



Rensselaer

Shaping the Web's Future with Decentralized Knowledge Graphs

WWW Day

Organized by IEEE WIE Student Branch Affinity Group, University of Jaffna

Aug 01, 2024

Oshani Seneviratne

Associate Director, Tetherless World Constellation

Assistant Professor, Department of Computer Science

Rensselaer Polytechnic Institute, Troy NY USA



About Me

- Assistant Professor in Computer Science at RPI
 - RPI = Rensselaer Polytechnic Institute
 - Oldest technological university in the USA
- I teach:
 - AI and Blockchain
 - Introduction to AI
 - AI in Fiction and Fact
- I research:
 - Decentralized Systems (Web, Blockchain, Knowledge Graphs)
 - Applied AI (Health Informatics and FinTech)

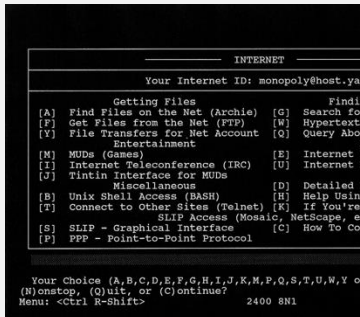
Background

- PhD in Computer Science from MIT
- Doctoral and Masters Supervisor: Tim Berners-Lee

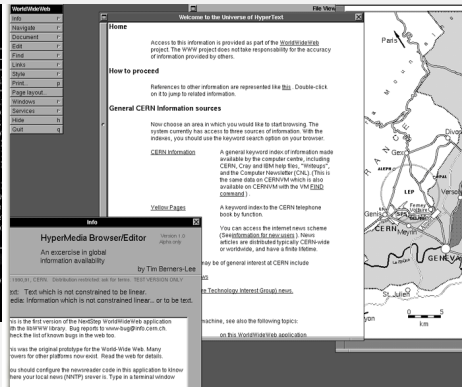




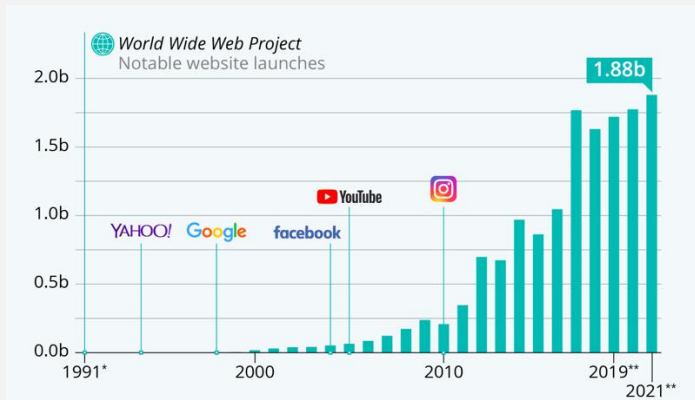
Early Browsers



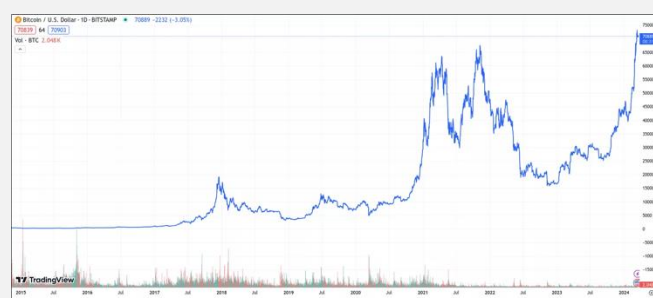
Web Applications



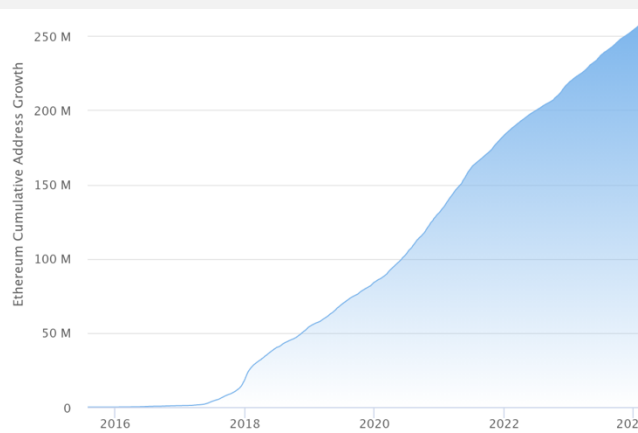
Website Registrations



Bitcoin Valuation



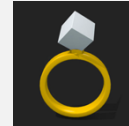
Ethereum Addresses



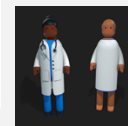
Decentralized (Blockchain) Apps



Banking



Supply chain



Health care



Voting



Property records



Insurance





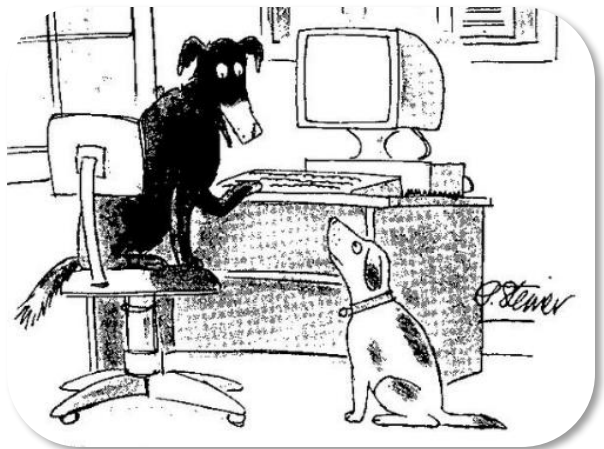
(Decentralized) Knowledge Graphs and AI

- Enhanced Search and Information Retrieval
- Natural Language Processing
- Recommendation Systems
- Fraud Detection and Risk Management
- Healthcare and Biomedical Applications
- Knowledge Management and Integration
- Explainable AI (XAI)
- Robotics and Autonomous Systems



Web3/Blockchain and AI

- Enables decentralized data sharing.
- Incentive mechanisms for model training and inference.
- Decentralized Autonomous Organizations (DAOs) can streamline many computational processes.



On the Web, nobody knows you are a dog!



On Blockchain, nobody knows you are an AI!



Web 1 →

Web 1

“static web”

Read

Web 2 →

Web 2

“dynamic web”

Read

Write

Interactable

Web 3

Web 3

“semantic web”

“ownable web”

Read

Read

Write

Write

Interactable

Interactable

Meaningful

Ownable

Verifiable



Tim Berners-Lee's TED Talk (2009)

