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

SIEMENS

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

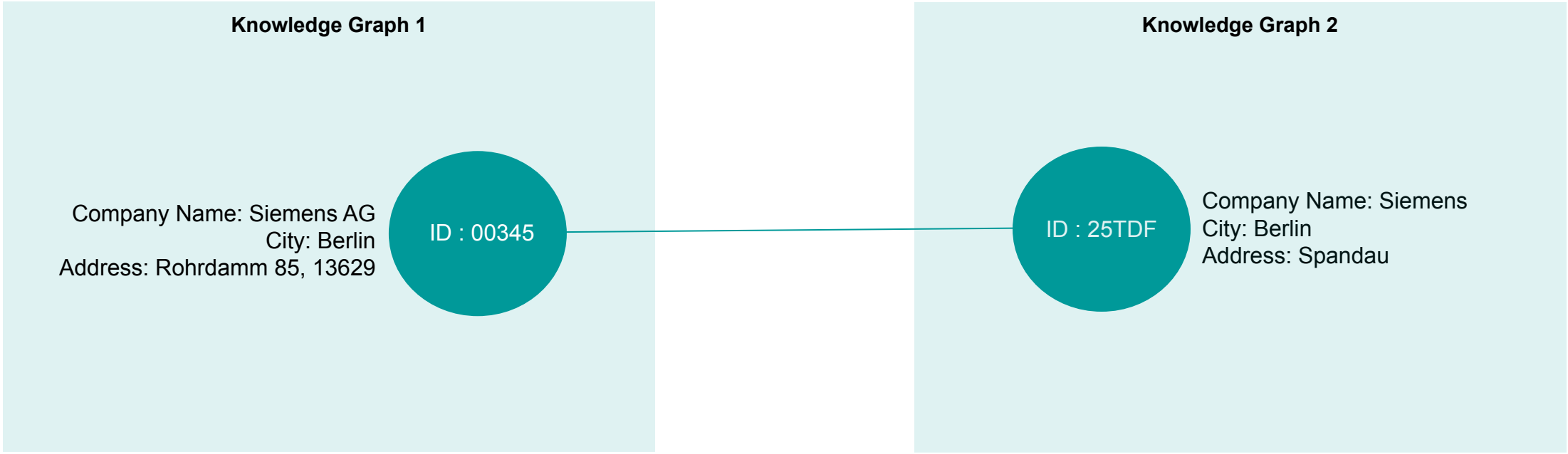
Approach

Integration of macroeconomic information into supply chain data enhances risk assessment

