

Entrevista

Héctor García Molina

Por Claudio Gutiérrez

Héctor García Molina es uno de los científicos de la computación más importantes. Nacido en México, hoy mexicano-estadounidense, es actualmente profesor de la Universidad de Stanford. Fue director del Departamento de Ciencias de la Computación de la Universidad de Stanford, institución en la que fue profesor de Larry Page y Sergey Brin, creadores y fundadores de Google. Su área de especialización son las Bases de Datos Distribuidas. Y es uno de los autores con mayor impacto en Ciencia de la Computación.

Durante el Workshop sobre Fundamentos de Bases de Datos 2011, en Santiago, que lleva el nombre de Alberto Mendelzon en homenaje a uno de los teóricos más importantes de las Bases de Datos a nivel mundial, y gran amigo de la comunidad Latinoamericana del área, tuvimos oportunidad de conversar con Héctor sobre su trabajo y nuestra disciplina.



I. Los inicios en la disciplina

¿Cómo llegaste a la computación?

Desde que estaba estudiando mi carrera en México me empezaron a interesar las computadoras, aunque estudié ingeniería eléctrica originalmente, ya había computadoras y siempre me fascinaron.

Terminé la carrera de ingeniería eléctrica en México en el Instituto Tecnológico de Monterrey, en la ciudad de Monterrey que ahora tiene diferentes campus en México. Después me fui a estudiar a Stanford en 1974 una maestría, inicialmente en ingeniería

eléctrica pero al primer año me di cuenta que lo que me interesaba más era el área de computación.

¿Qué cosas te entusiasmaron para cambiarte al área de computación?

Pues el hecho de que era un poco más matemático, podía chequear las ideas un poco más fácilmente, ver que funcionaban y la verdad es que nunca tuve mucha suerte con los circuitos eléctricos. Construí algún circuito y no funcionaba, pero le movía los alambres y sí funcionaba, y luego tenía más

fuerte con los programas que escribía, que eran más sólidos y funcionaban mejor que lo que hacía en electrónica, así que pensé “mejor me voy por la computación”.

¿Cómo elegiste el área particular en que trabajas (Bases de Datos) dentro de la Ciencia de la Computación?

Cómo soy de México originalmente y todavía no había decidido dónde me iba a quedar en el largo plazo, quería estudiar algo que además creía que era un poco más práctico y que podía tener potencial más práctico, y en aquel entonces el área de Bases de Datos me pareció un área que tiene más aplicaciones que algunas de las otras áreas de Computación y entonces pensé que tendría la puerta abierta para regresar a México o para trabajar en compañías, porque todavía no había decidido qué quería hacer. Esa fue una de las razones.

¿Y luego de eso, te entusiasmoste y seguiste un Doctorado?

Sí, después de un año me gustó Stanford. Durante mi carrera no sabía bien de qué se trataba un Doctorado, cuando estudié en el Tecnológico de Monterrey había muy poca gente con Doctorado no se hacía mucha o nada de investigación en aquel entonces. No entendía bien en qué consistía hacer un Doctorado, hasta que llegué a Stanford y empecé a entender de qué se trataba la investigación, qué hacía un investigador, qué futuro había allí y eso me empezó a entusiasmar y decidí que me quedaba en el Programa de Doctorado.

¿Con quién hiciste la tesis, quién fue tu advisor y qué temas trabajaste?

Mi asesor fue Gio Wiederhold -quien está jubilado pero todavía está ahí y participa en las juntas de nuestro grupo- y fue en el área de Bases de Datos Distribuidas. En aquel entonces me impresionó mucho uno de los trabajos de Jim Gray sobre control de concurrencia y bloqueo; me pareció

interesante y tenía algo práctico pero con buena teoría detrás. Me interesó esa área y ahí trabajé inicialmente.

Terminé mi tesis en 1979. Y aún cuando empecé a trabajar en el área de Integración de Información, pues no existía la Web (fue hasta los noventa cuando empezó), o sea había muchas de las ideas pero no tenían la importancia de la Web.

Cuando terminé el Doctorado me fui a trabajar de profesor en la Universidad de Princeton, donde estuve desde el '79 hasta finales del '91, doce años dando clases. Luego tuve la oportunidad de regresar a Stanford como profesor y empecé en enero de 1992.

