Generating Linked Data by inferring the semantics of tables

Varish Mulwad (Computer Science), University of Maryland, Baltimore County

~ 50 years

Predicition – Column – Nationality

- 154 million high quality relational tables on the web
- 14 nations sharing data as spreadsheets
- Can we build systems to exploit this knowledge?

- Text analysis techniques do not work well with tables
- Structure may be captured, but what about semantics?

For every pair of linked strings in the two column, query the knowledge base (KB)
Generate a set of possible relations
Rank the relations and choose the best relation

Step # 1: Predict class labels

Step # 2: Linking the table cells

Step # 3: Relation Identification

The idea behind Evidence-based Medicine is to judge the efficacy of treatments or tests by meta-analyses or reviews of clinical trials. Key information in such trials is encoded in tables.

However, the rate at which meta-analyses are published remains very low—hampers effective health care treatment...