Generating Linked Data by inferring the semantics of tables

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

~ 50 years

UNDER THE GUIDANCE OF

Barack Obama (born August 4, 1961) is the 44th and current President of the United States, having taken office in 2009. He is the first African American to hold the office. Obama previously served as a United States senator from Illinois, from January 2005 until he resigned following his election to the presidency in November 2008.

http://en.wikipedia.org/wiki/Barack_Obama

Table are everywhere!

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

Step #1: Predict class labels
- For every string in the column, query the knowledge base (KB)
- Rank the classes and choose the best class
- Generate a set of possible classes

Step #2: Linking the table cells
- Re-query KB with predicted class label as additional evidence
- An SVM-Rank classifier ranks the result set
- A second SVM classifier decides whether to link to the top-ranked instance or not

Step #3: Relation Identification
- 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

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.