II. La investigación

En tu larga experiencia como investigador, eres una de las personas que tiene uno de los altos índices de citación en el área, ¿qué desafíos actuales hay para los investigadores de Ciencia de la Computación? ¿Qué has hecho como investigador para mantenerte en las grandes ligas?

Uno de los factores -no sé si desafíos, retos o factores- importantes es estar bien conectado con los problemas de la actualidad: qué está haciendo la industria, de qué problemas está hablando la gente, porque es una buena fuente de problemas y de ideas, y el trabajo que haces entonces es más relevante, le va a interesar más a la gente si estás trabajando problemas que les interesan a las compañías. Creo que esa es una forma de tener éxito en el campo, estar pendientes, estar yendo a conferencias, de visita a compañías, tener interacciones con otros investigadores para estar al tanto de lo que está pasando.

Trabajaste en Princeton y ahora en Stanford, donde el ambiente intelectual y científico es de primer nivel, ¿cuánto incide eso?

Para mí es sumamente importante estar así, en un ambiente de muchas ideas. Tú sabes que muchos de los proyectos que hemos

hecho han empezado ya sea cuando uno de mis alumnos -de los cuales aprendo mucho siempre- o alguna visita viene. Por ejemplo teníamos un programa sobre Data Warehousing (almacenes de datos) hace muchos años y eso empezó cuando Dan Fishman que trabajaba en Hewlett Packard Labs pasó a visitarnos, no me acuerdo en qué contexto fue (quizás en una Conferencia), y nos empezó a conversar sobre su interés en el ramo industrial de Data Warehousing. Entonces con mi colega Jennifer Widom nos sentamos y dijimos “no hay nada ahí, pues debíamos hacer una copia, pero vamos a pensarlo un poquito más, a lo mejor a hay algo más de lo que está hablando la gente y por eso las compañías dicen que hay problemas, veamos cuáles son los problemas que tienen”. Entonces comencé a leer, a estudiar y sí descubrimos que había algunos aspectos interesantes en los cuales podíamos contribuir como académicos. Empezamos el proyecto y fue bien recibido, tuvo alto impacto. Ese es un ejemplo de una idea que empezó a partir del contacto con alguien de una compañía. Un problema real. Pensamos si la gente está empezando a hablar de esto, vamos a ver si podemos hacer algo o no.

¿Qué áreas dentro del mundo de los Sistemas de Información y de las Bases de Datos, piensas que son las de más perspectivas hoy, desde el punto de vista técnico?

Es mucha la cantidad de datos que se está generando. Es un problema muy interesante, hay muchas fuentes de información a través de los sistemas sociales, por ejemplo, instrumentos científicos están generando una infinidad de datos, los cuales no sólo se tienen que almacenar y hacer búsqueda de ellos, sino también explotar, hacer data mining y analizarlos. Hay muchos problemas interesantes para poder canalizar efectivamente todos estos datos, entonces es un área muy interesante. Estamos entendiendo más cómo los humanos -porque estos datos son generados por seres humanos- interactúan, se comunican; estamos aprendiendo mucho

sobre la psicología de los humanos y cómo operamos, cómo evaluamos a otros, cómo se propaga la información a través de las redes, que es un área muy importante y muy interesante.

Otra área que está empezando a interesarnos a nosotros y a otros grupos es sobre el uso de seres humanos en la computación, inspirados por “mechanical turk” y otros, donde uno puede dar trabajos pequeños a gente que por un pago pequeño contesta las preguntas o hace los trabajos. Estamos viendo cómo se organiza un sistema de cómputo o un sistema de bases de datos, por ejemplo, donde parte del trabajo y parte de la información esté en las personas, cómo se puede construir un sistema de ese tipo, híbrido entre computación y humanos, y es ahí donde estamos empezando a investigar.

El tema de este diluvio de datos que hablaba Jim Gray, es un tema que parece que está envolviendo a toda la disciplina de la Ciencia de la Computación...

Y más. Todas las disciplinas en general: medicina, economía... todos están queriendo analizar datos.

De repente ¿no te queda la impresión de que la comunidad de Bases de Datos se ha quedado un poquito atrás en eso?

No, pues es difícil distinguir entre las diferentes comunidades, muy diversas, que hacen Bases de Datos: machine learning, data mining, databases (clásicas), Web, etc. Si vas por ejemplo a una conferencia de KDD (Knowledge and Data Discovery) que es una de las principales en esa área, hay gente de Base de Datos y de Inteligencia Artificial que están trabajando ahí. Entonces no creo que nos estemos quedando atrás.