Develop a framework based on real-world scenario applicable to various use cases

Introduction
Entity Alignment

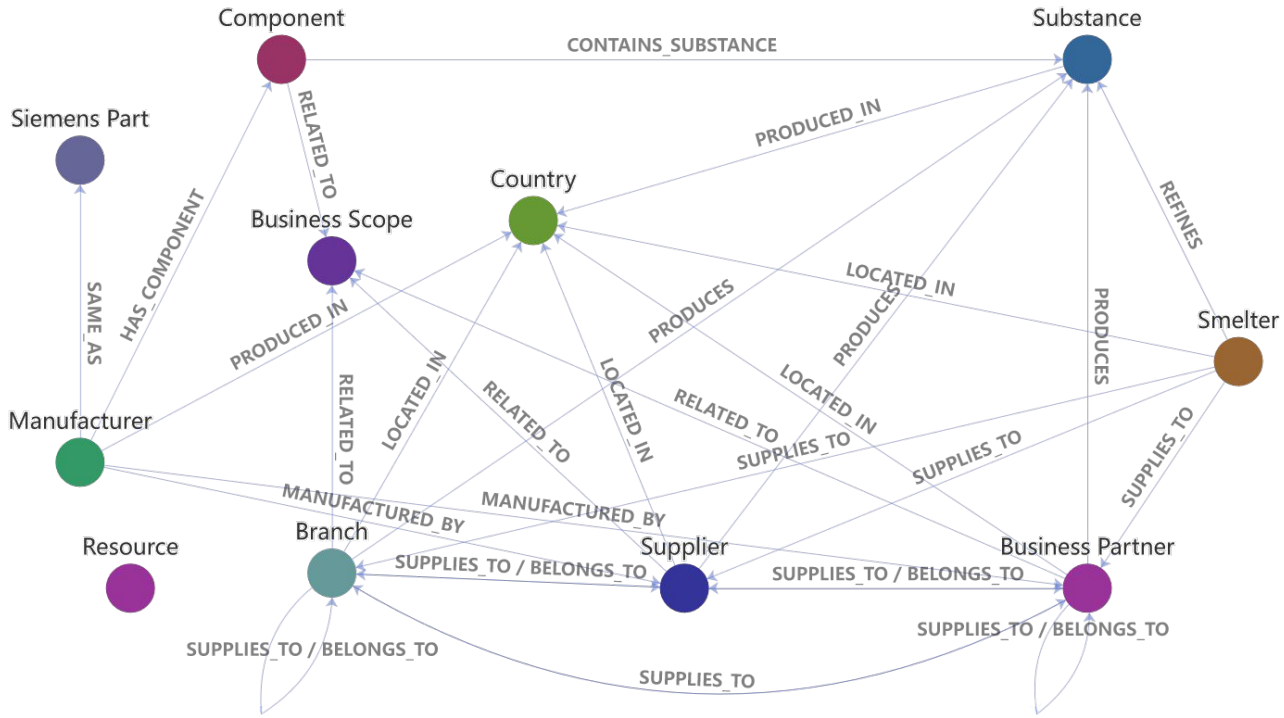
> Entity alignment is the process of linking corresponding entities across different knowledge graphs or databases to establish connections and improve data integration.



SIEMENS

Entity Alignment
Siemens Supply Chain

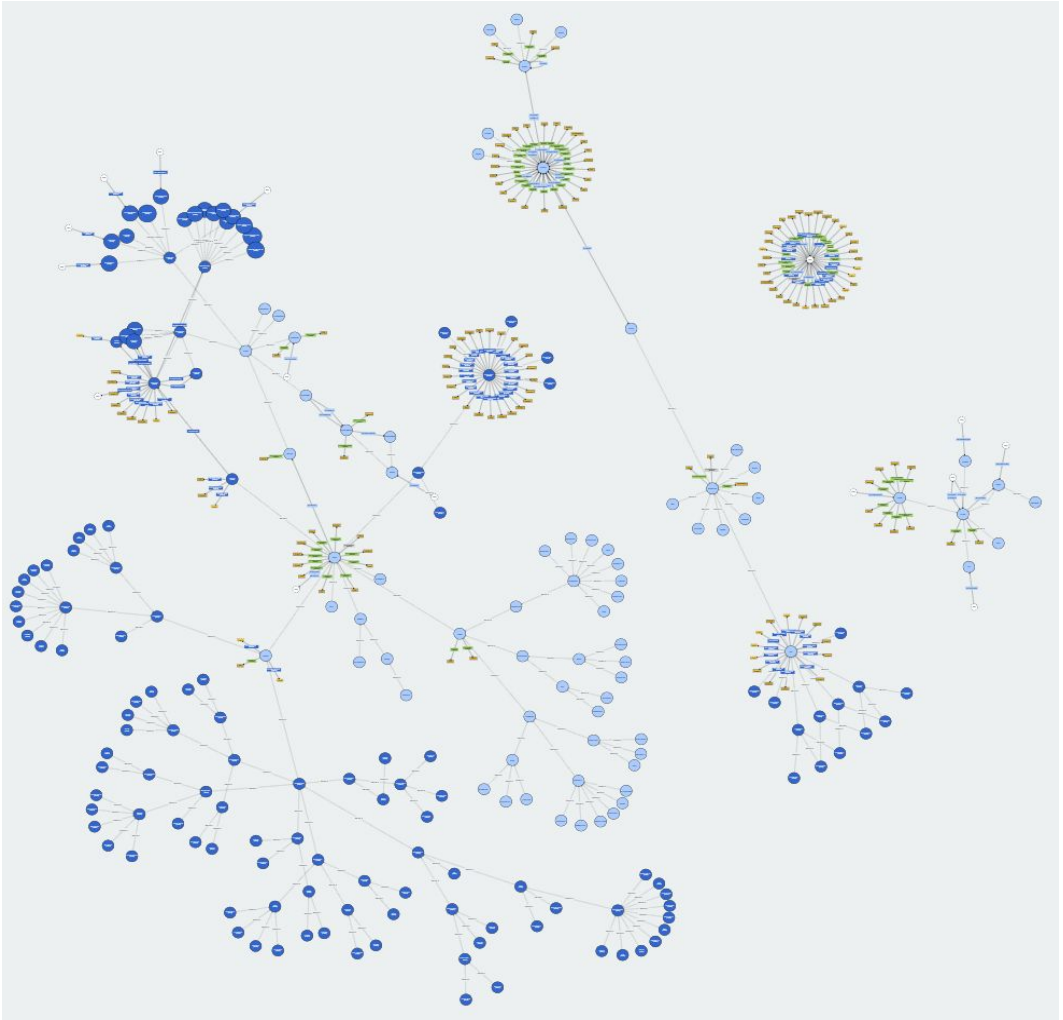
- Siemens stores its Supply Chain data as Knowledge Graph
- Contains over 60’000 suppliers
- 11 node types in total
- No risk data included