2009



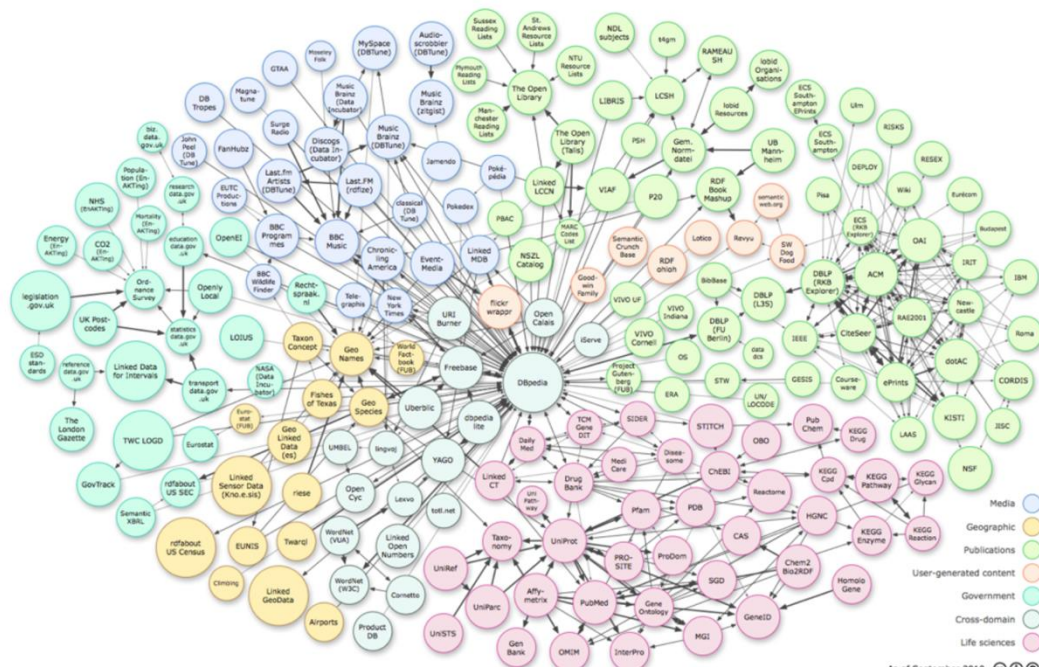
LINKED DATA

- ★ On the web, open license
- ★★ Machine-readable data
- ★★★ Non-proprietary format
- ★★★★ RDF standards
- ★★★★★ Linked RDF

IS YOUR DATA 5 ★ ?

Linked Open Data Cloud (2010)

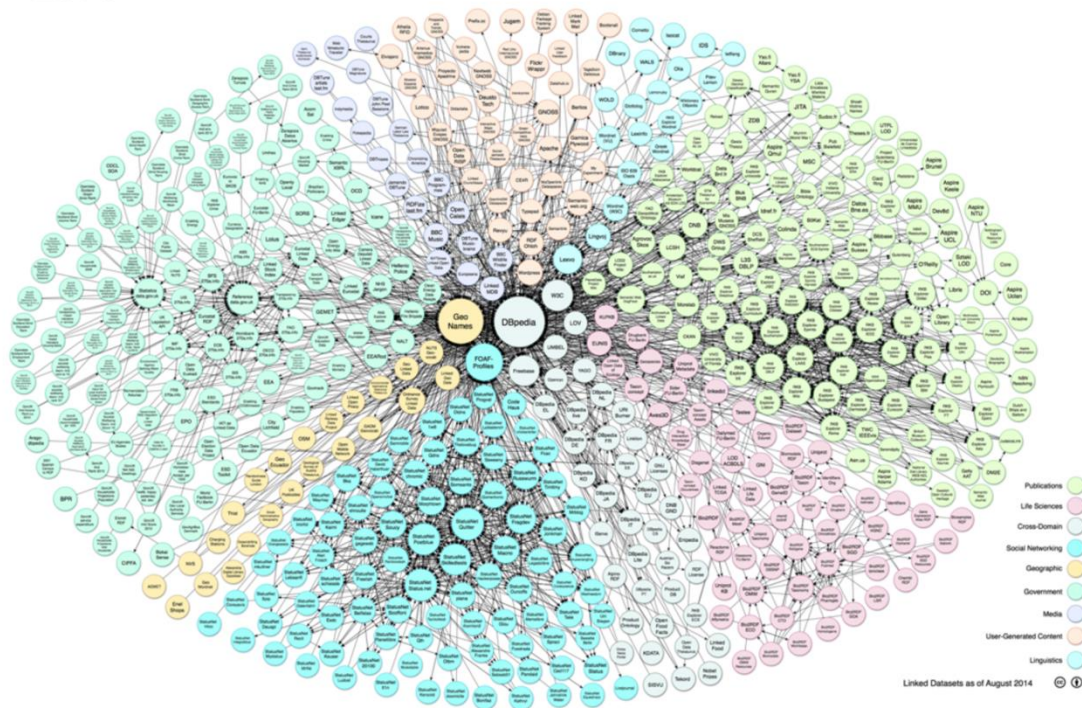
2010





Linked Open Data Cloud (2014)

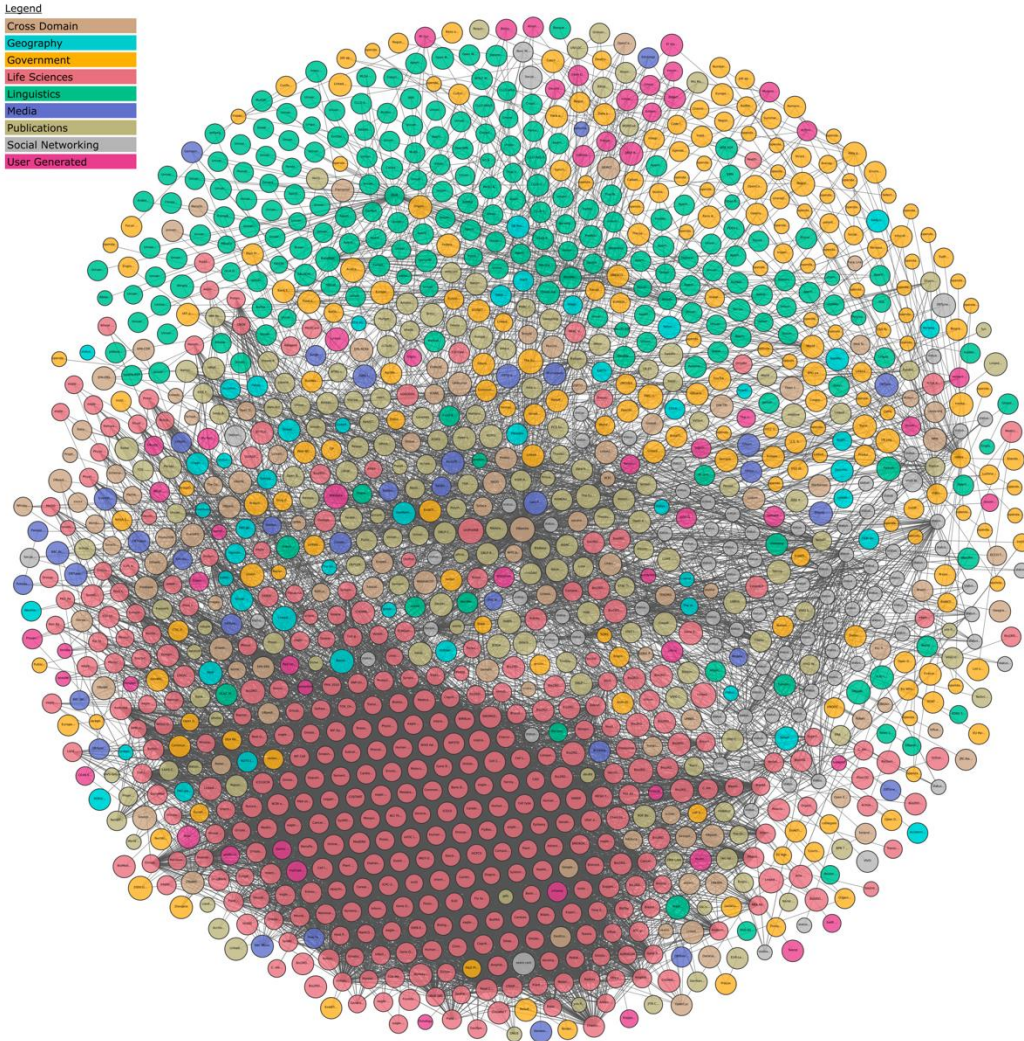
2014





Linked Open Data Cloud (now)

