

Publication: Asian Scientist Magazine
Date: 12 January 2015
Headline: Dealing with the data deluge

Dealing with the data deluge

SMU Associate Professor Zheng Baihua tackles the challenge of integrating multiple sources of data to help users find relevant information quickly and efficiently.

By Asian Scientist Magazine
12 January 2015

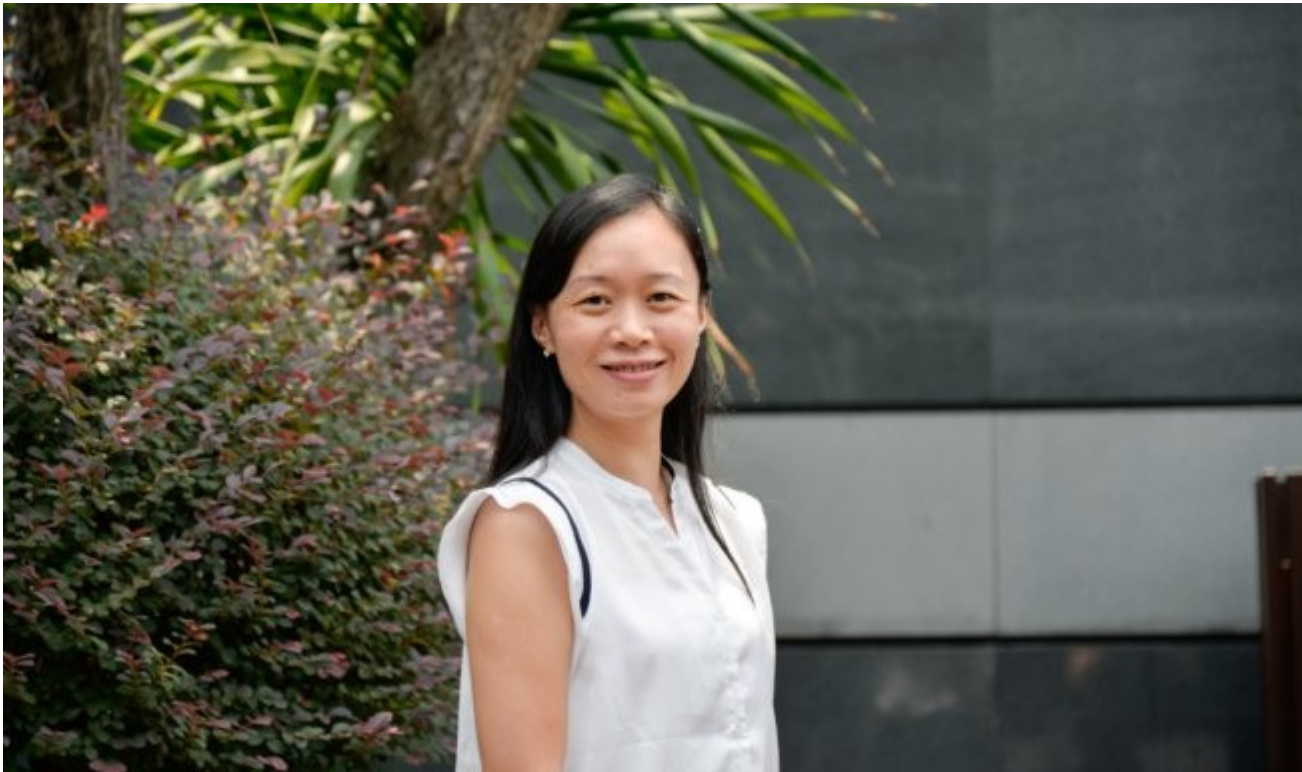


Photo: Cyril Ng

AsianScientist (Jan. 12, 2015) - By Rebecca Tan - In a few short decades, users have gone from being content with floppy disks that store kilobytes of information to powerful terabyte-sized hard drives. Multiply that demand by the growing number of Internet users and devices that each person owns and you might get a hint of the scale at which data is being generated today.

Faced with a veritable deluge of data, locating information that is relevant and useful can be more difficult than finding a needle in a haystack. In her research, Associate Professor Zheng Baihua from the School of Information Systems at the Singapore Management University (SMU) works on practical applications of database management to develop algorithms that help users find data quickly.

