Basel Shbita, Ph.D. **(he/him/his)**

CONTACT INFORMATION

Address: San Jose, CA, USA *WWW:* https://shbita.com/
E-mail: basel921@gmail.com

EDUCATION

University of Southern California; Los Angeles, CA, USA **Aug 2018 - May 2024**
PhD, Computer Science
Research Interests: Knowledge Graphs • Machine Learning • Geospatial Artificial Intelligence • Information Integration • Semantic Web • Data Science • Large Language Models (LLMs)
Dissertation: "*Transforming Unstructured Historical and Geographic Data into Spatio-Temporal Knowledge Graphs*" (Advisor: Prof. Craig A. Knoblock)

MSc, Computer Science **Dec 2023**
GPA: 3.771/4.0

Tel Aviv University; Tel Aviv, Israel **Feb 2011 - Jan 2015**
BSc, Electrical and Electronic Engineering
GPA: 93.33/100 (*magna cum laude*)

EXPERIENCE

IBM Research; Almaden, San Jose, California, United States **June 2024 - Present**
Research Scientist, *AI Research*

- Working on generative AI & large language models (LLMs) as part of the *Granite* project
- Developing debugging toolkits for LLMs that streamline issue analysis & resolution across workflows
- Integrating knowledge graphs (KGs) with LLMs to optimize data integration for enterprise
- Researching methods for synthetic data generation to enhance model robustness & evaluation
- Studying methods & techniques for semantic annotation of semi-structured & unstructured data
- Worked on model-agnostic & customizable solutions to safeguard LLM deployments

Information Sciences Institute; Marina del Rey, CA, USA **Aug 2018 - May 2024**
Research Assistant, *Center on Knowledge Graphs*

- Worked on knowledge graphs (KGs) with an emphasis on automatic spatio-temporal & semantic interpretation of topographic map data as a means to solve complex information integration problems
- Investigated new methodologies to leverage machine & deep learning techniques to establish automatic data understanding & KG construction in different domains
- Partook in several projects: MinMod (AI for mineral data), Linked Maps (constructing KGs for spatio-temporal data), MINT (data integration for scientific modeling) & Table Understanding (automated semantic interpretation of tables)
- Semi-finalist at the Amazon Alexa Prize Socialbot Grand Challenge 4 (knowledge integration)
- Supervised & mentored USC MS student-workers & summer interns

IBM Research; Almaden, San Jose, California, United States **May - Aug 2022**
Research Intern, *re*THINK Enterprise*

- Worked on knowledge graphs, ML & data integration in the domain of business transformation
- Designed, implemented & evaluated a user-assisted automatic & fully functional pipeline to convert unstructured client requirements textual data into a contextualized knowledge graph

General Electric Global Research; Niskayuna, NY, USA (online) **May - Aug 2021**
Research Fellow Intern, *Analytics Software & Knowledge Discovery*

- Worked at the Artificial Intelligence Technical Domain
- Designed, implemented & evaluated an infrastructure for an automated generation of classification & annotation rules for control concepts in Cyber-physical Systems software using Inductive Logic Programming & Semantic Technologies

Apple; Herzliya, Israel **Jan - Jun 2018**
Embedded Firmware Engineer, *Flash Storage Software Department*

- Designed & developed complex software modules for ultra high performance, real-time embedded systems in a multiprocessor environment for Apple products
- Defined firmware features & led HW-FW integrations
- Completed vertical integration with other modules in storage stack (i.e. drivers, file-system)

Mellanox Technologies; Tel Aviv, Israel **2011 - 2017**
Senior Firmware Engineer & Team Lead, *Switch Silicon Core Department* **Apr - Jun 2017**

- Managed team of 5 engineers
- Led the 100GbE Switch Systems firmware development process & software infrastructure activities - an operation involving 16 engineers
- Developed a distributed functional testing environment; debug tools & performance testing in C++ in Unix for both Ethernet & InfiniBand fabrics in OSI Data-Link & Network layers
- Implemented complex verification architectures consisting of static & dynamic analysis
- Delivered technical presentations to 50+ engineers & developers from various teams including: software, hardware, production & qualification
- Awarded for excellence & ranked "Superior" (top 5% out of 3000)

Team Lead Firmware Engineer, *Switch Silicon Core Department* **Apr 2015 - Mar 2017**

