Timos Sellis Speaks Out on Research in Australia and Greece

Marianne Winslett and Vanessa Braganholo

Timos Sellis

https://www.swinburne.edu.au/science-engineering-technology/staff/profile/index.php?id=tsellis

Welcome to ACM SIGMOD Record's series of interviews with distinguished members of the database community. I'm Marianne Winslett, and today we are in Snowbird, Utah, USA, site of the 2014 SIGMOD and PODS conference. I have here with me Timos Sellis, who is a professor at the Royal Melbourne Institute of Technology[1]. Before that, he was at the National Technical University of Athens and the University of Maryland. He is an ACM Fellow and an IEEE Fellow, and he has a VLDB 10-Year Paper Award. His PhD is from Berkeley.

---

[1] Timos Sellis is currently the Director of Swinburne Data Science Research Institute at the Swinburne University of Technology in Australia.

So, Timos, welcome!

Thank you, Marianne. It is very nice to meet you.

Your VLDB 10-Year Paper Award was for introducing R+ trees in 1987. Is multidimensional indexing a solved problem now?

Well, to tell you the truth, multidimensional indexing has been used mostly for indexing spatial objects where the number of dimensions was, say, three at most. So what became interesting was that after a while, people started using multidimensional indexing to index higher dimensionality objects. There we could see that R-trees and all the variations of R-trees were not enough. So, it's not surprising that people are still working on multidimensional indexing.

I actually see people working on different platforms instead of following the standard hierarchical indexing methods we've been using in the past. They've tried to accommodate over MapReduce and distributed file systems. So, the interesting issue is scalability. The multidimensional indexing structures we have built in the past do not scale easily to a large number of dimensions. So, to some extent, it has been solved for the standard types of applications which have been seen in the past. But the fact is that we are getting more applications where you have what we call feature vectors with thousands of dimensions, right? Then apparently you cannot use these kinds of indexing methods anymore. So (a) you have to reduce the number of dimensions, so we have all these problems of dimensionality reduction and (b) you have to think of techniques that scale very well in thousands of dimensions. So, in my opinion, it's not a solved problem.

I think the more we see new interesting applications coming up (like from social networking applications), we'll see more work on multidimensional indexing. What is also different are the types of queries that people ask on these multidimensional indexes: they are no longer the standard type of range queries we have seen in the past. So, they want to do more complex kinds of analysis on these indexes, which makes it quite interesting in my opinion.

I think computer vision apps will be the next big thing for our field and they have very challenging index requirements, so I think there will be a little renaissance of work in this area.

To tell you the truth the most interesting applications I have seen up to now is where you have thousands of queries coming at the same time with thousands of updates. So, anything around location-based kinds of services or applications have this characteristic, where you have lots of updates coming in from the moving objects, yet at the same time, you have all these continuous queries that have to access the same indexes. So, coming up with indexes that can scale to thousands of updates and thousands of queries at the same time is an interesting problem.

Is there a gold standard for indexing for those kinds of apps?

You know we always thought that the hierarchical indexing methods (because of the *log n* kind of performance) were the solution. Our experience in the last few years that we've been working on these location-based applications is that you need the combination between the hierarchal indexing and grid or cell partitioning techniques. So, the old grid file that we used to have in spatial databases, combined with hierarchical indexes is something that we see more often coming out now. I think it's no longer a game of having, as you've said, one kind of hierarchical B-tree extension that will do it all. You need to combine approximate kinds of representations for the data because you cannot afford to index everything with all the details. So, it's an interesting landscape, I would say. That's why I'm excited for my Ph.D. students in Australia now we're coming back to look at similar issues around multidimensional indexing in different contexts, for example, in tweet analytics. Again, we can say for sure that the techniques we had in the past are not appropriate.

Someone was commenting to me today that researchers from Europe and Australia don't have as many papers in top conferences in our field as their counterparts in the US do. Why do you think that is?