Legend



<https://lod-cloud.net/#diagram>

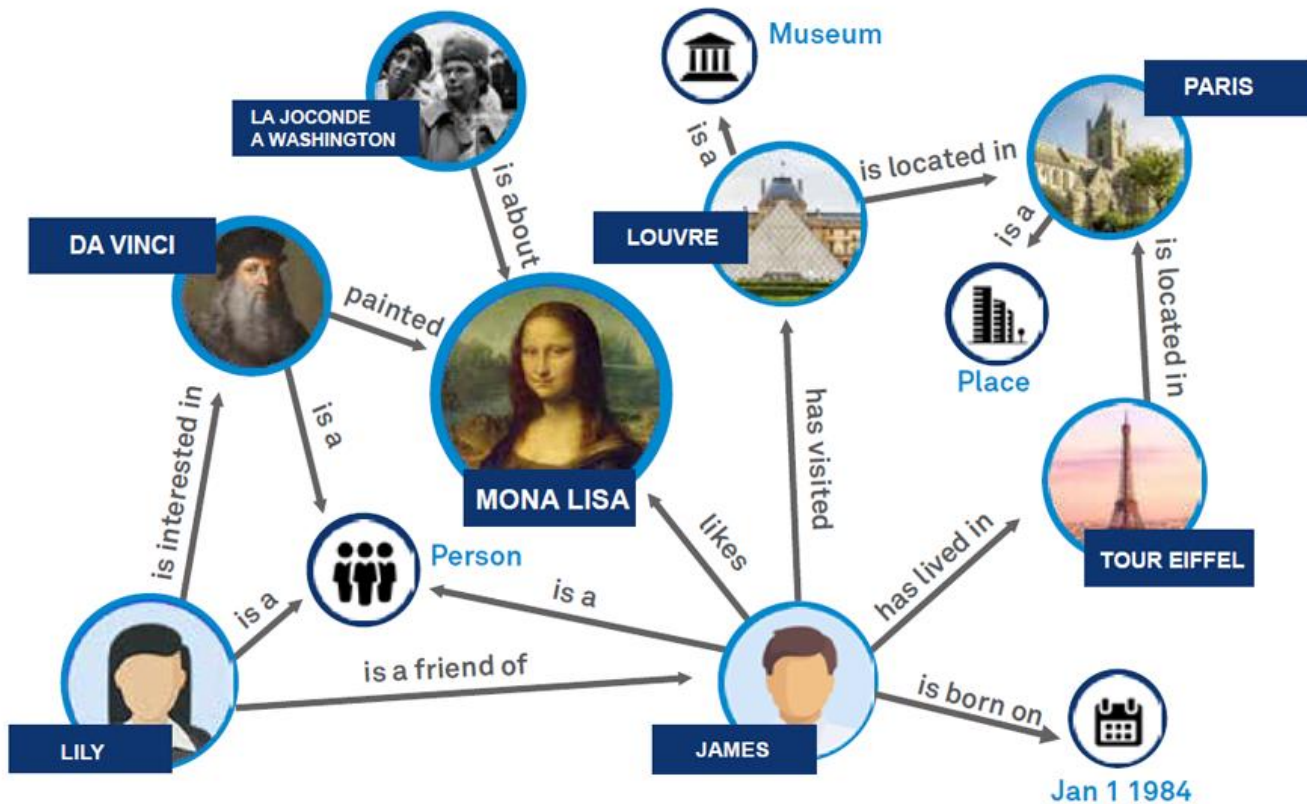


Limitations of Linked Open Data

- Data Quality
 - Datasets may contain errors, inconsistencies, or outdated information, or be incomplete which can affect the reliability of the linked data.
- Data Governance and Privacy
 - There is a need to share data in a privacy-preserving manner
 - Lack of robust access controls, and anonymization techniques to maintain trust and mitigate risks
- Sustainability and Maintenance
 - Relies on the willingness of data publishers to provide and maintain their datasets



What is a Knowledge Graph?





What Decentralized Knowledge Graphs (DKG) Enables

A DKG is a **global knowledge graph** comprising a shared set of verifiable assertions that are not tied to any central authority (can be both public and private)

Smart contracts allow users to contribute to the knowledge graph in a secure and incentivized manner.

- Knowledge sharing and collaboration
- Proper long-term sustainable data stewardship
- Ownership and Control of the resources in the KG
- Integration of advanced analytics



Some of my DKG research

Scientific Knowledge Sharing

Accountable Bench-to-Bedside Data-Sharing Mechanism for Researchers; *Oshani Seneviratne, Deborah McGuinness; Transactions on Social Computing, 2023.*

Assessing Scientific Contributions in Data Sharing Spaces; *Kacy Adams, Fernando Spadea, Conor Flynn, Oshani Seneviratne; Sci-K'23.*

Collaborative Decentralized Knowledge Graph Construction

Swarm Contracts: Smart Contracts in Robotic Swarms with Varying Behavior; *Jonathan Grey, Isuru Godage, Oshani Seneviratne. IEEE Blockchain Conference 2020.*

Decentralized Framework for Collection and Secure Storage of Google Street View Data: Case Study; *Sanjaya Mallikarachchi, Bonnie Ho, Iyad Kanj, Oshani Seneviratne and Isuru Godage; IEEE GlobCon2023 and ICCAR2023*

Generating Smart Contracts for Computable Knowledge Graphs

Translating Clinical Decision Logic Within Knowledge Graphs to Smart Contracts; *William Van Woensel, Manan Shukla and, Oshani Seneviratne;* Semantic Web solutions for large-scale biomedical data analytics (SeWebMDA-2023)



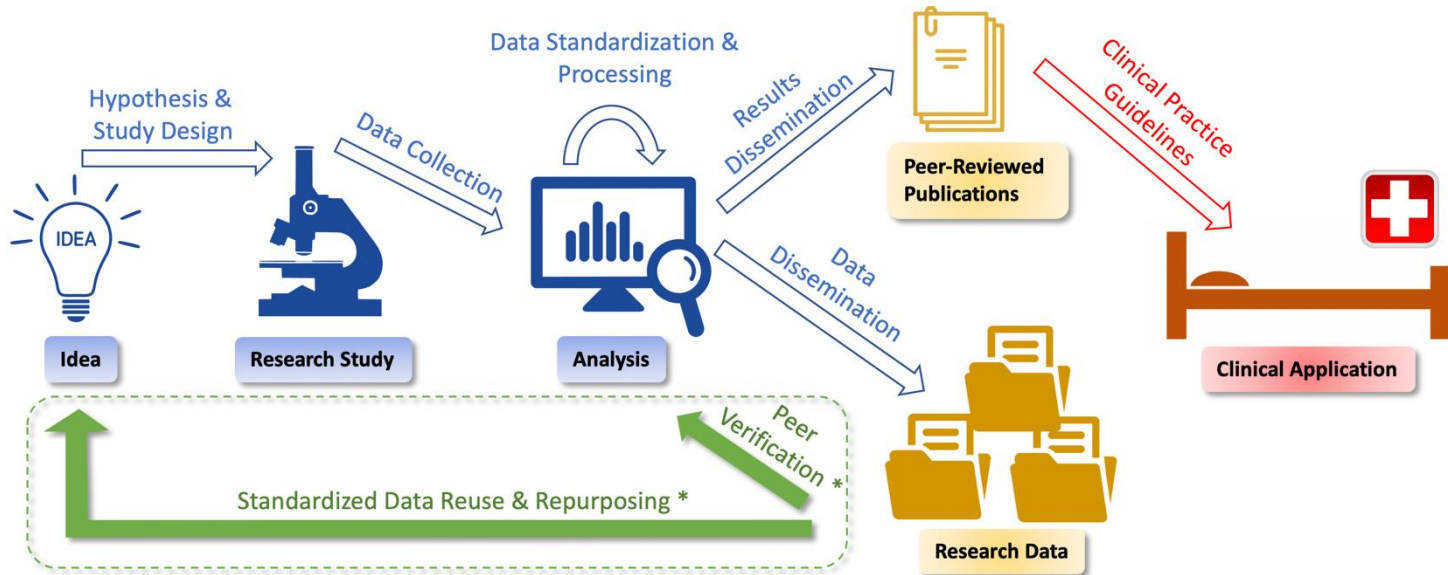
Rensselaer Incentivized Accountable Research Data Sharing Ecosystem as a DKG

Accountable Bench-to-Bedside Data-Sharing Mechanism for Researchers;
Oshani Seneviratne, Deborah McGuinness;
Transactions on Social Computing, 2023.

