



# Closed-Loop Link Mining of Textual Data

Zohreh Nazeri

(703) 983-5841

nazeri@mitre.org

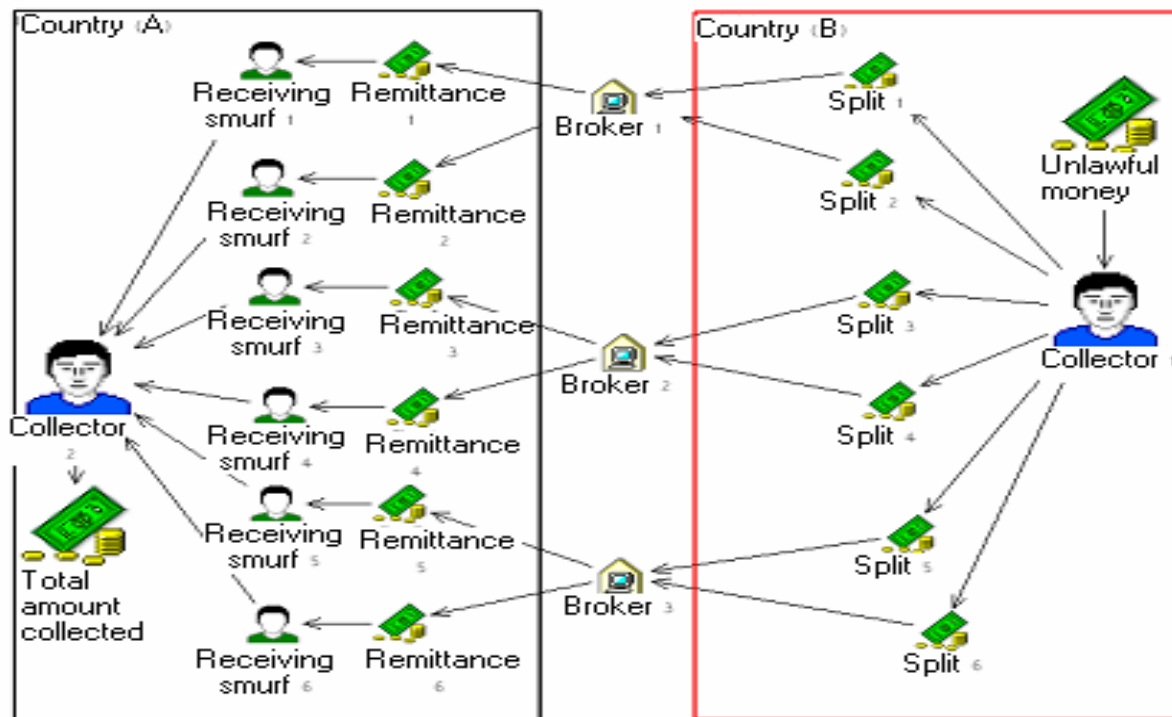
MSR

# Problem

- Analysts in a variety of domains are faced with the challenging task of **identifying and analyzing entity relationships** imbedded in large amounts of textual data.
  - In the **finance** domain, analysts need to uncover suspicious entities (individuals, organizations) and entity relationships in regard to money laundering and other fraudulent activities
  - In the **anti-terrorism** domain, analysts need to identify social links between individuals as well as track sequences of activities and uncover their links to certain events

# Background

- To evade transaction reporting requirements established by most regulations on cash activity above a specific threshold, money launderers often split up the total amount of the transaction into sums that are below the set threshold



# Objective

- **This research aims at developing a link mining system for textual data that:**
  - **extracts entities and relations from textual data**
  - **exploits disparate sources of data,**
  - **discovers communities and relationships among the entities embedded in textual and structured data, and**
  - **ranks the results in the order of their significance to the user**

# Activities

## ■ Developing software

- Common framework for Information Extraction (IE) and Link Mining (LM)
- Improving the entity/event extraction and link mining techniques
- Developing feedback mechanism between IE and LM