SIEMENS

Entity Alignment
CoyPu Knowledge Graph

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



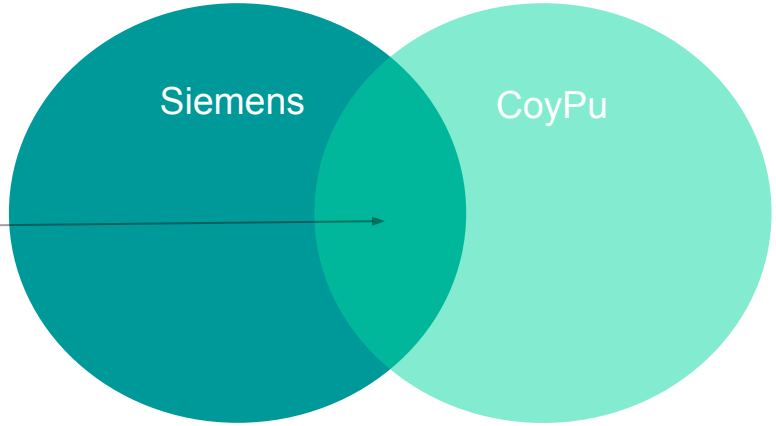
1. **Events and Incidents** (Demonstration, Disaster, Explosion…)
2. **Geographical Entities** (Airport, City, Continent, Country…)
3. **Business and Industry** (Company, Commodity, Industry Sector, Material…)
4. **Media and Information** (News, WikiNews…)

CoyPu Ontology: https://schema.coypu.org/global/2.2

SIEMENS

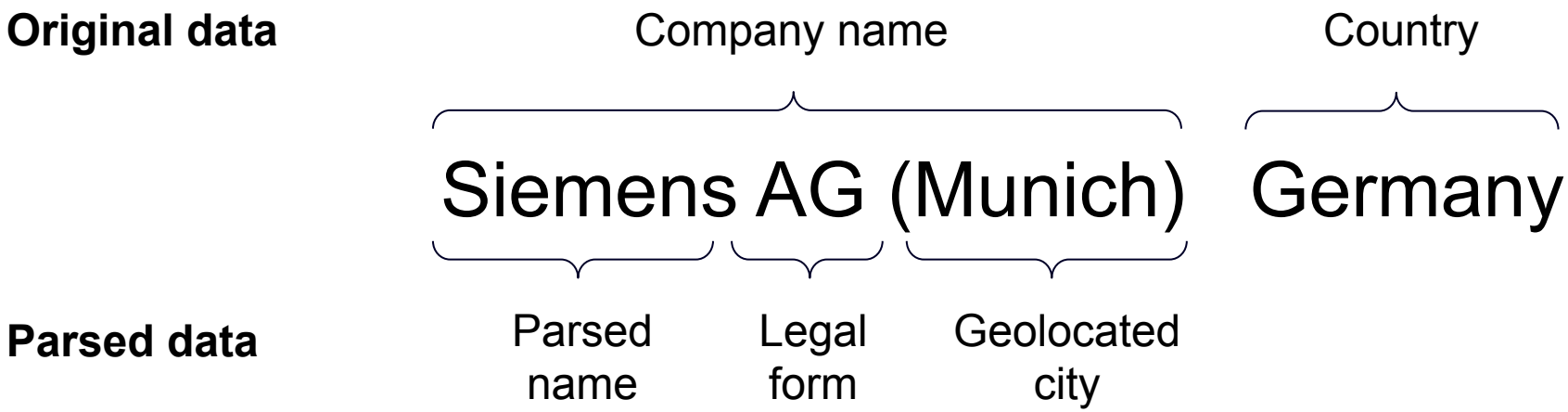
Entity Alignment
Schema Comparison - Supplier Matching