Goals:

- Handle “Researcher’s Dilemma”
- Reward Reproducible Research and Peer Verification
- Tokenization of Research Rewards

Bench-to-Bedside Biomedical Research Scenario:





Rewarding Reproducible Research with the SCIENCE Index

Assessing Scientific Contributions in Data Sharing Spaces; *Kacy Adams, Fernando Spadea, Conor Flynn, Oshani Seneviratne; Sci-K'23.*

SCIENCE

Capability-based

Intention-centric

Experiment-oriented

Networked

Collaborative

Expression

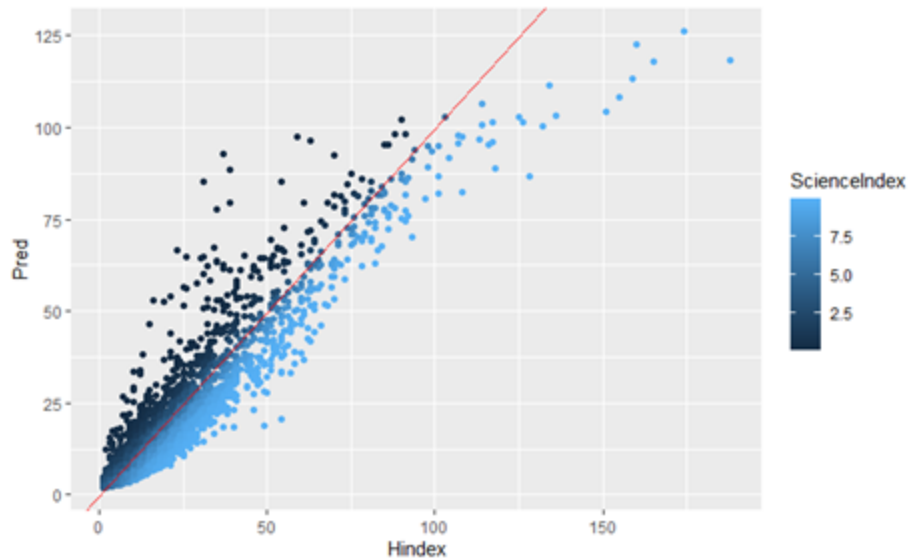
Application of the Data Sharing Ontology based Decentralized Knowledge Graph

- Mechanism to reward researchers for their data contributions
- Supplements the h-index
- To overcome the “cold-start” problem in our data-sharing dApp, we bootstrapped the SCIENCE-index with:
 - Publication data from the Microsoft Academic Graph and Semantic Scholar
 - Data citations from DataCite

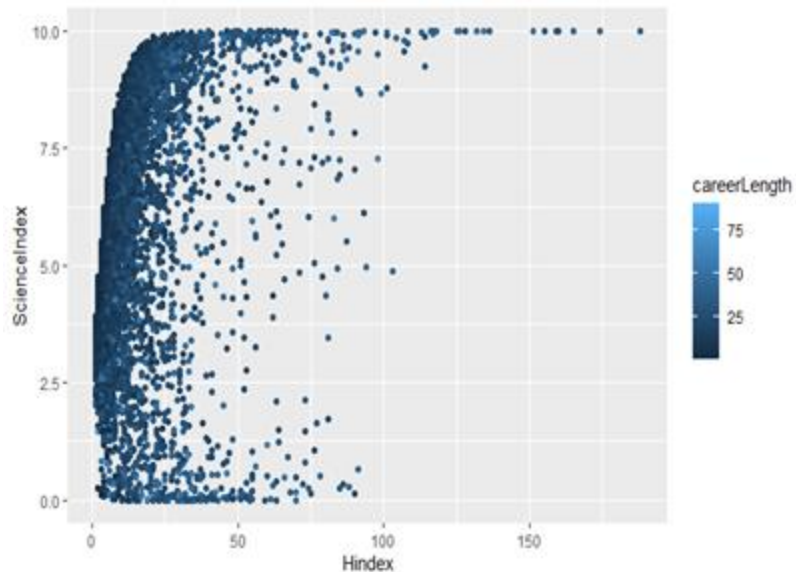


Results (1)

Predicted (proxy SCIENCE-INDEX vs Actual h-index



SCIENCE-index vs. Career Length

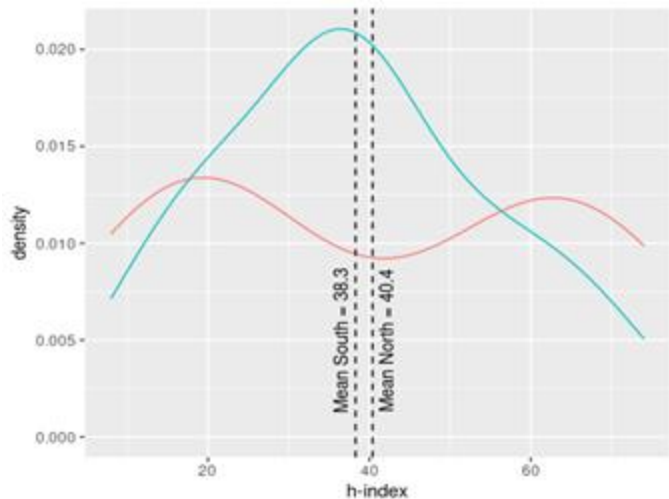




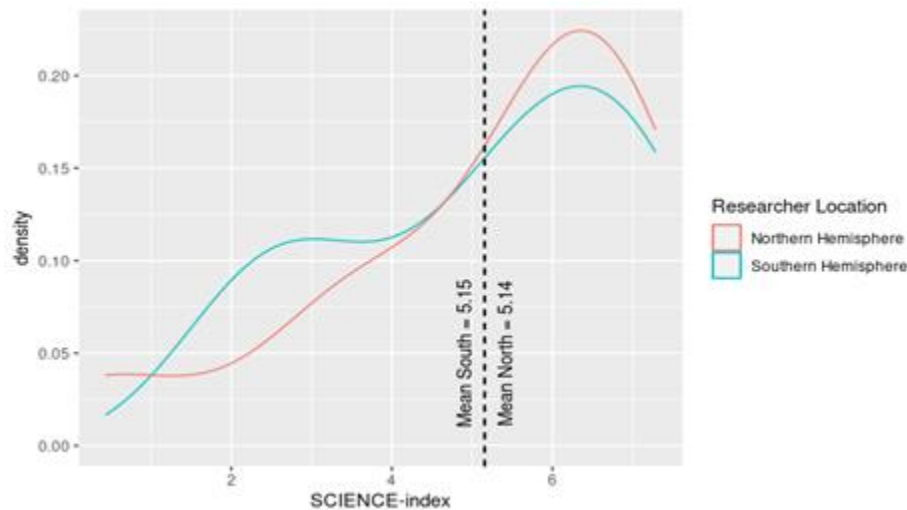
Results (2)

We present an evaluation of two groups of researchers (from the **Global North** and from the **Global South**) comparing their h-index to their SCIENCE-index to assess the **equity** of the two indices

