más poderoso [9]. Hoy en día ocupa el sexto lugar de la clasificación. Con sus 140 procesadores, se encontraba (hoy ya no) entre los 500 supercomputadores más potentes del mundo [10]. Está instalado en Los Alamos National Laboratory Center for Non-linear Studies and Theoretical Division. Se utiliza como supercomputador de propósito general, destinado a la investigación. Se le encomiendan problemas que abarcan desde la astrofísica a la dinámica molecular.

El honor de ser el Beowulf más potente a la hora de escribir estas líneas (diciembre de 2000) corresponde a *CPlant* (Figura 2) [11]. Ocupa el número 84 entre los 500 mejores supercomputadores. Instalado en Sandia National Laboratories, con sus 628 procesadores desarrolla una velocidad mantenida de 232,6 gigaflops y una velocidad punta de 580 gigaflops (a modo de comparación, el supercomputador más potente de la actualidad, *ASCI White* [12], de IBM, desarrolla 4938 y 12288 gigaflops como velocidades mantenida y punta, respectivamente).

COCOA [13], con 50 procesadores Pentium II, se destina a problemas de ingeniería aeroespacial. *Cluster Tux* [14] hace predicciones climáticas. *Chaingang* [15] realiza cálculos de química computacional para la creación de nuevos medicamentos. *Gargleblaster* [16] tiene fines docentes y de investigación.

Collage [17] tiene 18 procesadores. Se dedica al proyecto *ActiveMural*, un dispositivo de visualización de alta resolución (Figura 3). Actualmente utiliza 15 proyectores XGA y alcanza una resolución, que se espera incrementar, de 4920 x 2204 pixels.

Loki [18] es el hermano pequeño de *Avalon*. Sólo tiene 16 procesadores Pentium Pro. Realizó la simulación de la interacción gravitatoria de 9,75 millones de partículas (problema de interés en astrofísica y cosmología) en tres días, con una velocidad mantenida de alrededor de 1 gigaflop.

La industria del cine también saca partido. La empresa californiana Digital Domain [19] fue la encargada de los efectos especiales de la película *Titanic*, escrita y dirigida por James Cameron. En esta película, los efectos digitales se mezclan con imágenes reales de manera tan realista que resulta imposible discernir dónde acaba la realidad y empieza la magia digital. Para ello se contó con un sistema formado por 350 procesadores SGI, 200 procesadores DEC Alpha y 5 terabytes de disco, conectados por redes de 100 megabits por segundo o superiores (Figura 4) [20].

La UNED cuenta con, al menos, dos proyectos en marcha. *Smaug*, con 16 procesadores, es el Beowulf del Departamento de Informática y Automática. Por su parte, el Departamento de Física Fundamental cuenta con otra máquina de 8 procesadores.

Smaug tiene 16 procesadores AMD Athlon 500 MHz (Figura 5). Cada uno de ellos con una memoria principal de 128 megabytes, lo que totaliza 2 gigabytes. 15 de los PC's cuentan con un disco duro de 6 gigabytes, mientras que el PC restante, que actúa como servidor, tiene un disco de 12 gigabytes, lo que hace un total de 102 gigabytes de capacidad de almacenamiento. La red es una Fast Ethernet convencional de 100 megabits por segundo. Ha sido puesto en marcha recientemente por los autores, que a la hora de comenzar el proyecto contaban con muy poca experiencia con el sistema operativo Linux. Gracias a toda la documentación y la ayuda encontrada a través de Internet la puesta en marcha de la máquina se pudo realizar sin grandes contratiempos. Se piensa dedicar principalmente a problemas complicados de optimización y de control automático.

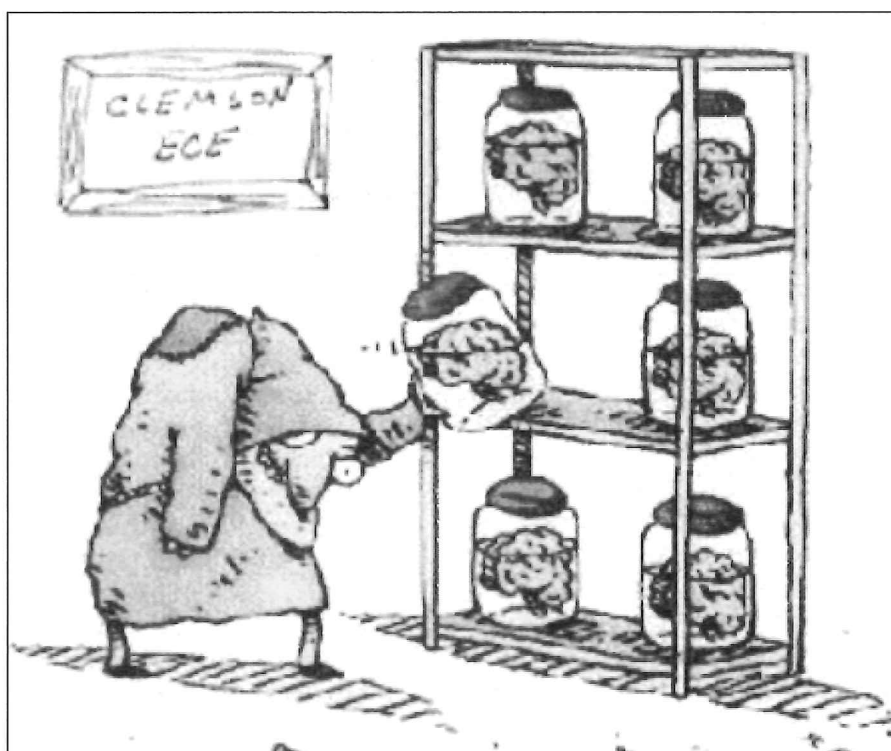


Figura 6. El supercomputador *Brainwulf*.

