Abstract:

A large number of wrappers generate tables without column names for human consumption because the meaning of the columns are apparent from the context and easy for humans to understand, but in emerging applications, labels are needed for autonomous assignment and schema mapping where machine try to understand the tables. Autonomous label assignment is critical in volume data processing where ad hoc mediation, extraction and querying is involved. We propose an algorithm Lads for Labeling Anonymous Datasets, which can holistically label tabular web document. The algorithm has been tested on anonymous datasets from a number of sites, e.g music, movie, political, demographic, athletic obtained through different search engines such as Google, Yahoo and MSN. The comparative probabilities of attributes being candidate labels are presented which seem to be very promising, achieved as high as 93% probability of assigning good label to anonymous attribute. To the best of our knowledge, this is the first of its kind for label assignment based on multiple search engines’ recommendation.

Biography:

Emdad Ahmed is a Ph.D candidate in the Department of Computer Science at Wayne State University. He received his Masters Degree in Computer Science & Engineering from UNSW, Australia. He has obtained MBA from IBA, Dhaka University and B.Sc. Engg (CSE) from Bangladesh University of Engineering and Technology (BUET). He has served as Lecturer, North South University, Bangladesh from 2001 to 2005. He has taught a wide variety of CS courses in NSU as well as in WSU. His research interests are in Web Data Integration, Bioinformatics, Computer Networks and simulation. He has published a number of research papers in ACM WIDM, SEDE, IEEE INMIC, ICCIT, ICECE etc.