Siemens Data	CoyPu Data
Country	Country
Business Scope	Exiobase Industry
Component	Product
Substance	Material/Commodity
Supplier	Company



Entity Alignment
Data Preparation

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



SIEMENS

Implementation
Method 1 - Dedupe

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

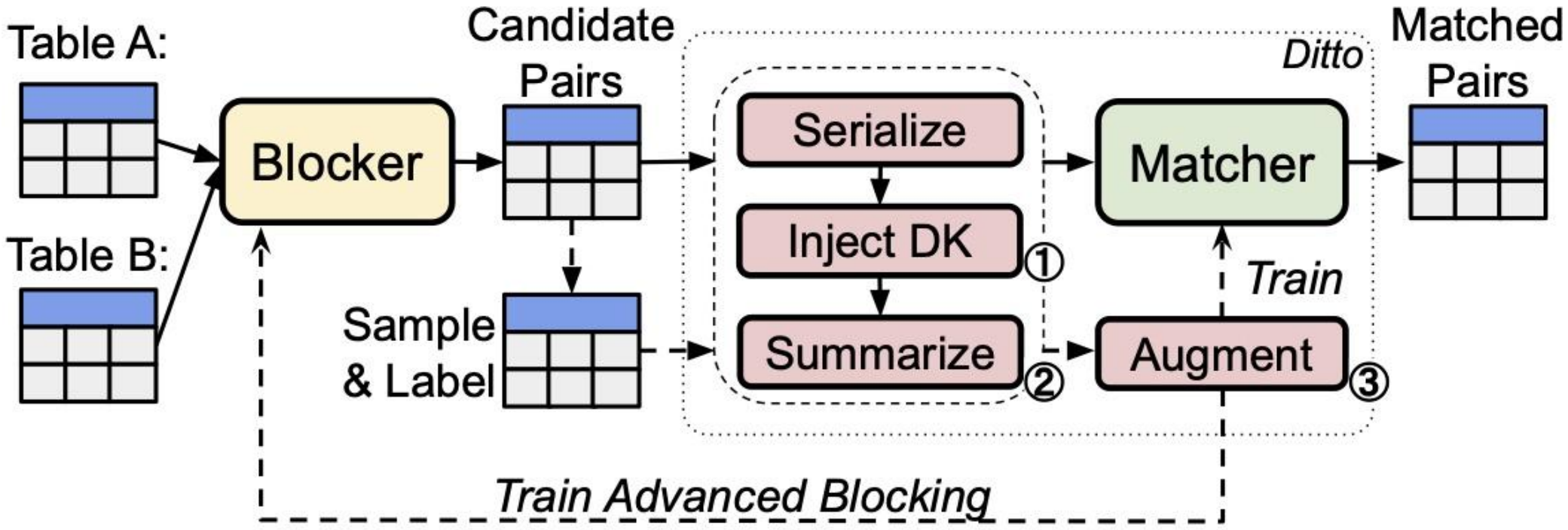
1. Step: Blocking	Predicate Blocking, Index Blocking
2. Step: Matching	Levenshtein text distance
3. Step: Human Input	Optional optimization for difficult matches

Repository: https://github.com/dedupeio/dedupe/tree/main

SIEMENS

Implementation
Method 2 - Ditto (Deep Entity Matching with Pre-Trained Language Models)

