Interpreting and Representing Tables as Linked Data
Varish Mulwad, Tim Finin, Zareen Syed and Anupam Joshi
University of Maryland, Baltimore County

We are laying a strong foundation for the Semantic Web

We develop a multi-stage framework for this task

We need systems that can (semi) automatically convert and represent data for the Semantic Web!

More than a trillion documents on the web
~ 14.1 billion tables, 154 million with high quality relational data
305,632 Datasets available on Data.gov (US) + 7 Other nations establishing open data; not all is RDF

Is it practical to convert this into RDF manually? NO!!

For every pair of linked strings in the two column, query the knowledge base (KB)

Generate a set of possible relations

Recall >= 0.6 for 81% of the columns

For every string in the column, query the knowledge base (KB)
Generate a set of possible classes
Rank the classes and choose the best

- An SVM-Rank classifier ranks the result set
- A second SVM classifier decides whether to link to the top-ranked instance or not

A template for representing tables as linked data

MAP > 0 for 75% of the columns

Entity Linking

# of Tables 15
# of Columns 52
# of Entities 611