Sola S. Shirai
https://solashirai.github.io/
Email : solashakashirai@gmail.com
Publication List (Google Scholar)

Education

- **Rensselaer Polytechnic Institute (RPI)** — Troy, NY
  Ph.D. in Computer Science - Advised by Deborah L. McGuinness : GPA 4.0/4.0 — *Aug. 2019 – Present*
- **Worcester Polytechnic Institute (WPI)** — Worcester, MA
  B.S. with High Distinction in Computer Science, Minor in Data Science: GPA 3.9/4.0 — *Aug. 2015 – May 2019*

Research Interests

My research interests lie in the areas of knowledge representation and reasoning as well as natural language processing for artificial intelligence applications. My recent work has particularly focused on developing context-aware methods to more effectively utilize semantic web technologies, such as knowledge graphs and ontologies, in order to produce new knowledge through interpretable and interoperable AI models.

Research Experience

- **Rensselaer-IBM AI Research Collaboration Summer Researcher** — May 2022 - Aug. 2022
  IBM Research — *Yorktown Heights, NY*
  - Research extern at T.J. Watson Research Center, mentored by Debarun Bhattacharjya and Oktie Hassanzadeh.
  - Conducted research project involving the use of causal knowledge graphs to perform forecasting about news events, using reasoning-based methods to predict properties of unseen events based on similar past cause-effect events.
  - Contributed novel forecasting functions to an event forecasting and analysis toolkit.
- **Knowledgebase Intern** — May 2021 - Aug. 2021
  Robert Bosch LLC. — *Remote (Sunnyvale, CA)*
  - Research intern with Bosch's Human-Machine Interaction research group, mentored by HyeongSik Kim.
  - Proposed and conducted a research project involving flow graph construction and graph embedding methods applied to procedural instruction text, specifically for automatically performing ingredient substitutions in cooking recipes.
- **Research Assistant** — Aug. 2019 - Present
  RPI — *Troy, NY*
  - Researcher under IBM-RPI's HEALS (Health Empowerment by Analytics, Learning, and Semantics) project.
  - Lead and collaborator in projects that apply resources such as ontologies, knowledge graphs, and NLP tools to support personal health applications, especially surrounding food and eating habits.

Research Projects

- **Integrating Spatial Omics Data with Knowledge Graphs for Cancer Analysis** Sep. 2022 - Present
  HEALS Research Project — *Troy, NY*
  - Ongoing project to explore the application of semantic technology-based methods to spatial omics data of cancer images in order to perform prediction and analysis tasks.
  - Investigating methods to incorporate large-scale bioinformatics knowledge bases of proteins and their interactions together with the spatial topology and protein expression data of individual cells.
  - Developing reasoning-based methods to predict patient survival rates using explainable and generalizable models.
- **Event Prediction using Case-Based Reasoning over Knowledge Graphs** — May 2022 - Oct. 2022
  Extern Project with IBM Research — *Yorktown Heights, NY*
  - Curated and released a dataset of cause-effect events and their surrounding neighborhood, sourced from Wikidata.
  - Developed a case-based reasoning method to perform inductive link prediction, enabling interpretable predictions about unseen events without the need for model training.
  - **Outcomes:** To appear in WWW 2023 Semantics and Knowledge Research Track, "Event Prediction using Case-Based Reasoning over Knowledge Graphs" (Lead). Wikidata Workshop at ISWC 2022, "Rule-Based Link Prediction over Event-Related Causal Knowledge in Wikidata" (Lead). DL4KG Workshop at ISWC 2022, "Knowledge Graph Embeddings for Causal Relation Prediction" (Co-author). ISWC 2022 Industry Track, "Knowledge-Based News Event Analysis Toolkit" (Co-author). Demo under submission to IJCAI 2023.

- **Defining and Using "Context" in Knowledge-Driven Systems** Oct. 2021 - Present
  HEALS Research Project *Troy, NY*
  - Ongoing project to explore how the concept of "context" is framed and utilized in knowledge graph research.
  - Investigating methods to identify contextually "interesting" and "useful" information to support explainable AI.
  - Developing methods to unambiguously represent various facets and perspectives of context to enable greater interoperability and communication of scientific results surrounding context-aware methodologies.
  - **Outcomes:** Grand Challenges in Personal Informatics and AI Workshop at CHI 2022, "Towards Context Clarity in Personal Informatics Applications" (Lead), "Realizing the Potential of Personal Health Informatics Through A Personal Semantic Health Knowledge Graph" (Co-author)

- **Procedural Instruction Modification using Flow Graphs** May 2021 - Oct. 2021
  Intern Project with Robert Bosch LLC. *Remote (Sunnyvale, CA)*
  - Generated flow graphs from cooking recipe text using dependency parsing tools and domain ontologies.
  - Developed strategy to embed flow graphs to support the use case of automatic ingredient substitution.
  - **Outcomes:** Patent application filed. ISWC 2022 Research Track, "EaT-PIM: Substituting Entities in Procedural Instructions Using Flow Graphs and Embeddings" (Lead)