> **Ditto** is an entity alignment solution based on pre-trained language models such as BERT



Repository: https://github.com/megagonlabs/ditto

SIEMENS

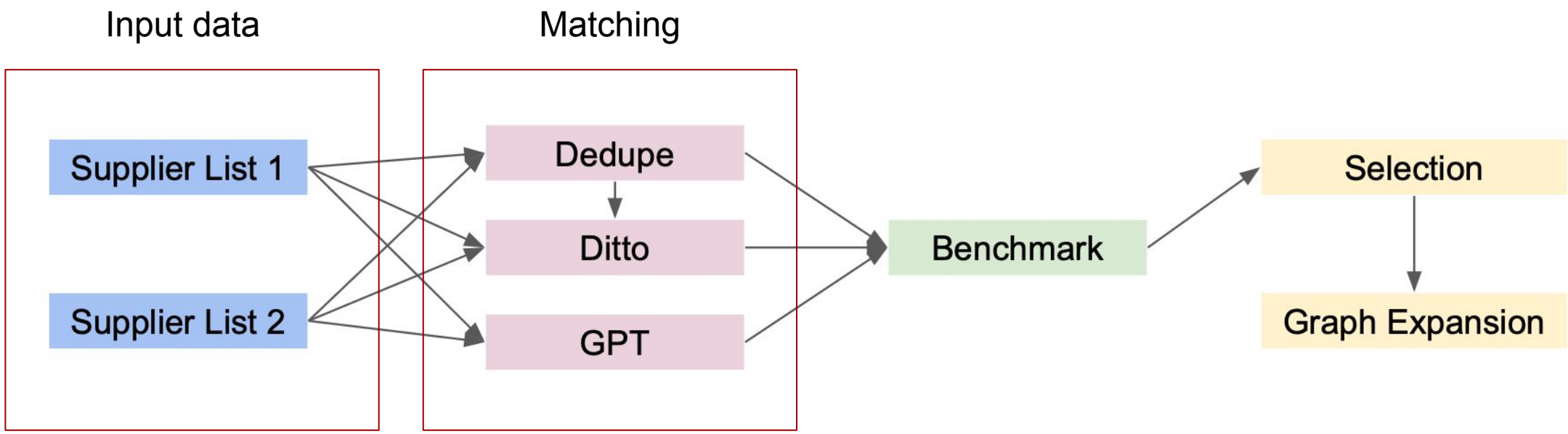
Implementation
Method 3 - GPT 3.5 Turbo

> **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

SIEMENS

Implementation
Framework for Entity Alignment between Knowledge Graphs



SIEMENS

Results
Evaluation

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

	Accuracy	Precision	Recall	F1
Dedupe	86.93%	83.27%	80.22%	81.72%
Ditto	82.80%	69.15%	95.24%	80.12%
GPT	90.27%	89.37%	83.15%	86.15%

