Similarity Searches on Non-ordered Discrete Data for Emerging Computer Applications: Why and How

Keynote Address

Qiang Zhu
Department of Computer and Information Science
University of Michigan – Dearborn, Dearborn, USA

Abstract

There is an increasing demand for processing similarity searches on non-ordered discrete data in numerous emerging computer applications including bioinformatics, biometrics, cybersecurity, social media, and image processing. To support efficient similarity searches, robust index techniques are required. In this talk, we will discuss why such similarity searches are important for contemporary applications, what the unique challenges are in processing them, and how efficient index schemes can be developed to tackle these challenges. We will also show how classical index methods developed for ordered continuous data fail to work for non-ordered discrete data. We will present some recent index methods specially developed for supporting efficient similarity searches on non-ordered discrete data. We will also discuss other related research issues including how to bulk-load large index trees, how to process similarity searches on hybrid discrete and continuous data, and how to support other types of searches such as box queries on non-ordered discrete data. In the end, we will highlight some future research directions in the area.

Bibliography

Dr. Qiang Zhu is the William E. Stirton Professor at University of Michigan - Dearborn. He was honored as an ACM Distinguished Scientist and an IBM CAS Faculty Fellow. He received his Ph.D. from the University of Waterloo in 1995. His research has been funded by highly competitive funding sources including NSF and IBM. He published extensively in top journals (e.g., IEEE and ACM transactions) and prestigious conferences (e.g., VLDB, IDCE, CIKM) in the database and other related areas. His work was included several well-known database text/research books. He received numerous distinguished research awards including IBM Faculty Awards and BCS Wilkes Award. He has served as editor-in-chief, associate editor, and editorial board member for a number of technical journals including the ISCA International Journal of Computers and Their Applications (Editor-in-Chief, 2010–2012). He also served as a program/organizing committee member for over 100 international conferences. His main research interests include database query optimization, big data processing, streaming data processing, spatio-temporal database, data integration, self-managing database, and data mining.