Tova Milo Speaks Out

on the Impact and Importance of Database Theory, Web Mash-ups, Open Source Software, Supervizing PhD students, Feminism in Science and More

by Marianne Winslett



http://www www.cs.tau.ac.il/~milo/

Welcome to this installment of ACM SIGMOD Record's series of interviews with distinguished members of the database community. I'm Marianne Winslett, and today I am at the 2010 SIGMOD and PODS conference in Indianapolis. I have here with me Tova Milo, who is a professor of computer science at Tel Aviv University. Tova and her coauthors are the recipient of the Alberto Mendelzon PODS 10 Year Paper Award for 2010. Her PhD is from the Hebrew University of Jerusalem. So, Tova, welcome!

What do you view as the impact of database theory in recent years?

I think the cleaning of information, the cleaning of concepts, cleaning of models, cleaning of ideas, and things like that. So it is the development of simple models, simple algorithms, checking what is possible and what is not possible, and so on, trying to have it clean.

Ok, clean and tidy. So, maybe the really practical people in our community would say, "Why do we need it to be clean as long as it works?"

I think that when you have a clean view of what the problem is, a clean model, know exactly what is possible, and what is not possible, you are going to build a better system. So it is true that even without all this you can make things work. But I think you make them better if you actually understand what the fundamental issues in what you are doing are.

So, transactions would be a good example of that.

Transaction are a good example, probabilistic information is another kind of example. When you know what you can do, and what the mathematical foundation below asking a query on unknown information or partial information is, the results will be sounder.

Those are good examples, do you have anything else?

Top-K query evaluation, for instance. There are these beautiful results, of Fagin and Naor, for example, on optimal algorithms for *top-K*. This is theory, but actually, it has a lot of impact in practice. And another area I can think about is data exchange. There is a bunch of really nice results that tell you how you can translate data from one form to another while really representing the information in a sound and complete way. I think it is important that if you build your system this way, it will really represent what you want. So, right, you can have whatever, but I think what you get will be less desirable.

What should database people know about business processes?

Well, everything that they don't know! I think database people mostly focus on data. You have the database, the data queries, indexes, you manage it, and so on. Often they forget that these databases, the usage of databases, is just a small piece in a big long business process where people interact with the system, they plug information from one side to the other, they consult with the data, they consult with other people, and so on. Once you realize that, and once you try to take that into consideration, you want to give them a much better system to play with. So there is tons of data there besides the data in the database, there is information about the business rules, there is information about where the data flows, what happens to it, what kind of interaction people do with it, and you can exploit it much better if you model it and if you allow to query this kind of stuff.

So the research challenges there, are they more about how to mine this kind of data, or are they some other aspect?

I think there are many types of research challenges. One big question is that of modeling and inference, because you have to somehow model in a reasonable way the processing that takes place, the interaction with information, the interactive nature of the whole process of data management. So modeling is one thing. Trying to infer things from that, like "what is the most likely thing that people will do once they get this kind of message", or "once they got this kind of query result from the database", as so on. You can use that to get more information, to recommend to people things that simplify the processing and so on.

Do web mash-ups have a theoretical foundation?

No, but they should have! Again, the construction of web mashups involves many aspects. It is using pieces of software, so there is the data that is behind these pieces of software, you need to put them together, you need an interface, and so on. People mostly focus, at least in our community, on the management of data, the data that comes from the different things, and needs to be mashed-up. But there are many different aspects like the creation of the mash-ups, and then the usage of these mash-ups. And

there are tons of crowd wisdom that one can use, because when you are creating the mash-up that uses a list of restaurants and a map, you are probably not the first person to create the mashup that uses a list of something and some map. So if you wire information in a similar way, or learn from what they tried and didn't work, you can do stuff much better. There is tons of information there on the web on how people constructed mash-ups, what they tried before, what kind of interface they found to do things. This is all data, so you can query it, you can use it, you can analyze it, you can put it together, to create better mash-ups.

Is that information in blogs, or is it from watching people who are trying to make the mash-ups?

Blogs are one sort of information, but also, it's the mash-ups that actually exist. There are mash-ups that are available publicly on the web, you can look at them, you can see what kind of data sources they used, you can see how the information flows from one mash-up to the other when you click, what happens there, what kind of wires people are using, and then do something similar, or say okay, I don't like that, I want to do something else. But you don't have to invent the fact that the address or the zip code of the restaurant is the address in the maps, someone did that, you just reuse and that is it.

The way you describe it, it seems very user oriented, close to the user.