Researchers' h-index



Researchers' SCIENCE-index

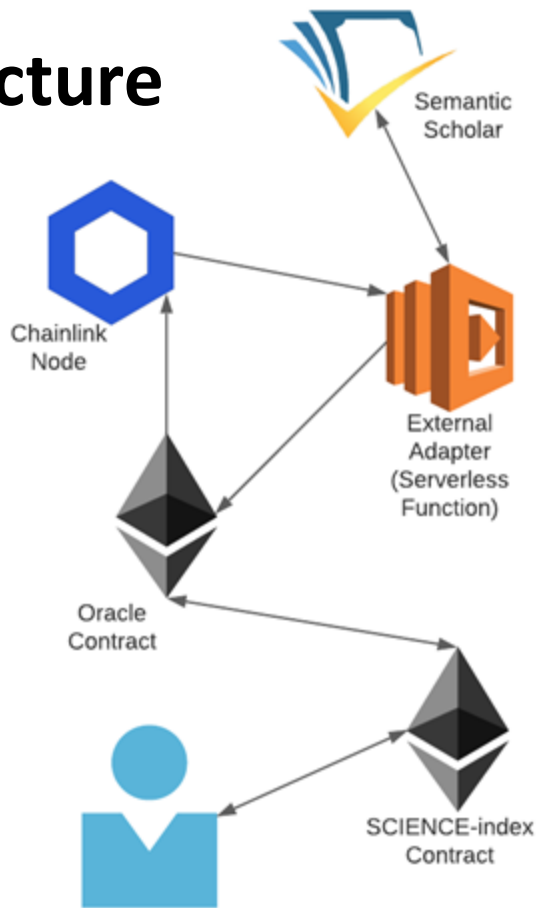




SCIENCE-Index Application Architecture

- SCIENCE-index is persisted as a **smart contract** and is exposed to a simple interface
- A researcher inputs into the contract their **Semantic Scholar ID**
- The contract utilizes a custom **oracle** that retrieves an author's career statistics from Semantic Scholar's API
- Using the statistics, our linear model calculates the researcher's **SCIENCE-index** and retrain itself from the new data
 - This keeps our index persistent and **up-to-date**

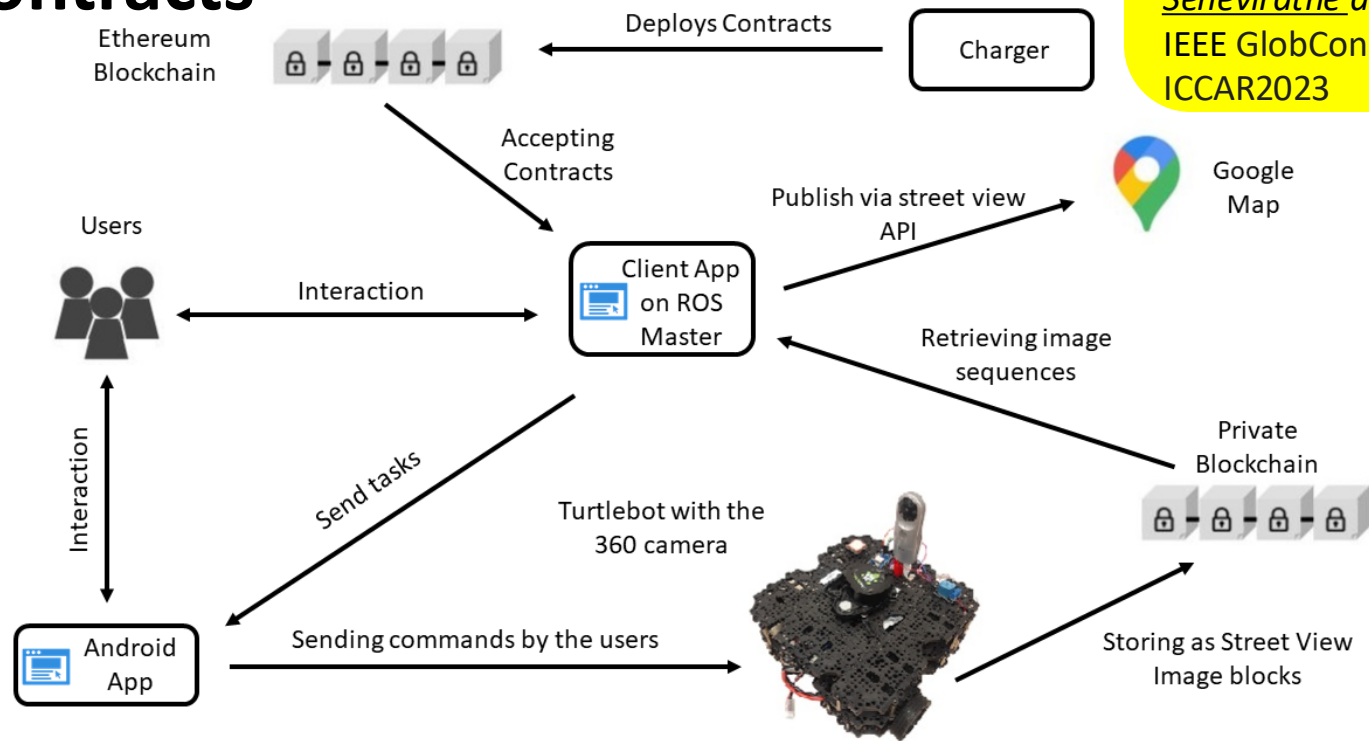
Code: <https://github.com/sharing-science>





Collecting Data for Decentralized Knowledge Graph with Smart Contracts

Decentralized Framework for Collection and Secure Storage of Google Street View Data: Case Study; Sanjaya Mallikarachchi, Bonnie Ho, Iyad Kanj, Oshani Seneviratne and Isuru Godage; IEEE GlobCon2023 and ICCAR2023





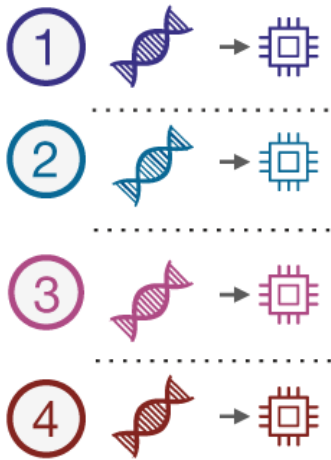
Knowledge Graph Utilization in AI Systems

- KGs provide structured, interconnected information that improves data representation, context understanding, and interoperability.
- KGs can help in:
 - Providing context for local data, helping to improve feature extraction and data preprocessing steps by incorporating domain knowledge.
 - Cleaning and validating the data by cross-referencing with the knowledge graph, ensuring higher data quality and consistency.



Different Learning Architectures

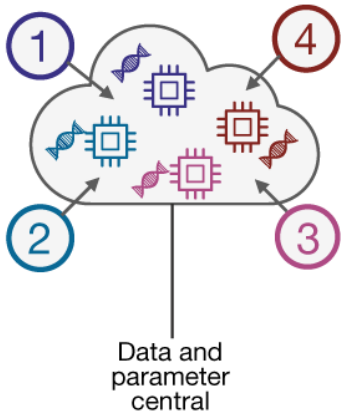
Local learning



Disconnected

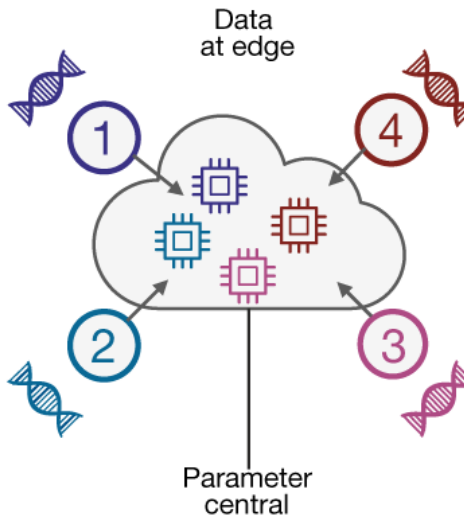
Localized data and model.