So, I've been in Australia only for 1½ years, so I can tell you what my impression is on this issue. Australia is far away from everywhere. So, to send a student to a conference, or even a professor to attend a conference, that is very expensive. Funding is not that bad in Australia, but it is not enough to cover expenses to go to many conferences. So, I've noticed that people are more directed towards publishing in journals rather than in conferences. I think it's (a) financial, and (b) there is a culture in Australia that conferences do not count that much for promotion in tenure. Even in computer science and we know this discussion has been going on for years now. In Australia, coming from the Anglo-Saxon kind of system, they give more emphasis to journals. So, I'm surprised on how many journal publications I've seen from several groups in Australia, but I must confess that some of these people I didn't know because I wouldn't

see them in conferences. So, it's only if this particular work they did was of interest to me, I would find it in a journal, but it's not the people that we meet very often in conferences. I think that's the major reason. Now that I've seen that, I've told my students that I don't want to follow this principle. I would like them to send papers to conferences. You know, I see they have no problem sending papers and they do very good work. So, there are some very good groups in Australia. It's not that they are hesitating to submit papers to good conferences. I feel that because of the financial issue, they don't do it as much as in the US.

> **I feel that my biggest accomplishment is the fact that we have generated a very good new generation of Greeks in databases...**

What about Europe versus the US?

Well, Europe you see, we have a lot of presence from European researchers in conferences, right? So, I don't think with Europe there is a problem of the same kind, especially since there is a lot of money going into research because of the European Union funded projects. So, I don't feel people are restricted to submit papers to conferences. I think in our community we see representatives from all over Europe, so it's not the same as with Australia. Now the difference that I see, for example, are the people coming from China. That's a huge difference compared to some years ago. I know very well that China is investing a lot in research. So many students have funding to come to conferences. This is not the case in Australia.

From the years that you spent in Greece what accomplishment are you most proud of?

So, I spent 20 years in Greece. That is a long time. I must confess that what I feel is my largest accomplishment is the fact that I helped create a culture in our database field. So, I'm very happy to see so many faces that have gone through my laboratory to do what you call in the US an "Honors Thesis" (the thesis that students do finishing their Undergraduate degree). I'm very happy that many of these students are now professors elsewhere or they are either doing their Ph.D.'s now. So, I feel that my biggest accomplishment is the fact that we have generated a very good new generation of Greeks in databases and of course after a while (I moved in '91 a few years after Yannis Ioannidis

and later on to recent years like Minos Garofalakis, Antonis Deligiannakis many well-known Greek researchers came back). So I'm very happy to see this going on and I'm very happy to see new faces. I'm always proud of myself that we managed to create the environment that can generate international level work in Greece.

So it sounds like the pipeline is in place and operating.

It is.

Well, a related question. What has been the impact of the financial crisis on the research environment in Greece?

This is an interesting question because unlike other areas in Greece, research has not been influenced too much by the financial crisis. I'm sorry to say that the reason was that the government never funded research at a significant level. All the money was basically coming from projects that we were getting from the European Union. Even if it was from projects that were coming from Greek ministries, the money would come to these ministries from Europe. So once this hose is still open, the money comes in. I would say that the community has not been influenced too much.

Of course, what I have seen as a difference is that many students would not stay to study in Greece anymore. For a Ph.D., for example, they would prefer to go abroad just because they have more opportunities once they get through their Ph.D. It is easier to get a job abroad than getting jobs in Greece. For example, very few academic positions are now available in Greece and very limited other positions for people with Ph.D. or Masters degrees are available in Greece. So, to some extent, research has not been influenced in reality, so you will see that the work that gets published at international conferences are pretty much similar. We get the same number of students that go for a Ph.D., but the only difference I would say is that I would expect that in a few years we might see the numbers dropping just because not too many students would like to continue for a Ph.D. in Greece.

While I was working in Singapore in the last few years, I visited Australia a few times to recruit summer interns for our Research Institute. But I didn't get any takers. From that, I conclude that Australian's level of interest in research is lower than in the US. Although maybe there are other factors I don't know about. What is the research environment and culture like in Australia?

It's interesting that you didn't find any takers. I would imagine that you would not find any takers from native

Australians. So, what happens is that the majority of the graduate students that I see in Australia are international students. Australians, somehow, don't feel the need to go for a Ph.D. Some of them go for a Masters. Many of them just go directly to the job market. The unemployment rate is something like less than 6%, maybe 5%, so there are jobs. People don't feel the need to study any further to get jobs.