I think it is close to the user, and also close to the developer. I think, again, that what people mostly focus on, in our community, is the data that runs through the mashup, but there are the two other aspects: how you construct this mash-up - how do you know which pieces of information to use and how to wire them together, and then, for the user, how to use this thing - what is the best way to get the address of the good restaurant that is close to your place? should you click here, and then it will show here, or should you pose query here and then…So there is a lot of interaction between the things, and there are things that bring better results than others.

What should database people know about managing open source software?

What we try often to do is to put out pieces of software on the web so that people can use them and so on. But there is also a lot of information about the dynamics of software (which is often ignored). For example, how do people improve UNIX packages - how do all these millions of programmers all over the world manage to bring together a coherent piece of software. And there is a lot of interaction there, there is a lot of logic, there is the flow of information from one person to the other, and I think there is no good data management support for these people. It works, but there is really need for infrastructure, so that people can understand better where are the problems, who's the person in charge of deciding whether this is going into the open-source package or not. This is not like in a software company where the hierarchy is known. This is kind of "cult" thing, and there are these people, these "mysterious" people whose names you don't even know, that decide. There is a lot of knowledge out there, about how things work. There is a need for support for this kind of information management.

So does it become a mining problem to make explicit the information that is kind of there but hidden?

I think it's a lot of distributed data management, because you have replicas of software and different versions, and different developers in different parts of the world, and there is some sort of relationship between them. But it's not the standard consistency that one may think of when you think of a database. So I think it is more a type of a weird distributed data management that one should look into. We worked with a European project that does this kind of open source development and tries to make coherent Linux editions, and it is really unbelievable how they manage all this information. It is weird, it's by e-mail messages, basically, and some forums and things like that. But it's definitely not done in a comprehensive manner.

So there is no concept of an overall version tree even if it is distributed?

Yeah, it is a very loose notion of that, and they live with that, but it is very difficult.

So, as database people, what angle should we be looking at?

I think we should talk to these people. Try to understand what the problems are, and where we can help. Sometimes the things they come up with are very close to what we (database people) do, but there are also many other interesting problems that can come out of there and where we can help.

Do you have suggestions for increasing the impact of database theory in the future? (could be particular topics, or styles, or actions…)

I think the first thing is public relations. We should make people understand that what we do is relevant, because it is, so I believe. And I think database practitioners meet theory with a mixture of fear and disrespect. Sometimes they think that what we do is irrelevant even though it's very relevant. I think really it's a lot a matter of public relations.

The PODS Executive Committee can take on that challenge!

Yeah, they should hire a couple of good PR people!

I've been told that you are an excellent advisor for students. Can you give us tips for being a good advisor?

So first, it is a great compliment. So, I don't know if I can give tips, I think it really depends on your personality. I treat my students like they were my children. So I do everything for them. From helping them draft their papers, sometimes writing the paper for them, sometimes solving for them the problems they are trying to solve, taking them out dancing when going to a conference, everything. And in a way, one may think that if your advisor does everything for you, then you don't have to work yourself, right? But it is a bit like children - if you are totally devoted (to them), they become totally devoted back. So they do everything to make you happy, they do everything to make you proud of them. And I'm really proud of them. So I think it that if you really invest a lot in them, you also get back a lot – like with children. I like them and they like me back.

I want to ask a follow-up question to that. I hadn't heard people talk that way before, so I have to think, so if it were your family and your kids you were talking about there, then with this close bond, sometimes it is hard for the kid to separate, or for the adult to want to separate when it is time to go off and start their career or whatever, so how does that analogy apply to grad students?

I think (that) also for kids, at some point you have to let go. You have to push them (away), so I do the same with my students. Slowly, somehow, I push them away.

Gradually, or when they finish their PhD?

I think that in the last year of their PhD I push them slowly away. But because, generally, I cannot work with people I do not like - all the people with whom I wrote papers, or at least more than one paper are my good friends, it is the same for my students - if I do not like them, I cannot work with them. So I like them, and when I push them away, in the sense that I make them work without other people and don't

write a paper for them anymore, we still stay good friends. So again, it is like with kids, they go away, but they are still your kids.

Many other places would like to emulate Israel's success in creating a vibrant startup culture. What are the key ingredients in making that happen?

They should send high school graduates to the Army before they go to the university! :-)
Well, "Israel is a small country surrounded by enemies" - that is how think about ourselves - so everybody goes to the army. You may think about it as a waste of time, on one hand, but on the other hand, it builds character. Because you have to subdue, you are 18, and you are sent away from home, and you have to manage for 2 or 3 years by yourself, and you are not very comfortable, safe. This builds character. So eventually, when you come out of that, you are very independent, you know what you want, you know how to manage, you don't have to be spoon-fed. So this is important for studying, so when we teach students we tell them, OK, you have to study that, and they manage, even if you don't give them the material in class, they manage somehow, they don't come to ask where should I take this material from, they manage. The same holds for research, and the same for industry. Somehow this independence builds creativity, and somehow, "I can do it". That impacts both industry and research.

