Entity Alignment for **Knowledge Graphs** in the Context of Supply Chain Risk Management

Rebeka Gadzo  
Humboldt University Berlin  
gadzo.rebeka@gmail.com

Dr. Yushan Liu  
Siemens AG  
yushan.liu@siemens.com
Agenda

1. **Introduction**
   a. Enhancing Supply Chain Risk Management
   b. Introduction to Entity Alignment

2. **Entity Alignment**
   a. Data Overview
   b. Schema Comparison
   c. Data Preparation

3. **Implementation**
   a. Method 1 - Dedupe
   b. Method 2 - Ditto
   c. Method 3 - GPT 3.5
   d. Framework

4. **Results**
   a. Evaluation
   b. Graph Expansion
Introduction
Enhancing Supply Chain Risk Management

Motivation
Crucial during economic, health, and political crises
Need for risk prediction algorithms for effective risk mitigation

Approach
Integration of macroeconomic information into supply chain data enhances risk assessment
Develop a framework based on real-world scenario applicable to various use cases
Entity alignment is the process of linking corresponding entities across different knowledge graphs or databases to establish connections and improve data integration.
- Siemens stores its Supply Chain data as Knowledge Graph
- Contains over 60'000 suppliers
- 11 node types in total
- No risk data included
Entity Alignment
CoyPu Knowledge Graph

Cognitive Economy Intelligence Platform for the Resilience of Economic Ecosystems (CoyPu)

1. **Events and Incidents** (Demonstration, Disaster, Explosion…)

2. **Geographical Entities** (Airport, City, Continent, Country…)

3. **Business and Industry** (Company, Commodity, Industry Sector, Material…)

4. **Media and Information** (News, WikiNews…)

CoyPu Ontology: https://schema.coypu.org/global/2.2
## Entity Alignment
### Schema Comparison - Supplier Matching

<table>
<thead>
<tr>
<th>Siemens Data</th>
<th>CoyPu Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
<td>Country</td>
</tr>
<tr>
<td>Business Scope</td>
<td>Exiobase Industry</td>
</tr>
<tr>
<td>Component</td>
<td>Product</td>
</tr>
<tr>
<td>Substance</td>
<td>Material/Commodity</td>
</tr>
<tr>
<td>Supplier</td>
<td>Company</td>
</tr>
</tbody>
</table>

![Venn Diagram](image)
Entity Alignment
Data Preparation

- Focus on three countries
- Company name, city name and legal form
- Geolocation
- Company names translation

Original data

Parsed data

Company name

Siemens AG (Munich)

Country

Germany

Parsed name
Legal form
Geolocated city
Implementation
Method 1 - Dedupe

**Dedupe Python Library** uses machine learning to perform fuzzy matching, deduplication and entity alignment quickly on structured data.

1. Step: Blocking          Predicate Blocking, Index Blocking
2. Step: Matching          Levenshtein text distance
3. Step: Human Input       Optional optimization for difficult matches

Repository: [https://github.com/dedupeio/dedupe/tree/main](https://github.com/dedupeio/dedupe/tree/main)
Ditto is an entity alignment solution based on pre-trained language models such as BERT.
Implementation
Method 3 - GPT 3.5 Turbo

Third-Generation Generative Pre-Trained Transformer model that understands and generates natural language or code and has been optimized for chat tasks and performs well for non-chat tasks as well.

- Configured to work as entity matcher: Utilizes system messages for configuration
- Configured to give confidence level output: Utilizes system messages for configuration
- Candidate pair input via user messages: User inputs are processed to identify companies
- Cost-efficient approach: Not fine-tuned due to cost constraints
Implementation
Framework for Entity Alignment between Knowledge Graphs

Input data
- Supplier List 1
- Supplier List 2

Matching
- Dedupe
- Ditto
- GPT

Benchmark

Selection

Graph Expansion
Results Evaluation

Total population of 750 → 273 positive matches, 477 negative matches

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dedupe</td>
<td>86.93%</td>
<td>83.27%</td>
<td>80.22%</td>
<td>81.72%</td>
</tr>
<tr>
<td>Ditto</td>
<td>82.80%</td>
<td>69.15%</td>
<td>95.24%</td>
<td>80.12%</td>
</tr>
<tr>
<td>GPT</td>
<td>90.27%</td>
<td>89.37%</td>
<td>83.15%</td>
<td>86.15%</td>
</tr>
</tbody>
</table>

Results across all countries

<table>
<thead>
<tr>
<th></th>
<th>Siemens Suppliers</th>
<th>Matched Suppliers</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td>4'184</td>
<td>2'341</td>
<td>55.95%</td>
</tr>
<tr>
<td>US</td>
<td>3'819</td>
<td>452</td>
<td>11.84%</td>
</tr>
<tr>
<td>China</td>
<td>3'374</td>
<td>224</td>
<td>6.64%</td>
</tr>
<tr>
<td>Total</td>
<td>11'377</td>
<td>3'017</td>
<td>26.52%</td>
</tr>
</tbody>
</table>

Dedupe’s matches per country
Results
New Supplier’s Properties
Many thanks for your attention!
Let’s discuss..
Appendix
### Appendix

Results for each country

<table>
<thead>
<tr>
<th>Country</th>
<th>True Pos.</th>
<th>False Neg.</th>
<th>True Neg.</th>
<th>False Pos.</th>
<th>Acc.%</th>
<th>Prec.%</th>
<th>Recall%</th>
<th>F1%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GER</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dedupe</td>
<td>71</td>
<td>25</td>
<td>131</td>
<td>4</td>
<td>87.45</td>
<td>94.67</td>
<td>73.96</td>
<td>83.04</td>
</tr>
<tr>
<td>Ditto</td>
<td>93</td>
<td>3</td>
<td>104</td>
<td>31</td>
<td>85.28</td>
<td>75.00</td>
<td>96.88</td>
<td>84.55</td>
</tr>
<tr>
<td>GPT-3</td>
<td>88</td>
<td>8</td>
<td>121</td>
<td>14</td>
<td>90.48</td>
<td>86.27</td>
<td>91.67</td>
<td>88.89</td>
</tr>
<tr>
<td><strong>USA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dedupe</td>
<td>100</td>
<td>16</td>
<td>149</td>
<td>2</td>
<td>93.26</td>
<td>98.04</td>
<td>86.21</td>
<td>91.74</td>
</tr>
<tr>
<td>Ditto</td>
<td>111</td>
<td>5</td>
<td>145</td>
<td>6</td>
<td>95.88</td>
<td>94.87</td>
<td>95.69</td>
<td>95.28</td>
</tr>
<tr>
<td>GPT-3</td>
<td>93</td>
<td>23</td>
<td>142</td>
<td>9</td>
<td>88.01</td>
<td>91.18</td>
<td>80.17</td>
<td>85.32</td>
</tr>
<tr>
<td><strong>CHN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dedupe</td>
<td>48</td>
<td>13</td>
<td>153</td>
<td>38</td>
<td>79.76</td>
<td>55.81</td>
<td>78.69</td>
<td>65.31</td>
</tr>
<tr>
<td>Ditto</td>
<td>56</td>
<td>5</td>
<td>112</td>
<td>79</td>
<td>66.67</td>
<td>41.48</td>
<td>91.80</td>
<td>57.14</td>
</tr>
<tr>
<td>GPT-3</td>
<td>46</td>
<td>15</td>
<td>187</td>
<td>4</td>
<td>92.46</td>
<td>92.00</td>
<td>75.41</td>
<td>82.88</td>
</tr>
</tbody>
</table>
Supplier Matching
Germany

Siemens Suppliers

CoyPu Suppliers