Given that most of the graduate students are international students, it would be very rare for these international students to leave Australia once they came, to go, for example, to the US. One thing that I've seen in the last couple of years that I've been there is that because the Australian dollar was kind of high, many students would not come any more to Australia, they would prefer countries where life is cheaper. That is why I can see that the number of students coming from China, for example, is declining. We have lots of students from India, Bangladesh, Iran, various places in the area, but I would suspect that the people from China, just because most of them come with scholarships from the government, they would prefer to go to the US because it's cheaper. So, I wouldn't say that Australians not interested in research. They are interested. Most of them prefer to go to the job market, get some experience first to see what the job market looks like, and then they may come back to do a Masters and perhaps a Ph.D., but it's not the typical thing I would see, for example, in Greece or in other countries where you would go immediately after your bachelors for a Masters and a Ph.D.

Is there a startup culture there?

Yes, there is very much a startup culture. I've seen more and more money from the government going to help startups. I've seen startups established in Australia come to the Silicon Valley and there are some very good examples. I think the culture is there, but I don't see that much in RMIT with our students in our undergraduate program. I cannot sense it that much, and that's the reason we decided to add entrepreneurship types of courses in our curriculum to help the students towards this direction. But it seems that the Australians do have this culture of taking the risk and getting something started up. I've seen many young people doing it.

Do you have any words of advice for fledging or mid-career database researchers?

I think what I will say will sound very traditional. My advice to both my students and as you say, researchers because I was lucky for the last six years to fund and run a new research institute in Athens in information systems and data management. So, I had the opportunity

to help people come to our institute from fresh PhDs to mid-career people. I'm always trying to push the same thing that all of us are saying. Do something that you like. Don't follow necessarily the trend in terms of what is funded or don't hesitate. For example, I was telling them, "Come to me, I will support your research", to get some prototype out, for example. Then we can go for funding. Don't try to change your research to be closer to something that is being funded. Instead, do what you can do best. I can help you find the leads to an application area. We have this experience after so many years in the field. My advice would be to not change their career based on what seems to be fashionable.

This brings me a bit to this Big Data area. Somehow Australians are trying to advertise me as a Big Data person. I don't like that. Wherever I sit in a panel with various CEOs or CTOs or whatever, I usually make this introduction that Big Data is not a term that I would like for us in data management to cover ourselves under. I'm aware it's something that everyone understands when you talk about Big Data, but I'm just telling them that we were always dealing with Big Data.

> *Don't try to change your research to be closer to something that is being funded. Instead, do what you can do best.*

True, we always were.

I don't want them to think that Big Data is something necessarily new. It's a new setting. It's an opening to many other areas. It's very interesting to work with data from various other areas, but I wouldn't like them to think of Big Data as the revolution that is happening nowadays.

Why is that bad? If they think it's a revolution, then they can get all excited about funding it for example or supporting it.

So, the reason I'm saying that is because I see that with many people, when they think about Big Data, they have the impression that it is something totally new and that we expect to hear something new. Of course, it has become almost a synonym for analytics, which is not false and I wouldn't say that is not the case. Most people think of Big Data in terms of data analytics. What I've seen is that people, at least in Australia, have seen this wave becoming large over there. The government,

funding, everybody talks about Big Data over there. It's good for us and I enjoy it, I must say. On the other hand, I don't want them to think that it is something totally new. They expect to hear new buzzwords. Personally, I find that difficult. I would like people to understand what we are doing and how this is related to Big Data or what they think is Big Data, rather than having to reinvent new kinds of terms or even methods to convince them that what we are doing is something totally new.

Anyway, Australia, I think at this time, is a good place for research in our area. That is the reason why I accepted to run SIGMOD there in 2015. So, I think it's good for our community to get the exposure in this country. It's not a coincidence that CIKM will also be in Melbourne next November. So, we will have both SIGMOD in June and CIKM in November. It's going to be a very interesting setting for people to come to Australia.

> **Big Data is not a term that I would like for us in data management to cover ourselves under [...] we were always dealing with Big Data.**

Among all your past research do you have a favorite piece of work?

