Matching Metadata on Blockchain for Self-Sovereign Identity

Frederico Schardong, Ricardo Custódio, Laércio Pioli, João Meyer
Federal University of Santa Catarina, Brazil

Summary

- Introduction
  - Example/Motivation
  - Objectives
- Systematic Literature Review
- Proposal
- Experiments
  - Planning
  - Execution & Results
- Conclusions & Future Works

Introduction

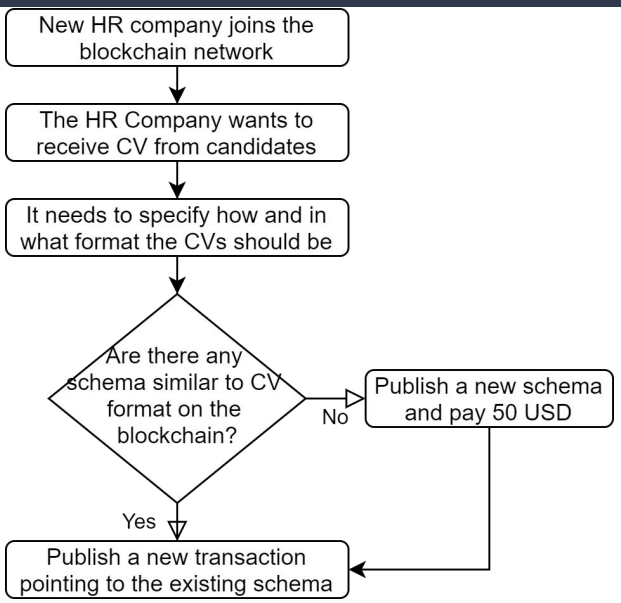
- Privacy is perhaps the biggest concern that people have when using online services
  - Personal data is critical
  - There are huge data silos of personal information (facebook, google, twitter, etc)

Introduction

- Privacy is perhaps the biggest concern that people have when using online services
  - Personal data is critical
  - There are huge data silos of personal information (facebook, google, twitter, etc)

- Self-Sovereign Identity (SSI) [1]
  - Decentralized/Blockchain identity
  - Transactions describe new users, connection info, data sharing formats (schemas), and more

Introduction
Example



Introduction
Objectives

- To systematically review the scientific literature for works that perform schema matching on blockchain in the context of self-sovereign identity
- To create a tool for schema matching on Sovrin, a blockchain-based self-soverign identity solution
- To compare the F-score of the proposed tool with the existing ones over schema matching queries

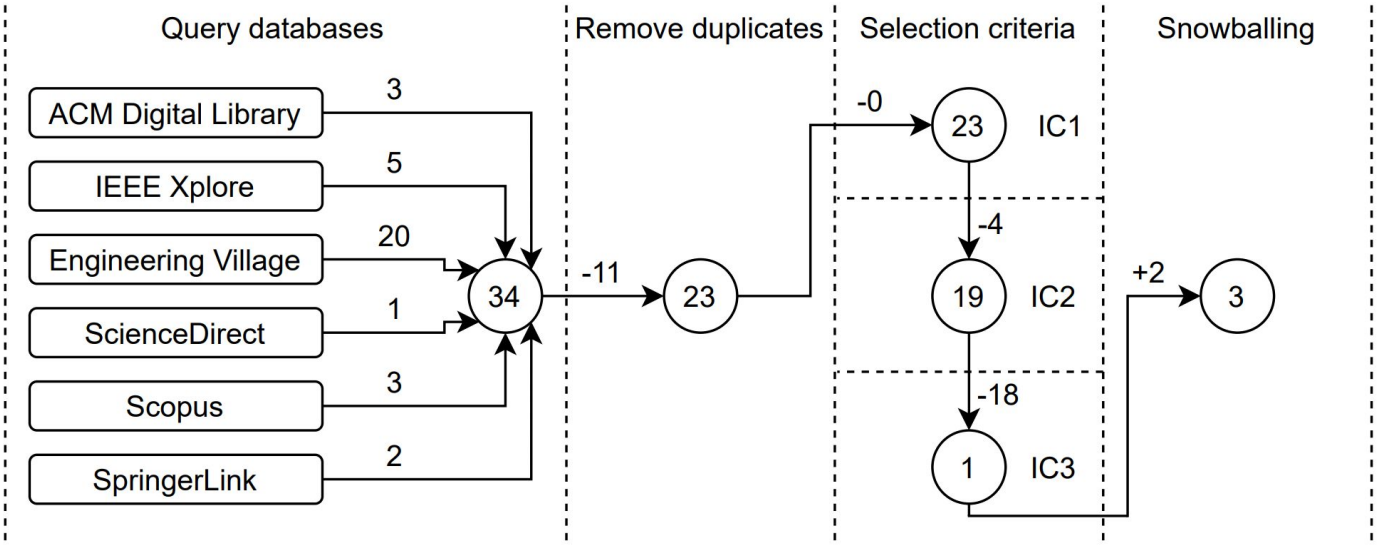
Systematic Literature Review
Search String