Central learning



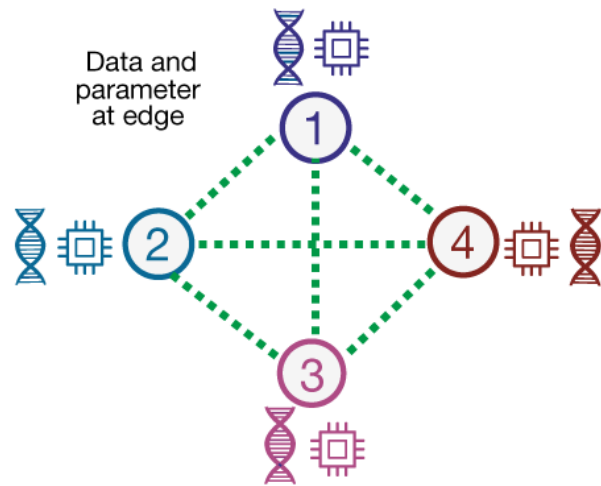
Cloud-based model with good accuracy, but at the cost of privacy and other issues.

Federated learning



Models are trained at the edge; only parameters are shared and merged by a central coordinator.

Swarm Learning

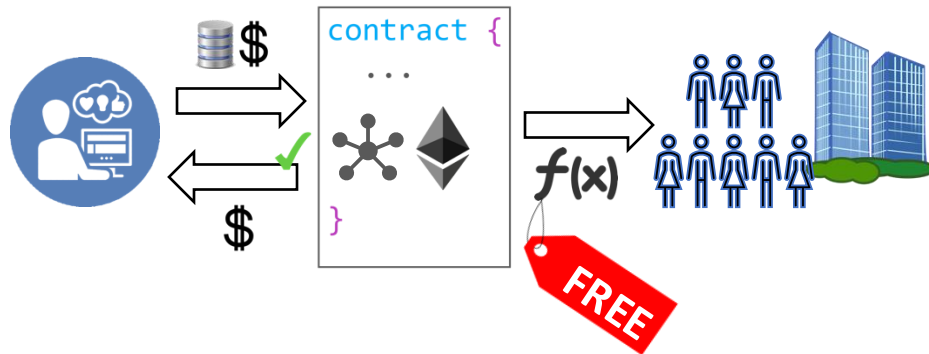


“Democratic [and Incentivized] Machine Learning” with both the **data and parameters** at the edge.



Blockchain-based AI/ML Model Training and Inference

- 1) Deploy an initial model
- 2) Contributors submit data + deposit
- 3) Contributors can get a reward after submitting good data
- 4) The model remains free to use for inference

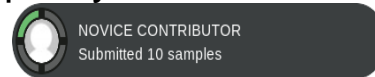




Incentivizing Quality Data

There are many ways to encourage contributors to submit good-quality data.

1. **Gamified** (non-financial, points + badges like Stackoverflow)
2. Based on established theory in **Prediction Markets**
3. Deposit, Refund, and Take: **Self-Assessment**

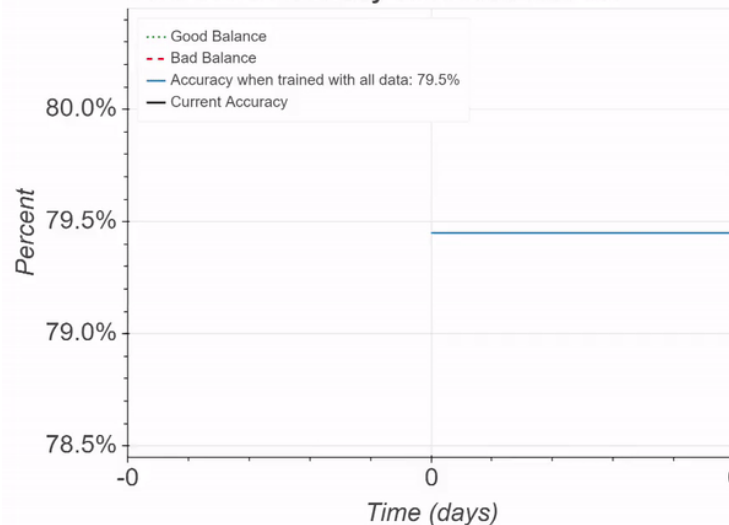


Our Simulation (PredictChain, Ledger journal 2024)

Assumption: “Bad Agent” frequently adds incorrect data.

- The model can still maintain accuracy.
- Honest contributors can still profit.

Balances & Accuracy on Hidden Test Set



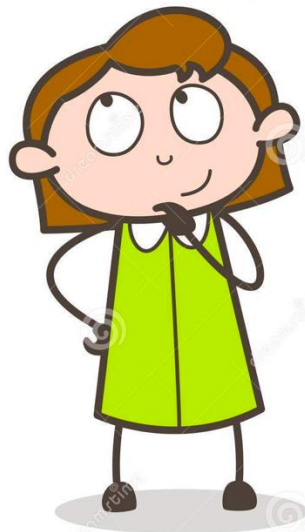


Food Knowledge Graph

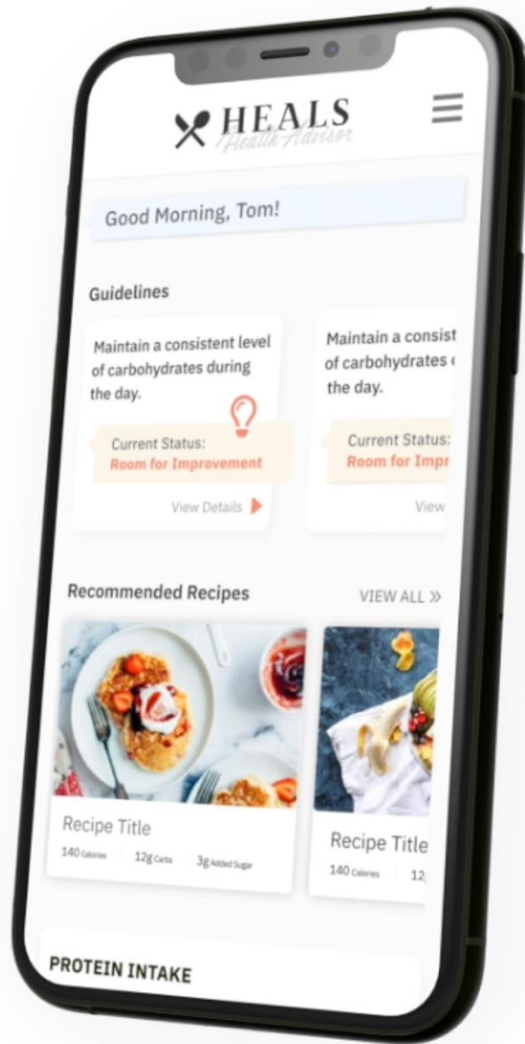
Food BOT

Your question:

Send



<https://foodkg.github.io/demo.html>





Customization with Personal Health Knowledge Graphs



HEALS *Health Assistant*

Bot: Hello Jennifer!

Jennifer: Suggest a good breakfast with Eggs

Bot:

- Sue's Goldenrod Eggs
- Featherbed Eggs
- Cheese Puffs
- Hole in the Wall
- Barefoot Contessa's Breakfast Bread Pudding
- Sausage Egg Casserole
- Southwestern Casserole
- Dunelm of Chicken
- Sylvia's French Toast
- Stuffed French Toast

Type a message... Send

Jennifer's PHKG:

- 35 years
- Female



HEALS *Health Assistant*

Bot: Hello Robert!