## ■ Studying data sources and use cases

- FinCEN data
- Enron data

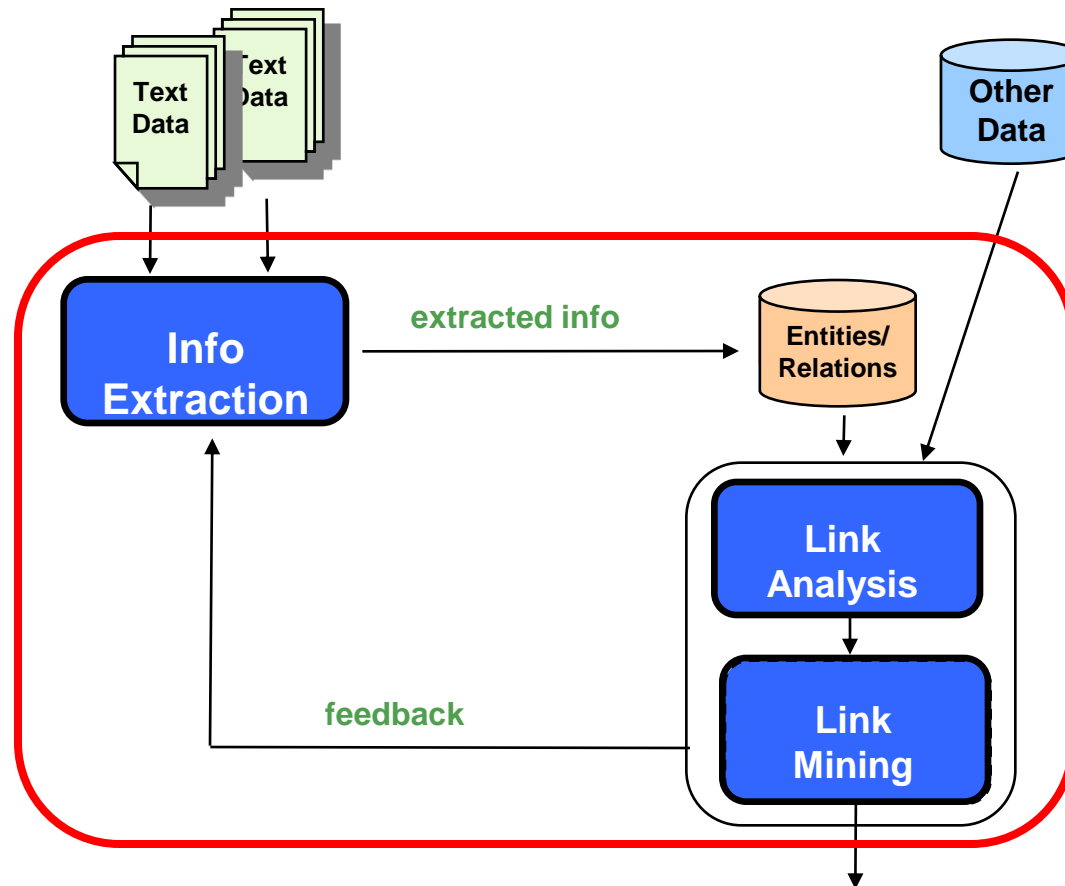
## ■ Collaborating with Arizona State University

- Developing single-community discovery algorithms

## ■ Demonstrating a proof of concept on the studied data sources

# Highlight

Linking the relationships depends on the effectiveness of information extraction from the text.