So going back to the advising thing, and treating the students like your kids, it sounds like these are not your typical kids, worldwide, I mean.

Yeah, I agree. I think Israeli students, undergrad and graduate students, are much more mature.

You say you do everything for them, but then you say they are so mature that they don't need the hand holding that, for example, a US undergraduate might expect.

It is a different kind of ``doing everything for them''. Pushing them to get scholarships, pushing them to meet people. (For instance) when I go to a conference with students, in 5 minutes, they know all the important people in the area, they are invited for dinner with us. It is a different kind of holding hands.

The students who went to the army at 18, by the time they are maybe 23 or 24 and are grad students, is it more equalized in their life than grad students in other countries, or are they still different at that stage?

I think they are still different. I think it is more independence than just aging. It is really surviving difficult situations, managing no matter what. (The feeling that) nothing is too difficult and you can always manage. I think this is an Israeli thing, in general, this survival thing. And also being optimistic. As an Israeli, if you think of all the bad things that may happen to your country, to your life, unless you are totally optimistic and believe that everything will be good, eventually, there is no way to survive. And I think that for research, for instance, and also for start-ups, this (optimism) is a great thing. You write a paper that gets rejected, it makes you very depressed, but you can also think (optimistically), okay, it will get better to a different place, better place, more reviewers will read it, etc. So there are ways to think optimistically about it, and this helps.

Is there such a thing as a mini-skirted feminist?

Well, at least one!

So what is it? I guess some people might claim that is a contradiction in terms.

I don't know why, actually. I love shopping, I go with mini-skirts, yet I am a feminist, never-the-less. I think it's more of a question of: ``can you survive here in research or whatever while being yourself''. Do

you have to be like the men? Do you have to put in your papers examples about basketball players, or can you put something different that you actually care about. I think that, in a way, it is easier if you mellow, if you go with the crowd, if you dress non-provocatively, if you look serious, if you don't talk about shopping. I will makes things is easier - people will think that you're serious. Yet, I think you can survive - I survived - without that. And it is okay. I think the more (women) will be like that, the easier it will be for the rest. You can do both, it is not difficult.

I certainly think about shopping, but I wouldn't talk about it with my male colleagues, because all they have to do is buy a pair of khaki pants a blue shirt and they are set, so I don't think they understand the challenge of finding that perfect whatever.

Well, I don't understand anything about basketball, soccer, but still they talk with me about it! So I can talk with them about whatever I want to, and if they don't want to, then they can go.

Is it true that you were the boss of the person who later became your advisor?

Yes, I was a captain in the Army, I was 4 years in the Army, and in Israel, men, after they finish the Army, they still do one month per year as an Army reserve. So Catriel Beeri, that later became my advisor, used to do his military service in the group that I managed. That is how I met him, actually. I told him about the project we were doing there, from there I started my Master and then my PhD. So the common joke is that in order not to clean the toilets, he wrote my PhD thesis!

Do you have any words of advice for fledgling or midcareer database researchers or practitioners?

I think the usual, do what you like, enjoy, have fun, and then the rest comes.

Do you have any words of advice specifically for young women?

The same, do what you think is fun, don't think about what people will think of you, just do what you want.

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

Actually, it is the paper for which I won the prize this year. I think it is very cool. This of course is very objective ☺

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

I don't think I want more time to do work; I want to go shopping, I want to go to the beach, I want to play with my kids, absolutely not work more.

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

I would read more. I am lazy, I don't read. That is why I often choose to work on new things that nobody worked on before (so that I don't have to read about it). But reading is very important. I should change that.

So, you mentioned your kids, do you have any tips for us on how to maintain the work and family life balance?

You should marry the right husband, that is the obvious! I think I was very blessed with the men in my life, from my father, my first boyfriend, my husband, my colleagues, I think I owe them a lot for everything, tons of support... So you just have to find that. And marrying young also, that is a good thing.

How does marrying young help?

I think that when you marry young, you sort of grow together. So if you are advancing in your career, it's more easily acceptable, somehow, than if someone comes and has to accept at an older age that his wife is a workaholic, that she travels a lot, or whatever. Somehow you grow into that, and that makes life - the splitting of duties, and so on - much more natural. But that (marrying young) is not very common these days. Maybe I was lucky.

Thank you very much for talking with me today!

Well, thank you very much for interviewing me.