A veces es difícil identificar cuál es la comunidad de Bases de Datos, porque está desperdigada en diferentes subdisciplinas.

Hace años la gente nos decía “es que la gente de la comunidad de Bases de Datos se quedó atrás con lo de la Web ¿verdad?” No nos quedamos atrás, es que hay otras disciplinas que están tomando esos trabajos. Pero un poco para presumir ¿de dónde salió Google? La compañía Google salió de un grupo de Base de Datos, por ejemplo. Así que no nos quedamos tan atrás como comunidad.

III. Políticas científicas

Vamos a hablar de políticas científicas. Me gustaría conocer tu opinión, lo pregunto particularmente sobre la computación chilena, sobre esta relación entre la ciencia y la computación como disciplina nueva, y el respeto -o la falta de respeto- con que otras disciplinas la miran en términos de evaluación, publicaciones científicas y estas cosas. ¿Sientes que la Ciencia de la Computación se ha ganado el respeto dentro de las otras ciencias clásicas?

Pues poco a poco. Pero ese problema no es exclusivamente chileno. Fue un problema en Estados Unidos por muchos años, en universidades batallaba la gente en computación para salir adelante en las promociones, porque ocurría el mismo problema: los que estaban tomando las decisiones, los decanos, los rectores de universidades veían por ejemplo el número de publicaciones en journals o revistas y en computación no existe esa tradición porque estamos en conferencias, en sitios “más informales” según ellos.

¿Cómo ganaron esa pelea en lugares como Stanford, porque esa pelea aquí en Chile todavía la tenemos?

No sé cuál es el secreto, pero en algunas universidades se empezaron a tomar más factores, por ejemplo, la opinión de la gente importante, de los que han tenido más influencia en el campo, más que el número de publicaciones o dónde han publicado. Creo que las universidades donde

han surgido mejores departamentos de computación es donde hay administradores que aprecian, entienden el campo y saben evaluar las contribuciones e interpretarlas. Por ejemplo en Stanford para promover a alguien el factor más importante son las cartas de evaluación, se piden como a quince personas, especialistas del campo, que evalúen el candidato y eso cuenta más que el número de publicaciones. Sí vemos el número de publicaciones y en qué sitio son, pero es más importante ese otro tipo de evaluación. Creo que el resto de las universidades está cambiando, tendiendo a evaluar la computación en forma diferente.

Hay otro tema que empieza a aparecer en la academia y es esta relación en que el científico más independiente, más puro, se está mezclando mucho con la aplicación y de alguna manera hay un tema económico que tiene que ver con la innovación, con la ligazón con la empresa; aquí alguna gente reclama que esto le hace perder el rol al académico clásico, ¿qué piensas de ese fenómeno?

Me parece que es bueno, como dije inicialmente, tener nexos con el mundo real porque los problemas que generan son interesantes y es la clave para tener impacto, que es una palabra que se usa mucho en las evaluaciones en Stanford y otras universidades. Entonces es más fácil tener impacto si uno está trabajando en problemas reales, donde hay gente esperando la solución. Ahora no es bueno irse al extremo y estar trabajando demasiado solamente en problemas que van a servirle a la industria, porque ésta muchas veces quiere soluciones inmediatas, problemas a corto plazo. Lo que tienen que hacer los académicos es tener los problemas, escoger cuáles son los a más largo plazo en qué pueden contribuir los académicos y no preocuparse tanto por los a corto plazo, porque muchas veces las compañías quieren software, quieren algo inmediato

y muchas veces hay que decirles que no, esa es la clave para poder identificar los problemas que vale la pena atacar.

Y por otro lado, ¿qué opinas de esta competencia casi demencial por el paper que existe hoy? ¿A lo largo de tu carrera notas que ha cambiado, que hay más énfasis en la publicación?

Sí, ha cambiado bastante en los ramos de computación la presión por publicar más y más artículos. Lo veo, por ejemplo, en el número de publicaciones de los que están solicitando empleo. Hace años los currículums llegaban con dos o tres publicaciones y era alguien muy bueno, ahora si no tiene diez o veinte publicaciones alguien que se está recibiendo entonces no es tan bueno.

¿Es parte del fenómeno que estamos viviendo o ves alguna salida a esto?