“We are now in the era of big data, where massive amounts of data generated each day make search performance a major issue,” Professor Zheng says. “My main research interest is in finding the data we want as quickly as possible, given the large sizes of datasets common in many situations today.”

Ahead of the curve

Attracted by the real and immediate applications of the field, Professor Zheng was first drawn to data management more than ten years ago.

“When I started doing my PhD, my supervisor did not give me any topic but asked me to go through the literature and find out what I was interested in. I found the topic of data management very interesting because I could relate it to real life, using algorithms to address real-world problems,” she says.

The subject she chose to tackle was the provision of location-based services, such as a system that could help users find the nearest Chinese restaurant or ATM machine. In this scenario, a user submits his or her query to a server, which would then respond based on the user’s location. The server would have to be able to simultaneously handle a large number of users, providing different responses to queries from the same location.

“At that time, mobile devices were relatively rare, leading many people to ask us why we wanted to embark on such a study. However, we felt that being able to provide users with location-specific responses could potentially be applied to a large number of situations, so we tried to develop the kind of services we thought people would want,” Professor Zheng shares.

“Now, we are very happy to see that there are so many popular location-based apps available on mobile phones. It goes to show that the study we did about 10 years ago can actually be implemented, and is in fact, very helpful to users.”

Pervasive computing: not just a concept

Another scenario where data management becomes crucial is pervasive computing, otherwise known as the Internet of Things. First proposed by theorists 20 years ago, pervasive computing envisions a situation where sensing devices are embedded in the environment, communicating with each other to help users.

“Imagine you have a fridge that is embedded with a device that knows you will run out of milk today. With pervasive computing, your fridge will be able to communicate with your mobile device, reminding you to buy a fresh carton of milk as you drive by a supermarket,” Professor Zheng explains.

Up until recently, pervasive computing was only a concept with no system available to demonstrate that it could be achieved, says Professor Zheng. However, now that mobile phones and sensing devices have become commonplace, the challenge for developers has become one of how to integrate data from multiple sources.

“Data could come from dedicated sensors, mobile phones or even the users themselves. With so much data available in such a large variety of formats, how do you find useful data quickly?” she asks.

Accordingly, Professor Zheng’s current research in data management has expanded beyond addressing questions in location-based services to dealing with social media data, a rich source of information with a particularly high demand for data integration.

“We have social media data, which most likely expresses the users’ opinion. Then we have the network itself, containing information such as which users ‘like’ and ‘follow’ each other, pointing to how different users are linked. On top of these are dimensions such as timestamp and location data,” she notes.

“In our research, we try to combine all these different types of data together so as to provide more useful services ultimately. For this project, I am currently in discussions with SMU’s Living Analytics Research Centre.”

Breaking down barriers

Looking forward, Professor Zheng predicts that more and more people will begin to realise the potential of integrating different types of social media. However, she also foresees that there are sizeable barriers preventing seamless data integration from happening.

“For example, a user could have a Facebook or Twitter account and use many different mobile apps. The data generated via these different avenues are owned by different parties who may be unwilling to share the data because of privacy concerns. Even if one party agrees to let us use their data, they may not agree to let us combine their data with those from other sources. This can be a problem because data becomes a limited resource,” she says.

Citing hospital settings as an example, Professor Zheng notes, “There are a lot of things we can do if there is more data about patients, for instance, predicting whether a patient is likely to be re-admitted in the near future. Such information is highly relevant to hospitals, as it can help them plan their resources better.”

“Nevertheless, even within a hospital, not all departments may have the same level of data access. A project that I am currently working on aims to resolve this by creating a platform whereby one can plug all the data in, so that all the data corresponding to one patient is integrated.”

Professor Zheng also hopes that technology barriers will come down for senior citizens, who currently are excluded from the benefits of pervasive computing because they are not adept at using mobile apps.

“As pervasive computing becomes a reality, I would like to explore how sensors in community spaces can be used to provide services even to the older generation. I think there is great potential for us to provide remote healthcare and other critical services.”

Asian Scientist Magazine is a media partner of the Singapore Management University Office of Research.