I think the work that we did in the last 6 years with one of my Ph.D. students (Kostas Patroumpas) which is around what we call positional streaming data which means coordinate data streaming in, no matter what the application is… (like XY coordinate data streaming). I think I have enjoyed the work in this area because it's the first time in my life where we've started from data models that you need, to query languages, to indexing, all the way down to the applications. I enjoyed the fact that, at least for me, it was the first time I addressed a data management area and I saw the whole thing from its theory, like what kind of algebra and operators we need to run in these kinds of applications, all the way down to building systems. I think it's probably the longest engagement I have had in my career. I've gone through various areas like query optimization, spatial

data, data warehousing, personalization, but with this one, I liked how the idea evolved from geometry problems to query processing, to indexing, etc. So, I think I'm very happy that we managed to do work in this area for like 5 or 6 years, which has resulted in quite some interesting outcomes.

There are two papers that I think are quite interesting. Both of them appeared in SSDBM. The first one[2] has to do with techniques for positional steaming data where you get trajectories, and you would like to compress them online. So, imagine that you have all the cars and you want to record all trajectories. This is too much information. So, we looked at various techniques like, for example, if I can predict the location where you will be after a few seconds, then I don't need to store it. So, what are the samples that I take that I store, or I drop them because I can predict them?

The other one was in the same paper of what we call amnesic compression where you would like to have an online method where, as the data comes in, you approximate the past with enough information so that you have an idea what the trajectory was like, but you keep all the details for the present. So, the present is very accurate. As you go to the past, you get less detail, that's why we call it amnesic. This is work that I find interesting because you don't find equivalents for other types of data. Trajectory data brings opportunities for very interesting problems. So, compression was one.

The second one[3] is the one we got the Best Paper Award in SSDBM 2012 that was about privacy issues. So, with position streams, if you wish to instead of sending your actual location, to send a blurred kind of area where you're in, can you still answer interesting continuous queries? If I want to continuously know how many of my friends are around me or who is around me within a distance of say, 1 kilometer, where they don't send their actual locations, but they blur them a little bit and send me an area where they are, can I still answer this query with some probability? For example, I would like to say that 5 of your friends are in this area with a probability of 75%. We wanted to show that if you want to compromise, to some extent, accuracy but to gain privacy, you could do it with these methods. Again, these problems are interesting because you have the issue where the queries move with the users, and the objects move also move. You need to come up with these kinds results fast but at the same time, we wanted to show that even under uncertain situations, you can still answer these queries. So, I think that out of my

[2] Michalis Potamias, Kostas Patroumpas, Timos K. Sellis: Sampling Trajectory Streams with Spatiotemporal Criteria. SSDBM 2006: 275-284.

[3] Kostas Patroumpas, Marios Papamichalis, Timos K. Sellis: Probabilistic Range Monitoring of Streaming Uncertain Positions in GeoSocial Networks. SSDBM 2012: 20-37.

work, although I said we have done several things, I would pick these two as interesting.

If you magically had enough extra time to do one additional thing at work that you're not doing now what would it be?

I can tell you now that I've reached my mid-50s. What have I missed? I have missed working with systems people. When I was in Berkeley, I enjoyed working with Michael Stonebreaker. He was my advisor. I could see a person who understood and actually wanted to build systems out of every idea that came out. I think I missed that. I missed it in the course of my work in Greece. I see this opportunity now and I try to take it with the people now in Australia, to work with systems people (so, people who understand systems issues very well).

The second thing is the multidisciplinary environment. So, in Greece for the first time, after 20 years in my career, I started working with biologists and we worked for about 6 years together. I enjoyed that very much. So, picking an area where we can contribute with our knowledge from data management… I like that very much. With systems people, I want to work with them because I want them to influence my thinking about my problems in data management, but with another science kind of area, I would like to work with them, so we can influence this area as much as we can.

If you could change one thing about yourself as a computer science researcher, what would it be?

As a computer science researcher, I think I would pick the same answer as with the previous one. I would love to have emphasized more on systems issues. You know, I'm an electrical engineer by training, so my degree in the National Technical University of Athens in '82 was electrical engineering just because there was no computer science back then in Greece. If you're an engineer, you carry this continuous kind of interest on how things work. If I could have found a way to combine more engineering work or more systems work with the things I have been doing in the past, I think I would have enjoyed that very much.

Thank you very much for talking with me today.

Thank you, Marianne.