Robert: Suggest a good breakfast with Eggs

Bot:

- Italian Breakfast Biscuit
- Farmer's Casserole
- Scrambled New York Breakfast Wrap (South Beach Diet P2)
- Perfectly Poached Eggs
- Southern Eggs En Cocotte
- Rum Custard

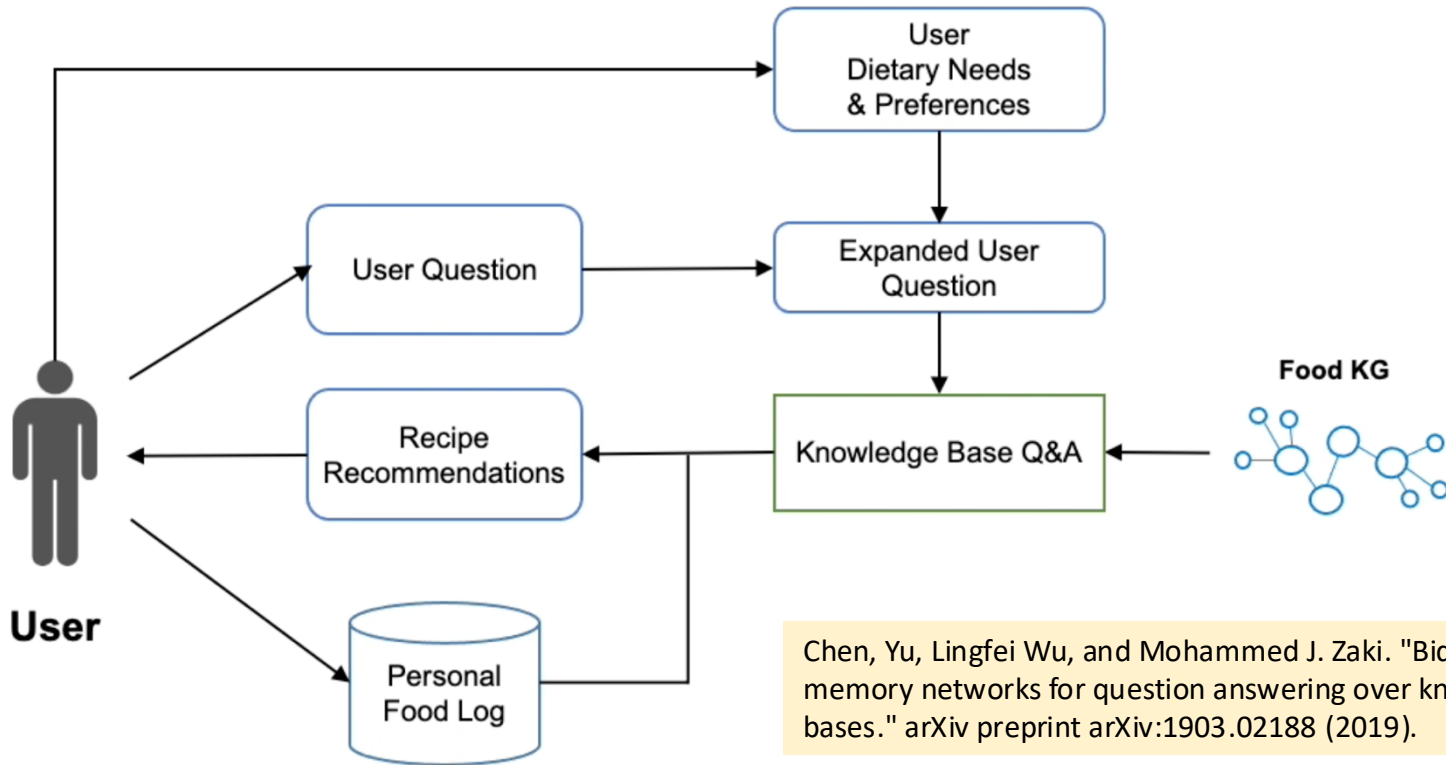
Type a message... Send

Robert's PHKG:

- 65 years
- Male
- Diabetic



Behind the Scenes



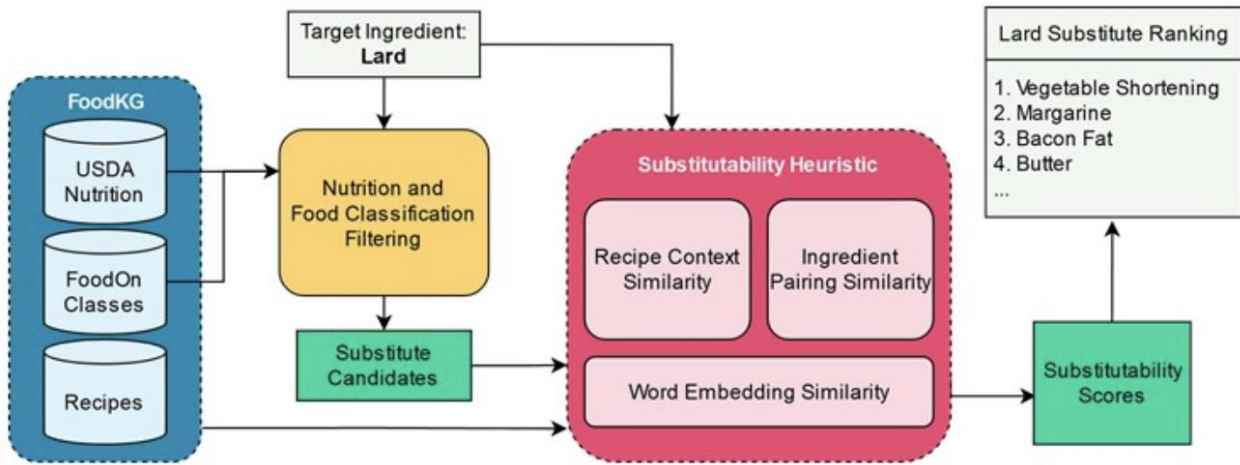
Chen, Yu, Lingfei Wu, and Mohammed J. Zaki. "Bidirectional attentive memory networks for question answering over knowledge bases." arXiv preprint arXiv:1903.02188 (2019).



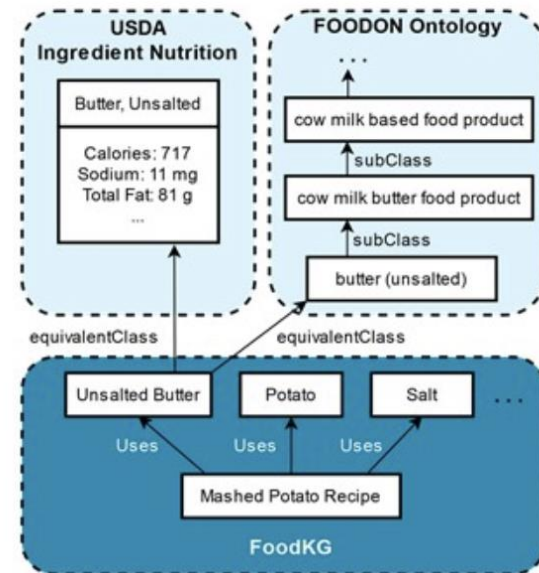
Ingredient Substitutions

Diet-Improvement Ingredient Substitution Heuristic (DIISH)