CONCLUSIONES

Al día de hoy ya nadie duda que el futuro de la supercomputación se llama Beowulf. Ese futuro ya está aquí. Gracias a Beowulf, científicos e ingenieros que antes debían comprar horas de supercomputador a las grandes instituciones ahora pueden disponer de su propio sistema a un precio razonable. Alguien ha comparado esta revolución a la de los años 70, cuando *Apple* puso los computadores al alcance de todo el mundo a un bajo costo.

De todas formas, montar un sistema con unas cuantas docenas de PC's sigue costando unos cuantos millones de pesetas. Si todavía le resulta excesivamente caro al lector, no todo está perdido. Su nombre: *Brainwulf*... (Figura 6).

REFERENCIAS

- [1] <http://www.lone-star.net/literature/beowulf/index.html>
- [2] <http://legends.dm.net/beowulf/index.html>
- [3] <http://www.linux.org/>
- [4] <http://mission.base.com/tamiko/cm/cm-text.htm>
- [5] <http://www.research.ibm.com/deepblue/home/html/b.html>
- [6] <http://www.beowulf.org/>
- [7] <http://www.hpcc.nasa.gov/insights/vol7/beowulf.htm>
- [8] <http://cnls.lanl.gov/avalon/>
- [9] <http://www.TopClusters.org/>
- [10] <http://www.top500.org/>
- [11] <http://www.cs.sandia.gov/cplant/>
- [12] <http://www.llnl.gov/asci/>
- [13] <http://cocoa.aero.psu.edu/>
- [14] <http://www.badcheese.com/~steve/clustertux.html>
- [15] <http://tonga.usip.edu/gmoyna/chaingang/index.html>
- [16] <http://www.cs.ucsb.edu/Facilities/Cluster/>
- [17] <http://www-fp.mcs.anl.gov/collage/>
- [18] <http://loki-www.lanl.gov/>
- [19] <http://www.d2.com>
- [20] <http://www2.linuxjournal.com/cgi-bin/frames.pl/ljissues/issue46/2494.html>

Ángel Pérez de Madrid
Pablo Sebastián Dormido Canto
Dpto. de Informática y Automática

Catenanos y rotaxanos. Máquinas en miniatura

La construcción de máquinas en miniatura o nanoescala es un objetivo de la Ciencia y Tecnología modernas, inspirado en el artículo de Richard Feynman (galardonado con el Premio Nobel de Física) "There's plenty of room at the bottom" [1] en el que propuso por primera vez la idea de una química basada en la construcción de máquinas miniaturizadas semejantes a las máquinas mecánicas y construidas átomo por átomo a escala molecular, y esta idea se ha hecho realidad. La elaboración de estas máquinas abre nuevos caminos a la investigación, ya que, como sus parientes de la biología, las enzimas, las máquinas moleculares están principalmente dirigidas hacia la transformación de una energía aportada del exterior para poder ser utilizada posteriormente.

La miniaturización de un dispositivo a nivel molecular no tiene únicamente valor académico sino que también está contribuyendo al desarrollo de la nanotecnología [2] (una molécula mide menos de un nanómetro, una millonésima de milímetro). Los progresos en la construcción de dispositivos moleculares a escala nanométrica, pueden tener importancia en muchos campos como, por ejemplo, en

ingeniería biomimética o tecnología de computadores.

A lo largo de las últimas décadas se han construido diversos tipos de moléculas semejantes a hélices, aspás, ruedas dentadas, engranajes, frenos, torniquetes, péndulos, balancines, giroscopios, y un largo etcétera, de los que se muestran más abajo algunos ejemplos [3].

Mientras los físicos e ingenieros exploran la ingeniería en nanoescala, los químicos utilizan la síntesis química para ensamblar compuestos por fuerzas intermoleculares preprogramadas. Uno de los retos más interesantes en este área es el de diseñar y sintetizar moléculas que puedan experimentar cambios en su forma, o en la posición de sus componentes, en respuesta a estímulos externos y, en principio, realizar un trabajo mecánico. La mayor parte de esta investigación se ha centrado en la clásica isomería *cis-trans* hasta llegar a los **catenanos** y **rotaxanos** [4].

Casi ha transcurrido un siglo desde que el visionario Willstätter especulara sobre una molécula con dos anillos entrelazados al igual que los eslabones de una cadena. Después de muchos intentos en 1960 Wasserman [5], sintetizó el primer **catenano**, que es así como se deno-