(blockchain **OR** ledger)
AND
(ontology **OR** retrieve **OR** matching **OR** similarity **OR** crosswalk **OR** mapping)
AND
(self-sovereign **OR** self-sovereignty **OR** self sovereign **OR** self-sovereignty **OR** decentralized identity **OR** decentralised identity **OR** distributed identity)

Systematic Literature Review
Inclusion/Exclusion Criterias

- IC1: Articles written in English; AND
- IC2: Articles that necessarily have a title and abstract; AND
- IC3: Articles that propose a technique to search or match schemas in blockchain-based identity management solutions.

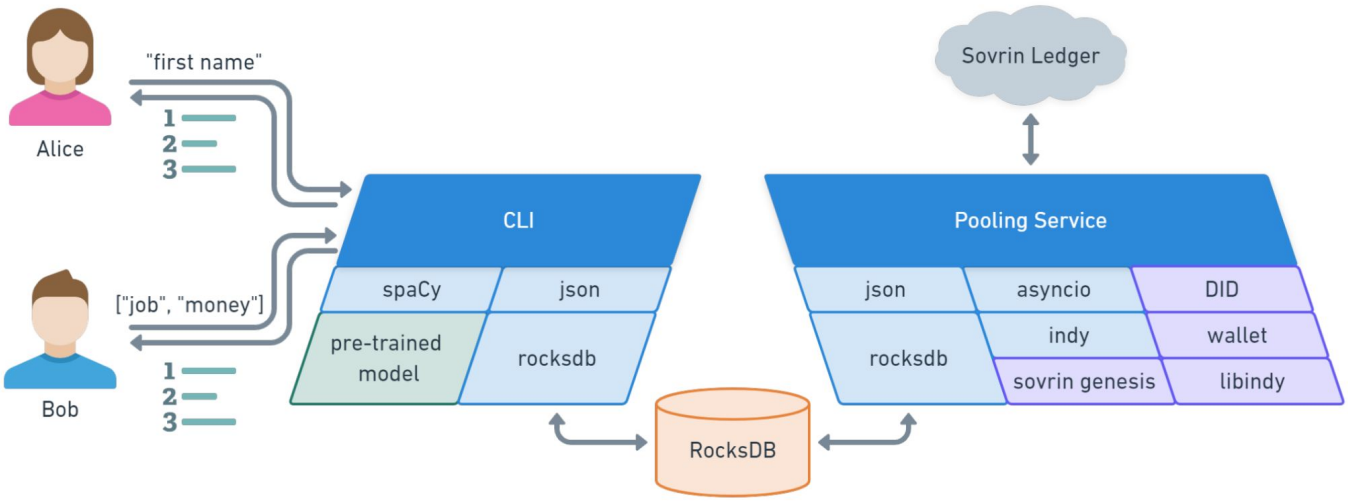
Systematic Literature Review
Mapping Conduction