Implicit Semantics



Explicit Semantics



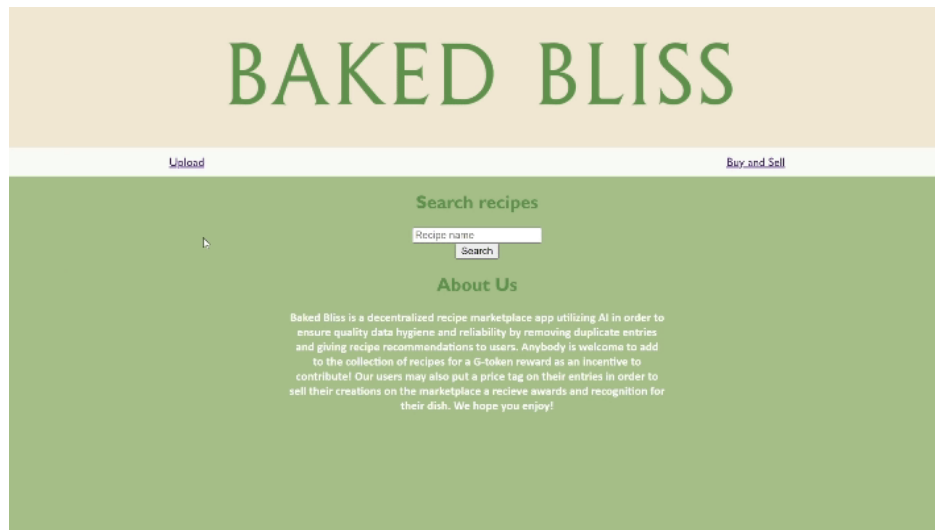
Sola S. Shirai, Oshani Seneviratne, Minor E. Gordon, Ching Hua Chen, and Deborah L. McGuinness. 2020. Semantics-Driven ingredient substitution in the FoodKG. CEUR Workshop Proceedings 2721 (2020), 243–247.



Incentivizing Good Quality Data Entry to the FoodKG



- Blockchain-based mechanism
- Users are incentivized to submit good-quality recipes.
- **Deposit, Submit, and Refund & Reward** model.
- **Reward** if the recipe is novel.
- (Novelty is determined using an embedding/vector similarity mechanism.)
- Other users can **buy** the “rights” to publish the recipe, just like an NFT.



Code: https://github.com/AI-and-Blockchain/F23_BakedBliss



Analyzing Crypto Twitter / Social Media

- There are lots of cryptocurrency scams on the Web!
- More than 320 transactions (with total value > 100K USD) had already taken place before Twitter took down the posts!

Can we figure out a way to flag these scam addresses?

Yes, by analyzing the Transaction Graph.

Example pattern:

- Scammers inject money into the scam address before the scam event.
- After some money is added, it is transferred through multiple accounts to obscure the scammer's identity.

← Tweet



Joe Biden ✓

@JoeBiden

I am giving back to the community.

All Bitcoin sent to the address below will be sent back doubled! If you send \$1,000, I will send back \$2,000. Only doing this for 30 minutes.

bc1qxy2kgdygjrqtzq2n0yrf2493
p [REDACTED]

Enjoy!

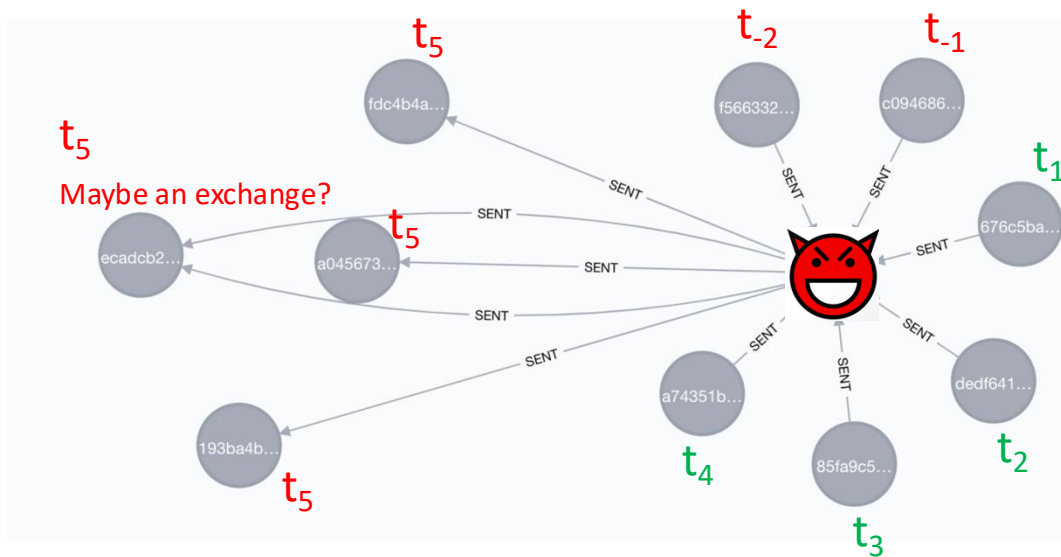
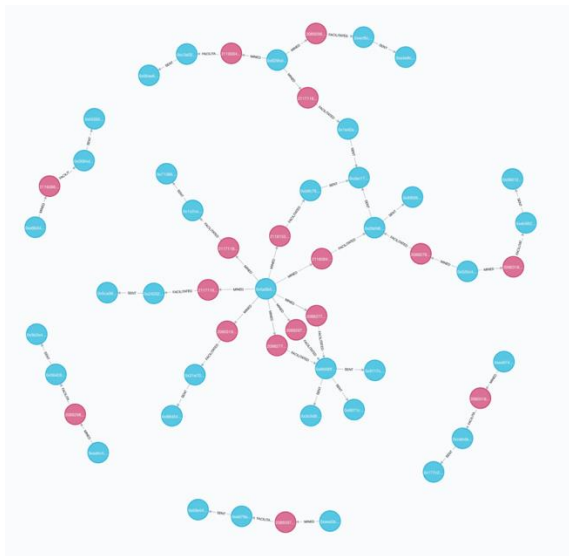
4:22 PM · 15 Jul 20 · [Twitter Web App](#)

2,375 Retweets and comments 1,739 Likes



Integrative Blockchain Provenance Analyzer

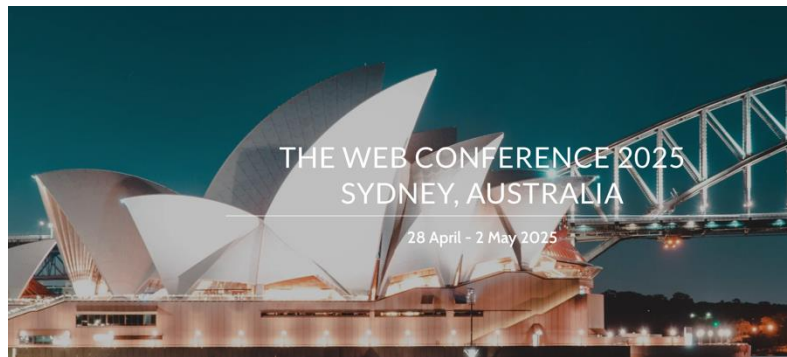
- Use provenance to identify suspicious addresses.
- Does an effective job at quickly calculating a reasonable suspicion flag for a selected address.





Several Upcoming Web-related Conferences

- ACM Web Conference'25
- ACM Web Science Conference'25
- ACM Web Search and Data Mining'25
- International Semantic Web Conference'24
- ...





Key Takeaway and Q&A

There are a lot of research potentials to explore the synergistic combinations of the Web, Knowledge Graphs and AI.

Any questions? Please feel free to contact me:
senevo@rpi.edu

Linking the World's Information

*Essays on
Tim Berners-Lee's
Invention of the
World Wide Web*

Oshani Seneviratne, James Hendler (Editors)



ASSOCIATION FOR COMPUTING MACHINERY