### Source

<table>
<thead>
<tr>
<th>Column 1</th>
<th>Column 2</th>
<th>Column 3</th>
<th>Column 4</th>
<th>Column 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bill Gates</td>
<td>Oct 1955</td>
<td>Microsoft</td>
<td>Seattle</td>
<td>WA</td>
</tr>
<tr>
<td>Mark Zuckerberg</td>
<td>May 1984</td>
<td>Facebook</td>
<td>White Plains</td>
<td>NY</td>
</tr>
<tr>
<td>Larry Page</td>
<td>Mar 1973</td>
<td>Google</td>
<td>East Lansing</td>
<td>MI</td>
</tr>
</tbody>
</table>

### Semantic Model

<table>
<thead>
<tr>
<th>Column 1</th>
<th>Column 2</th>
<th>Column 3</th>
<th>Column 4</th>
<th>Column 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bill Gates</td>
<td>Oct 1955</td>
<td>Microsoft</td>
<td>Seattle</td>
<td>WA</td>
</tr>
<tr>
<td>Mark Zuckerberg</td>
<td>May 1984</td>
<td>Facebook</td>
<td>White Plains</td>
<td>NY</td>
</tr>
<tr>
<td>Larry Page</td>
<td>Mar 1973</td>
<td>Google</td>
<td>East Lansing</td>
<td>MI</td>
</tr>
</tbody>
</table>

Beneficial to automate source discovery, data integration, and service composition

---

**Step 1: Build a Graph from Known Models**

- Create a component in G for each known source model
  - Only add if the model is not subgraph of an existing component
- Annotate links with list of supporting models
- Connect graph components using all paths inferred from the ontology
- Assign low weight \( \varepsilon \) to links within a component
- Weight other links according to their popularity (gray links)

**Step 2: Learn Semantic Types**

- A CRF-based model to assign a Semantic Type to each column from its data
- Semantic Type: Ontology class or Ontology data property + domain

**Step 3: Generate Candidate Models**

- Map learned semantic types to nodes in graph
- Compute Steiner tree (minimal tree) for each mapping

**Step 4: Rank Semantic Models**

- Rank the candidates based on:
  - Cost: sum of the weights
  - Coherence: prefer models with more supporting models

### Evaluation

- **Dataset 1**
  - 17 data sources modeled manually
  - Ontologies: DBpedia, FOAF, GeoNames, WGS84
  - Total number of attributes: 68
  - GED in previous work: 68
  - GED in new approach: 29

- **Dataset 2**
  - 6 museum sources modeled by domain experts
  - Ontologies: EDOM, SKOS, FOAF
  - Total number of attributes: 49
  - GED in previous work: 29
  - GED in new approach: 20

57% **↑**

31% **↑**