- Managed team of 3 engineers
- Contributed to end-to-end development & defined version-release procedures for the company's 100GbE Switch products; conducted personal training, code reviews; defined coding-style & methodologies of software engineering for team of 30
- Optimized shared library cross-platform code, resulting in ∼40% reduction in project compilation time for 90+ department developers
- Led full silicon bring-up process, both Pre-Silicon & Post-Silicon stages
- Supervised integration processes operating in Beijing (China), Seattle & Sunnyvale (US)

Firmware Engineer, *Switch Silicon Core Department* **Oct 2011 - Mar 2015**

- Developed ANSI-C compiler-specific code to run on Switch Systems RISC
- Implemented C++ & Python-based regression testing scripts
- Introduced simulation tool to reduce ∼50% time in firmware development cycle
- Developed "Stress" tool that has become major tool for system production/screening, power measurements & debug process
- Led the planning & priorities coordination procedure with software architecture & SDK teams
- Collaborated with chip-design & software engineers to create a Post-Silicon Random Verification Environment for the 56GbE & EDR (100Gbps Infiniband) technologies

HP Autonomy (Virage); Cambridge, MA, USA **May - Aug 2009**
Computer-Science Research Intern, *Advanced Technology Group*

- Studied Statistical Natural Language Processing, Automatic Speech Recognition & Machine Learning; performed model training & testing with Python
- Wrote paper titled "*Improving Text-Independent Speaker Identification Performance Using Gaussian Mixture Speaker Models*" under direction of Dr. David Palmer (HP Autonomy (Virage), Inc)
- Conducted on-campus coursework in scientific theory & off-campus work in scientific research over 6 weeks under sponsorship of the Center for Excellence in Education (CEE) & the Massachusetts Institute of Technology (MIT) as part of the Research Science Institute (RSI)

TEACHING EXPERIENCE

University of Southern California; Los Angeles, CA, USA **Spring & Fall 2020**
Teaching Assistant, *DSCI 558: Building Knowledge Graphs*

- Designed & evaluated course examinations, written bi-weekly assignments & weekly quizzes
- Held weekly office hours (2 hours each)
- Designed & delivered 3+ sessions of 2-hour lectures (includes core-material classes & guest lectures)

AWARDS & SCHOLARSHIPS

University of Southern California; Los Angeles, CA, USA **2022**
University Outstanding Teaching Assistant Award (Highest Achievement, grant valued $1k)

Alexa Prize Socialbot Grand Challenge 4; Seattle, WA, USA **2021**
Our team was the recipient of a research grant valued $250k as part of the competition

Modeling and Managing Complicated Systems Institute; Pittsburgh, PA, USA **2019**
Recipient of the Ford Foundation Graduate Student Grant (value $1.6k)

Tel Aviv University; Tel Aviv, Israel **2013**
Faculty Dean's List of Outstanding Undergraduate Students

Tel Aviv University; Tel Aviv, Israel **2013**
Recipient of the Freescale Semiconductor Israel Excellence Scholarship (value $1.2k)

SELECTED PUBLICATIONS

B. Shbita, B. Vu, F. Lin, and C. A. Knoblock. Embedding Spatial and Semantic Contexts for Geo-Entity Typing in Smart City Applications. In *11th International Smart City Workshop co-located with the 2025 ACM Web Conference (WWW)*. ACM, 2025.

B. Shbita, A. L. Gentile, and C. E. DeLuca. Implementing Personal Knowledge Graphs. United States Patent Application. *Publication Number: 20250007723*. USPTO, 2025.

B. Vu, **B. Shbita**, C. A. Knoblock, and F. Lin. Exploiting Distant Supervision to Learn Semantic Descriptions of Tables with Overlapping Data. In *The Semantic Web–ISWC 2024: 23th International Semantic Web Conference (ISWC)*. Springer International Publishing, 2024.

B. Shbita, N. Sharma, B. Vu, F. Lin, and C. A. Knoblock. Constructing a Knowledge Graph of Historical Mining Data. In *6th International Workshop on Geospatial Linked Data (GeoLD) co-located with the 21st Extended Semantic Web Conference (ESWC)*. CEUR, 2024.

B. Shbita and C. A. Knoblock. Automatically Constructing Geospatial Feature Taxonomies from OpenStreetMap Data. In *2024 IEEE 18th International Conference on Semantic Computing (ICSC)*, (208–211). IEEE, 2024.

Y. Chiang, M. Chen, W. Duan, J. Kim, C. A. Knoblock, S. Leyk, Z. Li, F. Lin, M. Namgung, **B. Shbita**, and J. H. Uhl. GeoAI for the Digitization of Historical Maps. In *Handbook of Geospatial Artificial Intelligence*, (217–247). CRC Press, 2023.