La verdad no sé cuál es la solución, porque sí hay demasiada presión, demasiadas publicaciones. A mis alumnos trato de decirles que lo importante no es el número de publicaciones que deben tener si no que la gente lea, que tengan impacto, no nada más tratar de sacar un montón de artículos, mejor ir más lento pero tratar de sacar algo que valga la pena. Pero es difícil para ellos aceptar mis sugerencias porque ven que todo el mundo está publicando muchas cosas, entonces a veces les tengo que decir que no manden ese artículo, que no está listo, va a ser contraproducente, que si se lo publican es peor porque no va a ser bueno y a lo mejor van a agarrar mala fama.

¿Cómo ves desde Estados Unidos cómo ha evolucionado la computación en Latinoamérica?

Mis comentarios son más sobre México porque conozco más, no es tanto de Chile en particular, y creo que se aplica a Latinoamérica, pero no estoy seguro.

Parte del problema creo que es que en México, o en parte de Latinoamérica, no hay una cultura de apreciar los estudios de graduados y las ciencias. En México hay mucho más énfasis en tener un trabajo y ganar dinero, más que en tener una trayectoria de investigador, de científico. Por ejemplo, veo en Stanford que nos llegan solicitudes para el Programa de Doctorado de todas partes del mundo y son los mejores estudiantes del mundo que están solicitando, pero hay patrones muy marcados: de ciertos países nos llegan muchas solicitudes y de otros muy pocas, en general de Latinoamérica llegan muy pocas y he chequeado con mi colegas en otras universidades similares y también hay muy pocos solicitantes ¿y por qué?, cada vez que voy a México o a otro país de Latinoamérica trato de hablar con los estudiantes de por qué no quieren estudiar un Doctorado y me preguntan “para qué, para qué voy a estudiar un Doctorado si no me voy a hacer rico, si luego regreso a mi país y no hay puestos de trabajo en el área”, así es que no tienen interés. Por otro lado mis colegas en la Escuela de Negocios de Stanford tienen muchas solicitudes de Latinoamericanos. Todos los latinoamericanos aparentemente quieren sacar un máster en administración de negocios, no quieren estudiar computación o ciencias. Estoy exagerando, pero sí es una tendencia.

Tengo la impresión de que aquí en Chile hay un poco más de aprecio a la academia que en otras partes de Latinoamérica, así es que están mejor que en otras partes. Es lo que he visto.

¿Qué mensaje le darías a los profesores y a los investigadores en Ciencia de la Computación en Chile, en dos temas: en el desarrollo de área y en cómo contribuir a la computación desde Latinoamérica?

Hoy en día hay muchas oportunidades de contribuir fuera de los centros principales, porque a través de las redes y de la Web, la información está accesible, los artículos

que ustedes pueden buscar en Chile son los mismos que puedo buscar yo en la Web, así que se ha emparejado mucho más el campo, es igual de difícil o fácil trabajar aquí que en otras partes. Ese es un punto importante.

Como sugerencia, tratar de buscar problemas interesantes; ir a conferencias en otras partes del mundo; ir de visita a otras universidades; si tienen oportunidad pasar tiempo en Europa o en Estados Unidos; no preocuparse nada más por publicar muchos artículos sino tratar de tomarlo un poquito más lento y desarrollar las ideas que van a tener gran impacto, porque antes de lanzarse a escribir un artículo les digo a los alumnos que vamos desarrollar tres o cuatro ideas de posibles artículos, explorarlas y de esas cuatro escoger cuál suena más interesante, pero no irse nada más con la primera que se les ocurra.

En cuanto a trabajar solo o no, depende mucho del área y del estilo de cada quien, pero aún si uno trabaja solo es bueno tener colegas con los que se pueda conversar, aunque estén trabajando cosas diferentes, pero con mis alumnos aprendemos mucho cuando tratamos de explicar una idea, el sólo hecho de explicarla y tratar de contestar las preguntas muchas veces le hace ver a uno que a lo mejor esto ya no es tan bueno o hay un problema que aún no se había visto. Creo que es muy importante estar constantemente intercambiando ideas con otros aunque el trabajo sea de uno y los otros estén nada más dando opiniones. O aunque no hagan nada los otros son como el siquiatra: se sienta ahí y deja al paciente hablar. La ventaja es que uno al hablar y explicar entiende mejor las cosas, por eso es muy importante tener colegas que lo escuchen.

¿Qué consejo darías a los estudiantes que están iniciándose en la computación?

Que busquen problemas que los apasionen. Trabajar en un problema que los apasione, que les interese, es lo más importante. BITS