THE 9th INTERNATIONAL CONFERENCE ON INFORMATICS
AND SYSTEMS

(INfos2014)
Invited Talk

Wednesday, 17th December, 2014, 9:30-10:30

Cairo University Conference Center
(At Student City of Cairo University, Ahmed Zewail Street, Orman, Dokki)

Data Cleaning from Theory to Practice
Prof. Ihab Ilyas

Abstract
With decades of research on the various aspects of data cleaning, multiple technical challenges have been tackled and interesting results have been published in many research papers. Example quality problems include missing values, functional dependency violations and duplicate records. Unfortunately, very little success can be claimed in adopting any of these results in practice. Businesses and enterprises are building silos of home-grown data curation solutions under various names, often referred to as ETL layers in the business intelligence stack. The impedance mismatch between the challenges faced in industry and the challenges tackled in research papers explain to a large extent the growing gap between the two worlds. In this talk I claim that being pragmatic in developing data cleaning solution does not necessarily mean being unprincipled or ad-hoc. I discuss a subset of these practical challenges including data ownership, human involvement, and holistic data quality concerns. These new set of challenges often hinder current research proposals from being adopted in the real world. I also go through a quick overview of the approach we use in tamr (a data curation startup) to tackle these challenges.

Prof Ihab Ilyas
A Professor of Computer Science at the University of Waterloo. He received his PhD in computer science from Purdue University, West Lafayette in 2004. He holds BS and MS degrees in computer science from Alexandria University. His main research is in the area of database systems, with special interest in data quality, managing uncertain data, rank-aware query processing, and Information extraction. He is an IBM CAS faculty fellow since January 2006, and a recipient of the Ontario Early Researcher Award in 2008 and the David R. Cheriton Faculty Fellowship in 2013. Ilyas is a co-founder of Tamr, a startup focusing on large scale data integration and cleaning.