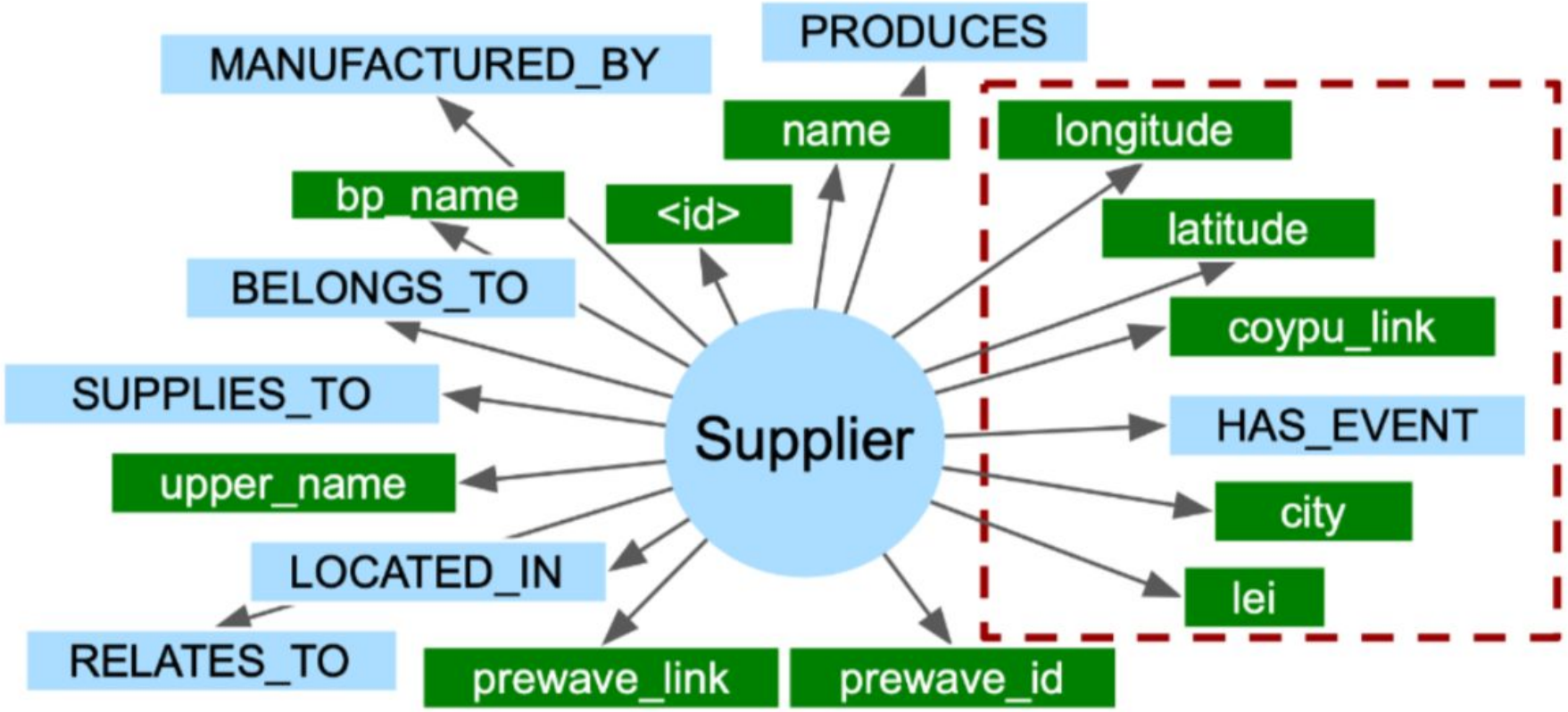
Results across all countries

	Siemens Suppliers	Matched Suppliers	Percentage
Germany	4’184	2’341	55.95%
US	3’819	452	11.84%
China	3’374	224	6.64%
Total	11’377	3’017	26.52%

Dedupe’s matches per country

SIEMENS

Results
New Supplier's Properties



SIEMENS

Many thanks for your attention!