F. Lin, C. A. Knoblock, **B. Shbita**, B. Vu, Z. Li, and Y. Chiang. Exploiting Polygon Metadata to Understand Raster Maps: Accurate Polygonal Feature Extraction. In *Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems*, (1–12). 2023.

B. Shbita, A. L. Gentile, C. Deluca, P. Li, and G. Ren. Understanding Customer Requirements – An Enterprise Knowledge Graph Approach. In *Extended Semantic Web Conference*, (625–643). Springer Nature Switzerland, 2023.

B. Shbita, C. A. Knoblock, W. Duan, Y. Chiang, J. H. Uhl, and S. Leyk. Building Spatio-Temporal Knowledge Graphs from Vectorized Topographic Historical Maps. In *Semantic Web, 14(3)*, (pp. 527–549). IOS Press, 2023.

B. Shbita, and A. Moitra. Automated Generation of Control Concepts Annotation Rules Using Inductive Logic Programming. In *Functional and Logic Programming*, (pp. 171-185). Springer, 2022.

J. H. Uhl, S. Leyk, Z. Li, W. Duan, **B. Shbita**, Y. Chiang, and C. A. Knoblock. Combining Remote-Sensing-Derived Data and Historical Maps for Long-Term Back-Casting of Urban Extents. *Remote Sensing, 13(18), 3672*, 2021.

H. Cho, **B. Shbita**, K. Shenoy, S. Liu, N. Patel, H. Pindikanti, J. Lee, and J. May. Viola: A Topic Agnostic Generate-and-Rank Dialogue System. In *Proceedings of the 4th Alexa Prize*, 2021.

M. Mann, F. Ilievski, M. Rostami, A. Aastha, and **B. Shbita**. Open Drug Knowledge Graph. In *Proceedings of the 2nd International Workshop on Knowledge Graph Construction co-located with the 18th Extended Semantic Web Conference (ESWC)*. CEUR, 2021.

Z. Hu, Z. Zhao, M. Rostami, F. Ilievski, and **B. Shbita**. Knowledge Graph-Based Housing Market Analysis. In *Proceedings of the 2nd International Workshop on Knowledge Graph Construction co-located with the 18th Extended Semantic Web Conference (ESWC)*. CEUR, 2021.

Z. Li, Y. Chiang, S. Tavakkol, **B. Shbita**, J. H. Uhl, S. Leyk, and C. A. Knoblock. An Automatic Approach for Generating Rich, Linked Geo-Metadata from Historical Map Images. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 3290-3298). 2020.

B. Shbita, C. A. Knoblock, W. Duan, Y. Chiang, J. H. Uhl, and S. Leyk. Building Linked Spatio-Temporal Data from Vectorized Historical Maps. In *Extended Semantic Web Conference*, (pp. 409–426). Springer, 2020.

B. Shbita, B. Vu, D. Feldman, M. Pham, A. Rajendran, C. A. Knoblock, J. Pujara, and Y. Chiang. Creating a FAIR Data Catalog to Support Scientific Modeling. In *Workshop on Advanced Knowledge Technologies for Science in a FAIR World*, 2019.

B. Shbita, A. Rajendran, J. Pujara, and C. A. Knoblock. Parsing, Representing and Transforming Units of Measure. In *Modeling the World's Systems*, 2019.

ACADEMIC SERVICE

Program Committee, AAAI Conference on Artificial Intelligence **Present**
Program Committee (Industry track), International Semantic Web Conference (ISWC) **2022-Present**
Reviewer (Research track), ACM Knowledge Discovery in Databases Conference (KDD) **2024**
Reviewer (Research & In-use tracks), International Semantic Web Conference (ISWC) **2023**
Program Committee (Industry track), International World Wide Web Conference (WWW) **2023**
Reviewer (Research track), International Semantic Web Conference (ISWC) **2022**
Reviewer, International Joint Conference on Knowledge Graphs (IJCKG) **2021**

LANGUAGES

English • Arabic • Hebrew • Russian

TECHNICAL SKILLS

`Python, C, C++, C#, SWI-Prolog, MATLAB, Assembly` • `RDF/OWL, SPARQL, SQL, PostgreSQL, PostGIS` • `PyTorch, TensorFlow, Keras, scikit-learn, pandas, SciPy, NumPy, Matplotlib, Jupyter, Flask` • `Ruby on Rails, HTML, CSS, JavaScript` • `Git` • `Docker`