# EECS 395/495 Project: Word Representations

### Integrating Extracted KBs

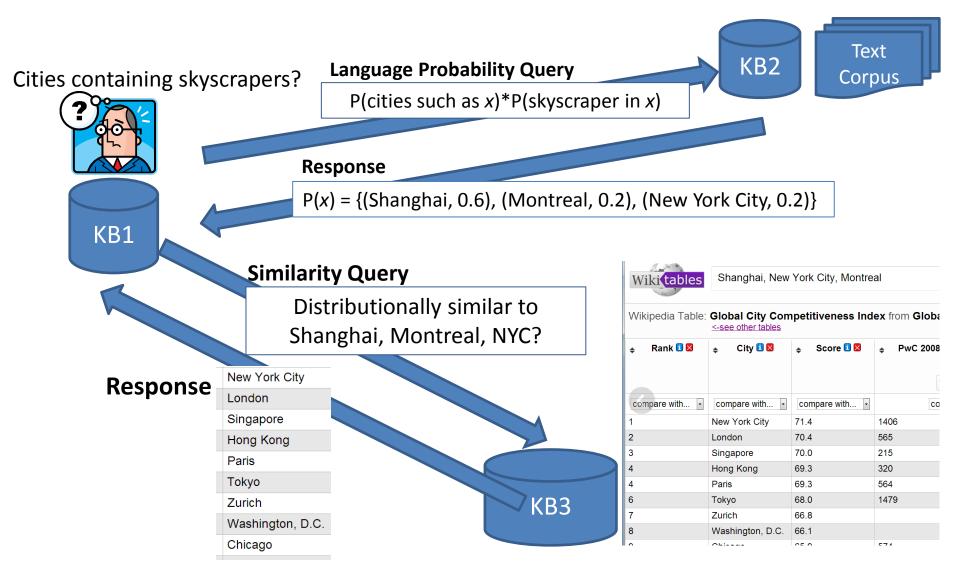
- Several extracted KBs exist:
  - Google Knowledge Graph
  - NELL
  - TextRunner
  - WikiTables
- Problem: KBs are separate silos, with different knowledge formalisms (or lack thereof)

If two KBs contain different sets of knowledge, represented in different ways, how can the KBs share knowledge with each other?

Idea: Use natural language.

Each KB K estimates  $P_K(\mathbf{w})$  over sentences  $\mathbf{w}$  e.g., City( $\mathbf{x}$ ) => increased P("cities such as  $\mathbf{x}$ ")

# Example



### Project

- Create word representations from KBs
  - Vectors of numbers
    - E.g., the probability distributions P(z | word) learned in hmwk #4
  - For use in "similarity queries"

Must make some use of ideas from course

#### **Possibilities**

- KB relations
  - Build graphical model as in homework #4?
  - Latent Dirichlet Allocation (we'll discuss this)
- Tables
  - numeric cells down to smaller space?
- Text
  - NLP tools?

## Goals for today

- Form teams with good distribution of skills
  - Each team needs some CS people for
    - Matching our dev/test sets with resources
    - Programming any models required

- Develop a reasonable "plan of action"
  - Including who will do what