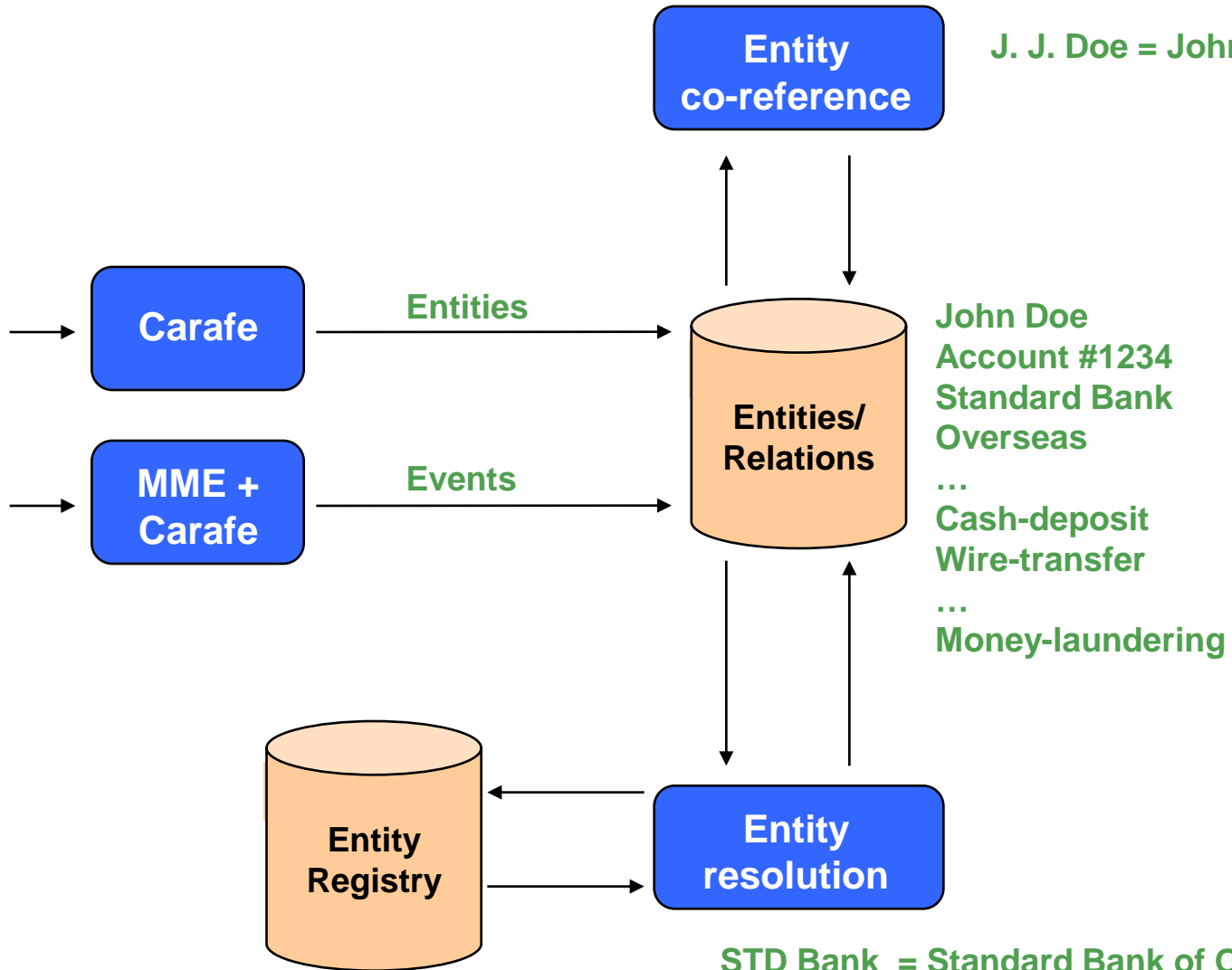


<u>extracted info</u>	<u>type</u>
person . . . . .	node
organization . . . . .	node
account . . . . .	node
date/time . . . . .	attribute
location . . . . .	attribute
currency . . . . .	attribute
event . . . . .	edge
deposit, wire transfer, etc.	

Discovered patterns of interest

# Highlight

Customer John Doe makes frequent cash deposits. Mr. Doe then wires large balances to account #1234, in STD Bank overseas.



# Impacts

- MITRE sponsors such as FinCEN, IRS, and DHS, can benefit from the outcome of this work
  - **Tracing suspicious activities** through multiple data sources in a short amount of time
  - **Discovering previously unknown cases** by recommending suspicious patterns for further investigation
  - **Prioritizing** the discoveries and allocating limited human resources intelligently
- Contribution of this work to the research community is a novel methodology to perform link mining on textual data
- The technology developed by this research could be made available to the public through the MITRE Technology Transfer Office

# Future Plans