- **Framework for Recommendations with Explanations** Jul. 2020 - Mar. 2021
  HEALS Research Project *Troy, NY*
  - Developed a lightweight pipeline framework for explainable recommendations using knowledge graphs,[1] demonstrated for a university course-recommendation use case and healthy food recommendation use case.
  - Supports object-oriented programming workflows using data from RDF data sources and ontology modeling
  - **Outcomes**: ISWC 2021 Poster&Demo Session, "Healthy Food Recommendation and Explanation Generation using a Semantically-Enabled Framework" (Lead)

- **Ingredient Substitution using a Knowledge Graph of Food** Dec. 2019 - Jun. 2020
  HEALS Research Project *Troy, NY*
  - Devised a heuristic model to identify good ingredient substitutions to empower patients to make healthier meals.
  - Utilized a knowledge graph of food and word embeddings to capture explicit and latent semantic information.
  - **Outcomes:** Frontiers in Artificial Intelligence Journal 2021, "Identifying Ingredient Substitutions Using a Knowledge Graph of Food" (Lead). ISWC 2020 Poster&Demo Session, "Semantics-Driven Ingredient Substitution in the FoodKG" (Lead). Presentation at AI & Food and Nutrition at AMLD EPFL 2021, "Utilizing a Food Knowledge Graph for Healthy Ingredient Substitutions"

- **Personal Health Knowledge Graphs (PHKG)** Feb. 2020 - Apr. 2020
  HEALS Research Project *Troy, NY*
  - Investigated existing literature and identified key challenges to develop PHKGs.
  - Assisted in organizing and presenting at the PHKG workshop at the 2020 Knowledge Graph Conference.
  - **Outcomes:** PHKG Workshop at KGC 2020, "Applying Personal Knowledge Graphs to Health" (Lead)

- **Generating Surrogate Facial Images for Crowdsourcing** Aug. 2018 - Mar. 2019
  WPI Undergraduate Capstone Project *Worcester, MA*
  - Developed generative adversarial networks to generate fake facial images that retained facial expressions.
  - Utilized surrogate images with crowdsourcing, enabling annotation while preserving the privacy of the original images.
  - **Outcomes:** CV-COPS Workshop at CVPR 2019, "Privacy-Preserving Annotation of Face Images Through Attribute-Preserving Face Synthesis" (Lead)

- **ASSISTments Open-Response Automatic Grading** May 2018 - Aug. 2018
  WPI Undergraduate Research Project *Worcester, MA*
  - Applied NLP methods for automatic grading of open-response math questions.
  - **Outcomes:** Contributed to a Poster presented at AIED 2019.

Additional Experience

- **DBpedia Hackathon - Knowledge-Graph Shiritori Application** Sep. 2020
  DBpedia Autumn Hackathon Project *Troy, NY*
  - Developed a game of "shiritori" with the goal of making connections between entities and facts.
  - Used Diffbot's APIs to parse user input, extract entities and facts, and identify connections [2].

[1] FREx github: https://github.com/solashirai/FREx – Additional FREx documentation: https://tetherless-world.github.io/FREx/
[2] Short demo video available at https://www.youtube.com/watch?v=BtSgWrNE7M8

- **Student Software Engineer - ASSISTments** May 2016 - Nov. 2018
  Student Software Engineer *Worcester, MA*
  - Performed various maintenance and improvements for front-end systems and interaction with student data.
  - Contributed to development of ASSISTments SDK for transitioning the system into Java.
  - Lead project to develop new user interfaces for teachers to create course content.

- **Big Data in Denmark's Waste Management Sector** Fall 2017
  WPI Interdisciplinary Project with Dansk Affaldsforening *Copenhagen, Denmark*
  - Assessed big data collection, management, and usage in Denmark's waste management industry.
  - Interviewed field experts and traveled to various municipalities to conduct on-site observations.
  - Project report: "Preparing for the Use of Big Data in Denmarks Waste Management Sector"

- **edX Internship** Summer 2014
  Research Science Institute Summer Internship *Cambridge, MA*
  - Developed a course component for edX to enable crowdsourcing of hints for student homework questions.

Skills

- **Programming:** Python > Java > C# > Javascript > C/C++
- **Misc. Tech.:** SPARQL, RDF, Ontologies[3], SQL, Git, Docker, Prolog, LaTeX, TensorFlow/Keras, Pytorch
- **Languages:** Native speaker of Japanese and English

[3]Example ontology for course recommendation: https://rpi-ontology-engineering.netlify.app/oe2020/course-recommender/ontology