Faculty Candidate Seminar

Kite: A Scalable Microblogs Data Management System

Amr Magdy, University of Minnesota

Abstract: Developers, researchers, and practitioners have been building a myriad of applications to analyze microblogs data, e.g., tweets, online reviews, and user comments. Examples of such applications include citizen journalism, events detection and analysis, geo-targeted advertising, and studying social influences in social sciences. Building such applications require data management infrastructure to deal with microblogs, including data digestion, indexing, and main-memory management. The lack of such infrastructure hinders the scalability and the widespread of such applications especially among users who are not computer scientists. This talk presents Kite; an end-to-end system that is able to manage microblogs data at a large scale. Using Kite, developers and practitioners can simply write SQL-like queries without worrying about the internal data management issues. Internally, Kite is equipped with scalable indexing and main-memory management techniques to support top-k temporal, spatial, keyword, and trending queries on both very recent data and historical data. Kite indexer supports scalable digestion and retrieval for incoming fast data in real time. Kite memory manager is monitoring the memory contents and smartly decides on which data is regularly moved to disk. Both in-memory and in-disk data are queried seamlessly through efficient retrieval techniques. Kite is open-sourced and available to the community to build on (http://kite.cs.umn.edu).

Bio: Amr Magdy is a doctoral candidate at the Department of Computer Science and Engineering, University of Minnesota. His research interests include database systems, big data management, spatial data management, indexing, and main-memory management with a particular focus on system building issues. His research is published in prestigious research venues, including ACM SIGMOD, ACM SIGSPATIAL, IEEE ICDE, and IEEE TKDE. Amr's research in microblogs data management is recognized among best papers in IEEE ICDE 2014 and has been incubated by several industrial collaborators. His research produced a successful system prototype that is patented and commercialized by a social media analytics company with access to all Twitter data. He has been selected as a finalist for Microsoft Research PhD Fellowship 2014 and he was awarded a prestigious Doctoral Dissertation Fellowship from the University of Minnesota in 2015.

Date: April 21, 2017

Time: 10:00 am

209 Computer Science Building

MISSOURI S&T