Proposal

- The creation of a Python tool that uses spaCy [6], a natural language processing algorithm with semantic similarity and multi-language support, to enable users to perform schema matching in the Sovrin blockchain
- Available at https://github.com/fredericoschardong/sovrin-schema-matching

Proposal



Proposal

Table 2. Results for the query "money"

Score	Trans. #	Schema Values
0.595	59556	State, Listing Agent, Lot Number, Buyer First Name, Contract Signed Date, Purchase Price, Postal code, Buyer Last Name, Subidivision, City, Street Address, Buyer Agent, Estimated Completion date, Model Name, Earnest Money
0.590	59555	Application Status, Loan Number, Loan Amount, Date of Approval, First Name, Subdivision, Lot Number, Last Name, Earliest Closing Date, Loan Term in Months, APR, Interest Rate, Lender Name
0.569	59551	Credit Score, Account Type, Institution Name, DOB, Total Deposits, SWIFT BIC, IBAN, Last Name, First Name, Statement Period, Average Montly Balance Last 12 months, Total Withrawwals, Account Number

Proposal

Table 3. Results for the query `["student","university","degree"]`

Score	Trans. #	Schema Values
0.660	54788	`degree, last_name, axuall_proof_id, institution, status, year, first_name`
0.660	54802	`first_name, institution, axuall_proof_id, last_name, degree, year, status`
0.620	33627	`DEMO-GPA, DEMO-Major, DEMO-Degree, DEMO-College Name, DEMO-Student Name`

Experiment
Planning

- Define three queries and manually choose which schemas are similar and should be returned by any schema matching tool
  - Query 1: “address”
  - Query 2: “first name”
  - Query 3: “company job”

Experiment
Execution & Results

Table 4. The number of correct and incorrect schemas based on our manual selection (support) and the f-score of predictions.

Query	Technique	Support		f-score		
		True	False	True	False	Avg.
"address"	[12]	36	112	-	-	-
	[21]			0.65	0.92	0.79
	[17]			0.46	0.90	0.68
	This work			0.29	0.84	0.56
"first name"	[12]	67	81	-	-	-
	[21]			0.66	0.83	0.74
	[17]			0.48	0.78	0.63
	This work			0.76	0.84	0.80
"company job"	[12]	23	125	-	-	-
	[21]			0.00	0.92	0.46
	[17]			0.00	0.92	0.46
	This work			0.55	0.94	0.74

Conclusions & Future Works

- We provided a systematic review of the scientific literature in the context of the blockchain-based SSI to identify the research materials that have been published considering the schema matching problem.
- We also proposed a novel tool to perform schema matching on Sovrin, which can be easily expanded to other blockchain-based SSI systems.
- For future works, we intend to experiment with more queries.

References

[1] - The Sovrin Foundation, Write To The Sovrin Public Ledger. [On-line]. Available: https://sovrin.org/issue-credentials/, Acessado em 25/03/2021.

[6] - Honnibal, M., Montani, I., Van Landeghem, S., and Boyd, A. (2020). spaCy: Industrial-strength Natural Language Processing in Python. https://doi.org/10.5281/zenodo.1212303, Acessado em 26/03/2021.

[12] - Lux, Z. A., Beierle, F., Zickau, S., and Gondor, S. (2019). Full-text search for verifiable credential metadata on distributed ledgers. In 2019 Sixth International Conference on Internet of Things: Systems, Management and Security (IOTSMS), pages 519–528. IEEE.

[17] - Stas, P. (2019). Hyperledger indy transaction explorer. https://indyscan.io/, Acessado em 21/03/2021.

[21] - Whitehead, A. (2019). Sovrin main net. https://sovrin-mainnet-browser.vonx.io/, Acessado em 21/03/2021.

Questions?