SIEMENS

Let’s **discuss..**

SIEMENS

Appendix

SIEMENS

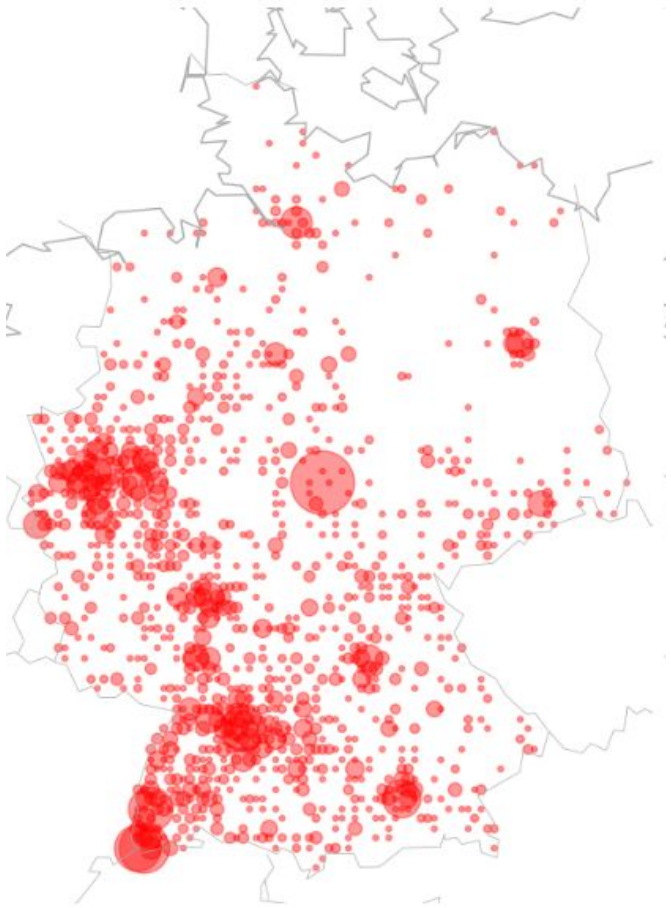
Appendix
Results for each country

Results per Country

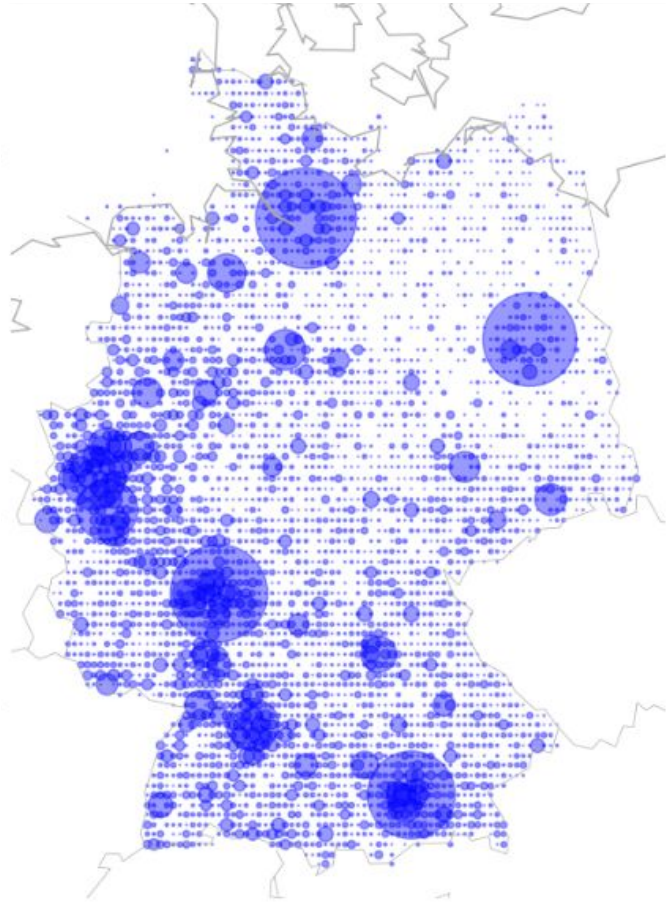
GER	True Pos.	False Neg.	True Neg.	False Pos.	Acc.%	Prec.%	Recall%	F1%
Dedupe	71	25	131	4	87.45	94.67	73.96	83.04
Ditto	93	3	104	31	85.28	75.00	96.88	84.55
GPT-3	88	8	121	14	90.48	86.27	91.67	88.89
	Pos.=96		Neg.=135		Total=231			
USA	True Pos.	False Neg.	True Neg.	False Pos.	Acc.%	Prec.%	Recall%	F1%
Dedupe	100	16	149	2	93.26	98.04	86.21	91.74
Ditto	111	5	145	6	95.88	94.87	95.69	95.28
GPT-3	93	23	142	9	88.01	91.18	80.17	85.32
	Pos.=116		Neg.=151		Total=267			
CHN	True Pos.	False Neg.	True Neg.	False Pos.	Acc.%	Prec.%	Recall%	F1%
Dedupe	48	13	153	38	79.76	55.81	78.69	65.31
Ditto	56	5	112	79	66.67	41.48	91.80	57.14
GPT-3	46	15	187	4	92.46	92.00	75.41	82.88
	Pos.=61		Neg.=191		Total=252			

SIEMENS

Supplier Matching
Germany



Siemens Suppliers



CoyPu Suppliers

SIEMENS